Proceedings of ASME Turbo Expo 2011
GT2011
June 6-10, 2011, Vancouver, British Columbia, Canada

GT2011-45061

TECHNOLOGY EVOLUTION OF TURBOFAN ENGINES BASED ON PATENT ANALYSIS, A NEW SIGHT TO TECHNOLOGY OPPORTUNITIES

V. Shamekhi*, K. Mazaheri
Center of Excellence in Aerospace Systems
Sharif University of Technology
11365-11155 Tehran, Iran
Email:
shamekhi_vahid@ae.sharif.edu,
mazaheri@sharif.edu

M. Ali
Faculty of Science
Lund University
Lund, Sweden
Email:
abi10mal@student.lu.se

A. Kokabi
Atinegaar Think Tank
Tehran, Iran
Email:
kokabi@atinegaar.com

ABSTRACT

In the current emerging knowledge economy, it is essential for industries to monitor knowledge progress in different technological areas. Here we use patent analysis and text mining techniques to develop a new approach to realize technology evolution in areas related to turbofan engines industry. Over three thousand patents from USPTO, UK-IPO, JPO, EPO and WIPO databases, are used. Some preprocessing algorithms are used to produce a list of high priority keywords and a few weighting matrices are developed to explore relation of different patents and keywords, and entities of these matrices are computed for three predefined time intervals. A clustering process is introduced to classify different patents and each cluster is named based on its contents. These results show us evolution of turbofan engines technology areas during these different time intervals. Two-dimensional maps are introduced to show correlation of different technology areas, and their relative importance. The introduced technique could be easily applied to other industries to generate valid meaningful information to support managers and policy makers to make decisions.

INTRODUCTION

In recent decade, there are several driving forces that increase the role of knowledge in economic growth. Rapid progress of science and technology, tough competitive business environments and incrementing level of customers' expectations are among driving forces that made it essential for governments, Industries, companies, policy makers, researchers and industry experts to be informed of the progress across different scientific and technological fields [1-4]. In such situation innovation and Research and Development (R&D) investments are becoming critical issues in maintaining competitive advantages [2, 5, 6]. On the other hands, innovation's rate are accelerating and product life cycles are becoming shorter [7]. Innovative ideas have to fulfill market needs, otherwise services or products based on them will fail in the market [8]. Senior management must decide how much of their R&D resources have to be spent on which type of technology [9]. So, in order to find the best business opportunities, strategic technology planning should be in convergence with business planning [2].

Being able to distinguish emerging trends and to anticipate direction of technology changes are critical issues for facilitating decision-making processes, prioritizing R&D projects, reducing the risk of technology development, and allocating resources [10]. Science and technology maps can be the ideal tools in these topics. Like 'geographic maps' that helps to understand environment in the physical world, by using science and technology maps, there will be more information about landscapes of science and technology in different disciplines [11, 12]. They help to have vision of targeted discipline.

Patents, as an output of innovative technology-related activities contain important research results that are valuable to the industry, business, law, and policy-making communities [13, 14]. Investigating patents' information help to clarify surrounding technological environments in different areas. In this study, by analyzing turbofan patents, try to develop a technology opportunity map for Turbofan Engines. With patent analysis techniques, try to provide a bird's eye view on high potential technical domains of Turbofan engines in future.

This paper is organized as follows. A brief review of patent analysis and application of text mining is presented in first part. Then the overall research process is described that is contained data gathering, data preprocessing, data analysis and interpretation and validation; and finally the case of Turbofan Engines industry and the results of this procedure are presented.

LITERATURE REVIEW

Patents are considered as tangible signs of technological progress and reliable measure of technical capabilities [3, 14-16].

Outputs of commercially useful innovative activities are usually registered as patents in international patent offices. These documents are open source and contain important research results that are valuable to the industry, business, law, and policy-making communities [13]. Due to patents' accessibility, scope of coverage and richness of information, substantial academic and industrial researches are accomplished with patent data for variant purposes [7, 9, 17-19].

In the next sections patent analysis is discussed briefly and applications of text mining in patent analysis are clarified.

Patent analysis

The detailed information contained in patent documents is valuable source for patent analysis. Two types of data are included in patent documents. Structured and unstructured data type [16, 20]. In structured part standardized fields with unique formats such as patent number, issue date, patent citation, inventors, assignees, technology fields, country and city of the assignee are provided. Unstructured parts such as claims, abstracts, titles or descriptions of the invention are free texts and have different length and contents.

Advantages of patent analysis can be discussed in macroscopically and macroscopically level [5, 7]. In macro view, patents are indicator of measuring techno economic growth. They have variant uses in Policy making issues such as determining R&D management strategy at the government level and comparison of countries' productivity based on their technology strength [2, 7]. In micro level patent analysis can help organization in their R&D portfolio management, designing new products, identification and assessment of future technological knowledge options, human resource management and monitoring competitor's technical capabilities [7, 9, 20-23].

It should be considered that although patents have extensive uses and contain valuable information, patent counts and analyses provide only a limited measure of the R&D activity and are not sufficient to be used as a sole metric of R&D. But when used in combination with other metrics, patents offer a manageable piece of data on the level of R&D effort by individuals and firms.

Citation analysis and network analysis are two main approaches in patent analysis [7]. Citation analysis has been the most frequently adopted method which is used for estimating knowledge flow [14, 24-27]. By using bibliographic citation fields of patents, this method aims to describe the relationship between patents by tracing the citations of a patent in the subsequent patents or investigates the relations of science and technology through monitoring the non-patent references in patents. This approach is based on the assumption that highly cited patents, supply basic knowledge to subsequent patents and thus highly cited patents considered as extremely important technological areas [10, 21]. However, citation analysis is subject to some crucial drawbacks. This metric provides very small attribute of the R&D activity and technological development that it takes only citing–cited linkages into account. Thus, the internal relationship among patents is not described and the quality of citations usually disregarded [7, 15]. Since it only uses the bibliographic citation field in patents, the scope of analysis and the richness of information are limited [28].

Network analysis examines the whole structure of patent documents and illustrates the relationship among patents as a visual network. Patents' information account for nodes and the relationships among patents represent edges in the network. by visualizing the locations of individual patents and linkage patterns among them, it becomes possible to view the overall landscape on a global scale from different perspectives [29]. The scope of analysis is wider in this approach because more information can be taken into accounts [7]. Text mining is the common technique applied in this method. In the next section, applications of text miming in patent analysis are discussed.

Application of text mining in patent analysis

Text mining can be used in patent analysis as information extracting tool and help researchers in their decision making process [7, 21, 30]. Text mining is often regarded as a process to find implicit, previously unknown, and potentially useful patterns from a large text repository. It consists of a variety of techniques which include text segmentation, summary extraction, keyword identification, feature selection, term association, term clustering, cluster generation and information visualization [13, 16]. Automatic techniques for assisting patent analysis are still in great demand [13] and text mining techniques are very useful in this matter.

THE OVERALL RESEARCH PROCESS

In this section all the steps involved in developed research process will be clarified. The overall research process that developed has been shown in Figure 1. It includes data gathering, data preprocessing, data analysis and interpretation and validation of results. In the following part the details of each step will be described. It should be noted that these are the different steps used in this study; most of these procedures are common steps in data mining researches and we customize them for solving our research question.

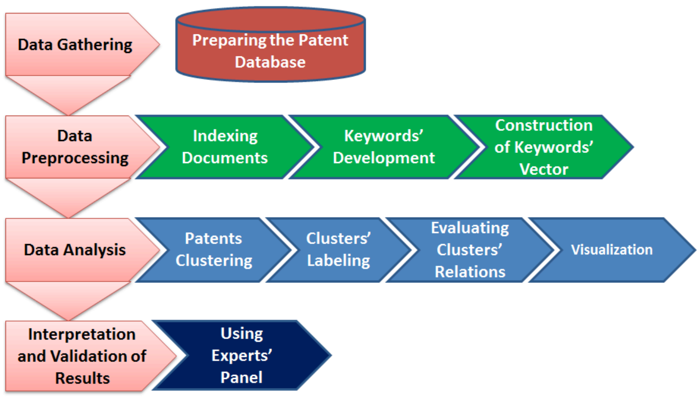

Figure 1. THE OVERALL RESEARCH PROCESS

Data gathering

The first step of the process is data gathering. Patents that related to the subject of research (specially the title and claim of patents) should be recognized and collected. Scopus database has been chosen for gathering data because it collects patents from different patent offices [31]. The multi-thread web downloader was made for this purpose and the needed data was downloaded and categorized successfully.

An application made to parse text elements that contain all the information of collected patents. This information includes patents' titles, abstracts, descriptions, claims, inventors, assignees and the filing dates.

Data preprocessing

Some preprocessing tasks have to be done in this step. Indexing documents, extracting the most significant or representative words and ranking the keywords are some of the preprocessing tasks that have been done in this research. The detail procedures of each step have been described in the following part.

Indexing documents: Indexing documents has been done by the Microsoft SQL Server 2008 Enterprise (trial edition) [32]. All patent documents were indexed using Microsoft SQL Server Full-Text Services [33]. The output of this section is indexed table which is containing of information that we need for next section.

Development of keyword: The main unit of analysis used in this research consists of the keywords that extracted from the unstructured text parts of the patents. The main task here was finding the suitable method for extracting key words from the text part of our documents. In order to do so, the MS SQL Server – Full-Text services are used. This service uses the TF-IDF index for extracting the keywords. TF-IDF index is calculated according to the following formula:

$$w_t = TF(d,t)IDF(t) \tag{1}$$

Where TF is the frequency of word t in document d and IDF is calculated with the following formula:

$$IDF(t) = -\log_2 ({DF_t}/{N}) \tag{2}$$

Where N is the total number of documents and DF_t is number of documents that contain term t. The importance increases proportionally to the number of times a word appears in the document but is offset by the frequency of the word in the entire document collection. As a result, terms that occur in a lot of documents are considered common terms and are down-weighted [34]. Some experts then revise these keywords; also experts added some key words to the list. For the case of turbofan engine the number of keywords that remain after revision was 136 keywords.

Construction of keyword vector (Matrix): Finally all the extracted keywords were transferred in a term-by-document matrix in which, each column represents the keywords and each row represents the documents. This matrix has been shown on table 1. It shows the relation between patents and keywords.

The MS SQL Server – Full-Text service helped us to create "Co-word matrix" and "Matrix of Features" which will describe in the following.

Table 1. Keyword vector matrix of features

	Keyword 01	Keyword 02	Keyword 03	...
Patent 01	Rank	Rank	Rank	...
	(0 to 1000)	(0 to 1000)	(0 to 1000)	
Patent 02	Rank	Rank	Rank	...
	(0 to 1000)	(0 to 1000)	(0 to 1000)	
Patent 03	Rank	Rank	Rank	...
	(0 to 1000)	(0 to 1000)	(0 to 1000)	
...	...	...	...	...

Data analysis

After preparing the suitable data, we go through the analysis procedures. Several tasks have to be done in this step which includes patents clustering, labeling the clusters, recognizing the relations between clusters, and visualizing the results. In the next section, each step has been explained briefly.

Patent clustering and identification Areas of Technology: The gCLUTO clustering software has been used in the clustering step. gCLUTO is a graphical clustering software package which combines clustering algorithms along with a number of analysis, reporting, and visualization tools [35]. It is an easy to use platform for cluster analysis of large datasets. The previously mentioned keyword matrix served as the input of this software. In this approach each cluster can be considered as one or maybe more areas of technology.

Labeling the clusters: Results of clustering should be prepared in the way that human users can interact with them efficiently. Labeling clusters is one way that can help users inform about the contents of each clusters. In this research we chose the labels based on the contents of each cluster. The information of 10 patents that have the highest rank in each cluster is extracted. This information includes titles, abstracts, claims, description of patents plus the keywords of each cluster. After extracting this information, we asked experts to choose the suitable name for each clusters and tried to reach to consensus on each name.

Evaluating the relations between areas of technology: For evaluating the relations between different areas of technology, matrix of clustered patents and the keywords of patents of each cluster can be used. For this aim a new big matrix will be created that the rank of each cluster with keywords of other clusters will be attained and the relation between clusters will be found.

Visualizing the results: The *Mountain Visualization of g-cluto software* has been used to visualize the clusters. In this visualization, each cluster is represented as a peak in the 3D terrain. There are some points that can be informative about the mountain visualization. First, the shape of each peak is a Gaussian curve. This shape is used as a rough estimate of the distribution of the data within each cluster. Second, the height of each peak is proportional to the cluster's internal similarity. Third, the volume of a peak is proportional to the number of patents contained within the cluster. Forth, the color of a peak is proportional to the cluster's internal deviation. Red indicates low deviation whereas blue indicates high deviation. Five, only the color at the tip of a peak is significant. At all other areas, the color is determined by blending to create a smooth transition[36].

As was mentioned before clusters are representative of the different areas of technology. We represented each of these technological areas by nodes and used the multi-dimensional scaling (MDS) graph to visualize them in two-dimension space. In this way distances between the areas of technology are interpretable and represents their similarities. The Areas of Technology that are close to each other, are very similar in their geodesic distances.

Interpretation and validation of results

The outcomes and results of analysis should be refined and validated. This is the final step of the research process. An expert panel will be convened for refining and validating the results. An expert panel consist of more than 3 experts that gathering in a meeting and they discuss on results and some of their feedbacks will be used for refining the outcomes.

ALGORITHM APPLICATION: CASE OF TURBOFAN ENGINE

Data gathering

In the current research, patents related to Turbofan engine have been used. There are two reasons to select turbofan engine industry. First, many people are interested to monitor this industry which has the highest effectiveness in increase and development of air transportation. Obviously there are many business and economical motivation for this. Second, there are few studies using text mining methods in the whole aerospace industries, and we would like to investigate how this could be applied to part of this industry, which has its special characteristics and limitations, and also to show how this could be effectively implemented.

The United States Patent and Trademark Office (USPTO), UK Intellectual Property Office (UK-IPO), World Intellectual Property Organization (WIPO), Japanese Patent Office (JPO) and European Patents Office (EPO) databases served as the source for collecting patent data. Overall, 3642 Turbofan engine related patents are gathered in the period of 1900–2010 (see Table 2). The issuance volume of patents is in this order, 1725 patents from (USPTO), 814 patents from (EPO), and 535 patents from (WIPO), 423 patents from (UK-IPO) and 145 patents from (JPO).

Table 2- Collected patents office distribution

Patent Office	No. of Patents
USPTO	1725
UK-IPO	814
WIPO	535
JPO	423
EPO	145
Summation	3642

Table 3 shows time distribution of these patents. One observes that about half of these patents (46.6%) are registered in the current decade (2000-2010), about another quarter (27.9%) during last decade (1989-1999) and the remaining quarter (25.5%) during (1900-1988). This observation is not in compliance with feelings of experts in this field, who believe most of innovations belong to 1950-1970. In fact these show that we have an explosion of innovation in this field, which shows importance and effectiveness of R&D activities in this field. This can also show that jet engine industry is a high-tech industry.

Table 3- Collected patents time interval distribution

Time Interval	Percentage of No. of Patents
1900-1988	25.5%
1989-1999	27.9%
2000-2010	46.6%

Data preprocessing

Indexing documents: One of the outputs of text mining for the set of 3642 Turbofan engine-related patent documents is a list of keywords extracted from patents. Without access to text mining tools, considerable time and effort has previously been required to change the unstructured text file into a structured format. As described in the previous section to do text mining, we construct an indexed database for transformation purposes.

Development of keywords: After application of text mining algorithms on the database, keywords will be generated which subsequently should be revised by some experts of that field. Here 136 keywords remained after experts' revision.

Construction of keyword vector (matrix): With the aid of indexed database and ranking algorithms (explained in the previous section), 136 keyword vectors are constructed. The keyword-patent matrix entries show the rank of each keyword in each patent document (explained in section, the overall research process).

Data analysis

Patent clustering and Identification of Areas of Technology: Patent clustering is used to identify areas of technology. This clustering procedure is performed based on the distribution of rank of different entries. The correlation process will define different areas of technology, each of them called a cluster of patents. Note that keywords will not define borders of a cluster. In each cluster a set of special keywords have a high rank. There are two unfavorable consequences for selecting a large period of time. First, we will end up with a lot of clusters, many of which are in fact similar. Second, the emerging areas of technology will hide among areas of common technologies. For this reason we have selected three different periods of time (1900-1988, 1989-1999 and 2000-2010), and intentionally we have selected the last period, a very short one. Therefore the process of data gathering and clustering in our algorithm is an iterative process, which is supervised by experts in the field.

This process has led to identification of 18 Areas of Technology at all, 5 clusters or Areas of Technology in the first period (1900-1988), 5 Areas in the second period (1989-1999) and 8 Areas in the third period (2000-2010).

Labeling the clusters: Tables 4 to 6 demonstrate Areas of Technology in all three periods of time. We name each cluster based on contents of the first ten high rank patents in each cluster.

Table 4- Areas of Technology for the first period of time (1900-1988)

Cluster	1900-1988
0	Fan Casing and Fan Blade Loading/Unloading
1	Rotor and Shaft Assembly, Bearing Structure
2	Afterburner
3	Nacelle Assembly and Mounting Structure, Inlet Guide Vanes
4	Positioning and Control System for Fan Thrust Reverser, Cooling a Part of Engine

Table 5- Areas of Technology for second period of time (1989-1999)

Cluster	1989-1999
0	Variable Blade Pitch, Rugged, Staggered
1	Thrust Reverser (Extendible, Blockerless, Synchronizing control system)
2	Microturbomachinery
3	Noise Reduction
4	Improving Cooling Systems and Methods

Table 6- Areas of Technology for third period of time (2000-2010)

Cluster	2000-2010
0	Inlet Guide Vanes Controlling Methods
1	Variable Geometry Guide Vane System, Variable Exhaust Nozzle
2	Interstage Sealing
3	Thrust vectoring
4	Variable Torque (Low Pressure and Booster aft of Counter Rotating fans)
5	Fuel System (Fuel Control, Fuel Efficiency,...), Reduction of Exhaust Emissions
6	Thermal Barrier Coating (Methods, Materials,...)
7	Engine Controlling Systems

Relation between Areas of Technology: After identification of eighteen Areas of Technology, to find the interrelation of different clusters and areas, we use the matrix of clustered patents and keywords of patents of each cluster, and relation is set based on rank of keywords of each cluster respect to keywords of other clusters.

Visualizing the results (mountain visualization): To present these clusters in a graphical interface, G-cluto Mountain Visualization is used. Figures 2 to 4 illustrate the mountain visualization for all clusters. Mountain visualization simultaneously provides different information for us. First, the vertical shape of each peak is similar to a Gaussian curve and provides a rough estimate of the distribution of data within each cluster. Second, the height of each peak is proportional to the cluster's internal similarity (i.e. how different patents in each cluster are similar). Third, the volume of a peak is proportional to the number of patents contained within the cluster. Forth, the color distribution on a peak is proportional to the cluster's internal deviation (i.e. how patents in this cluster differ from other clusters). Red indicates low deviation, whereas blue color indicates high deviation. Note that only the color at the tip of a peak is significant. The color of other regions is determined by blending to create a smooth transition[36].

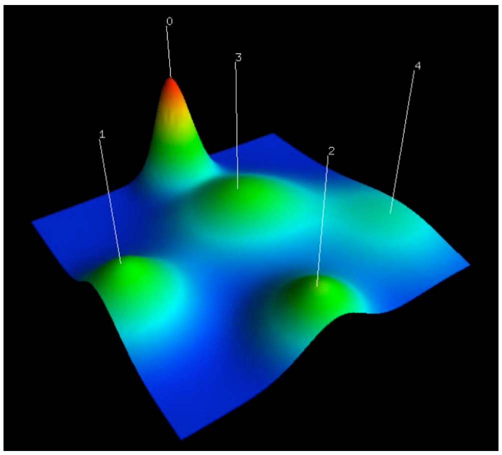

Figure 2- Mountain Visualization for Areas of Technology in the first period of time (1900-1988)

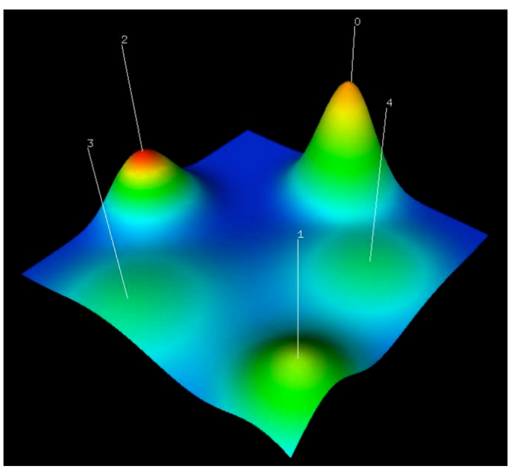

Figure 3- Mountain Visualization for Areas of Technology in the second period of time (1989-1999)

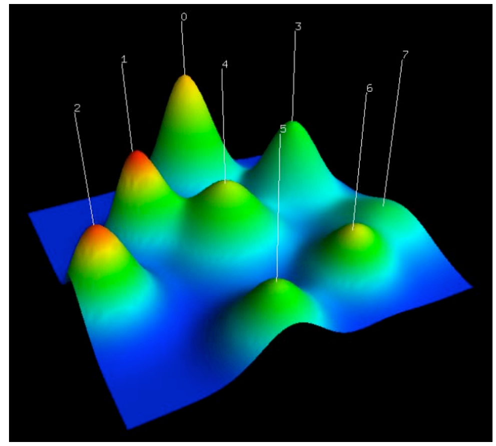

Figure 4- Mountain Visualization for Areas of Technology in the third period of time (2000-2010)

Visualizing the results (2D map): Nodes are representing clusters or Areas of Technology. The weight or size of each node has been determined based on volume of patents of each cluster or Area of Technology. This has been performed for all areas of technology. Figure 5 to 7 demonstrate the results. Multidimensional Scaling (MDS) has been used to represent distance of different nodes. The length and thickness of each line between any two nodes is representing their mutual and general similarities, which will be described in the results section.

Interpretation and Validation

Results achieved above could show us many important concepts, regarding relations of areas of technologies. We present these concepts using five different criteria.

Turbofan Technology Evolution: Comparison of areas of technology in three periods of time, is presenting the temporal *evolution of Turbofan Technologies*. As one sees in table 4, table 5 and table 6, technologies has been shown in table 4 has been focused on *manufacturing and integration of the engine*, whereas the technologies has been shown in table 5 has been focused on accessories and side sub systems *and/or scaling of the engine*. At the end, technologies have been showed in table 6 has been focused on *control and optimization of the engine*. To validate the whole process and results, a group of specialists in the field are consulted. The technology evolution is confirmed by them, and also the results are in conscience with [37]. This is also consistent with the general believe that innovations appear in creation of new products, and then on processes. For the case of Turbofan Engines, innovation in accessories, side subsystems and automation may be considered as process innovation.

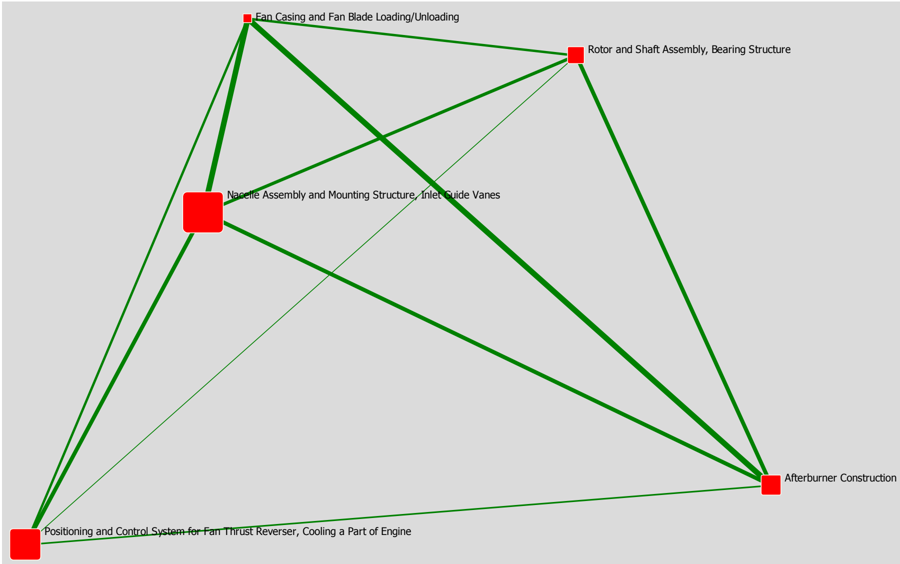

Figure 5- Relation between Areas of Technology in the first period of time (1900-1988)

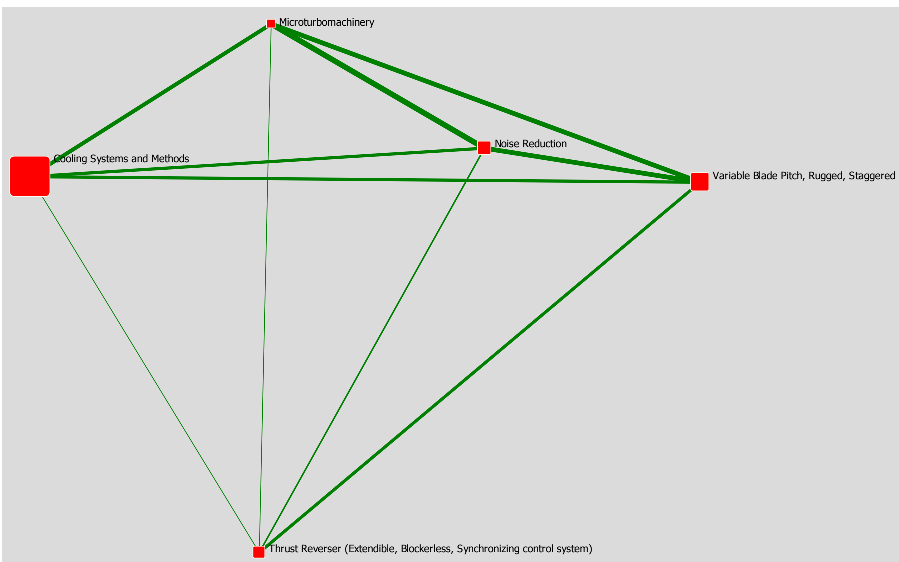

Figure 6- Relation between Areas of Technology in the second period of time (1989-1999)

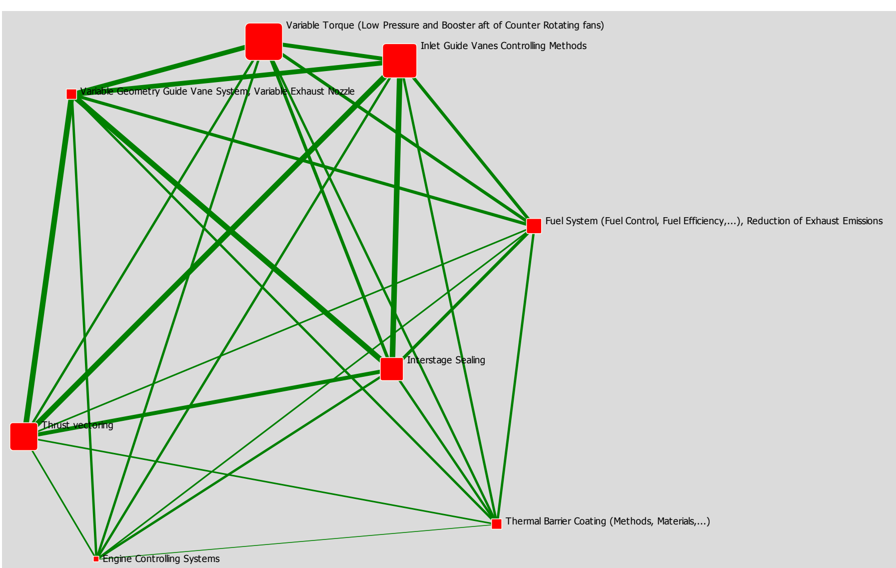

Figure 7- Relation between Areas of Technology in the third period of time (2000-2010)

Exploration of Technology Opportunities: We have already divided the period of time (1900-2010) into three sections. In each section we have some Areas of Technology. Only the last section (2000-2010) determines our Technology Opportunities. These are present evolving technologies. The areas of technologies in the corresponding map are our Technology Opportunities.

We can go further to determine priority of each Area of Technology. Priority of each area (node) is proportional to the weight of that area (node), since higher weight means higher number of related registered patents. So we attain priorities of each area of technology in each section or period of time.

Table 7- Prior Areas of Technology

Time interval	Prior Areas of Technology
1900-1988	Nacelle Assembly and Mounting Structure, Inlet Guide Vanes
	Positioning and Control System for Fan Thrust Reverser, Cooling a Part of Engine
1989-1999	Improving Cooling systems and Methods
	Variable Blade Pitch, Rugged, Staggered
2000-2010	Variable Torque (Low Pressure and Booster aft of Counter Rotating fans)
	Thrust vectoring

In the case of Turbofan Engine, as you see in Table 7, prior Areas of Technology in the first section are “Nacelle Assembly and Mounting Structure, Inlet Guide Vanes” and “Positioning and Control System for Fan Thrust Reverser, Cooling a Part of Engine”, which is in accordance with Figure 5. On the second section, according to Figure 6 prior areas of technologies are “Improving Cooling systems and Methods” and “Variable Blade Pitch, Rugged, Staggered”. On the third section, according to Figure 7 the prior areas are “Variable Torque (Low Pressure and Booster aft of Counter Rotating fans)”, “Inlet Guide Vanes Controlling Methods” and “Thrust vectoring”. The prior Areas of Technology in the third section are also the Technology Opportunities.

The relation of Areas of Technology: Thickness of lines shows amount of vicinity of Areas of Technology. One observes that it has a better meaning when assume relations as a system. The Areas of Technology that has connection to an Area of Technology. In the case of Turbofan Engine according Figure 5, “Nacelle Assembly and Mounting Structure, Inlet Guide Vanes” has strong relation by “Fan Casing and Fan Blade Loading/Unloading” and it is interesting that “Fan Casing and Fan Blade Loading/Unloading” has noticeable relation to “Afterburner Construction”; these are construction and manufacturing subject that effect on each other straightly. According Figure 6, “Microturbomachinery” has strong relation to “Noise Reduction” as well as “Microturbomachinery” has intensive relation to “Cooling Systems and Methods”, as you know when the size of engine become smaller, the cooling matters became more important and the size of engine have straight relation to noise matters. According Figure 7, “Thrust vectoring”,” Variable Geometry Guide Vane System, Variable Exhaust Nozzle” and “Inlet Guide Vanes Controlling Methods” have strong relation with each other, it is clear that thrust vectoring has inseparable relation with guide vanes especially exhaust nozzles.

Table 8- More intensity of vicinity

Time interval	More intensity of vicinity
1900-1988	Nacelle Assembly and Mounting Structure, Inlet Guide Vanes
	Fan Casing and Fan Blade Loading/Unloading
	Afterburner Construction
1989-1999	Microturbomachinery
	Noise Reduction
	Cooling Systems and Methods
2000-2010	Thrust vectoring
	Variable Geometry Guide Vane System, Variable Exhaust Nozzle
	Inlet Guide Vanes Controlling Methods

The position of Areas of Technology based on Multi-Dimensional Scaling (MDS): We used Multi-Dimensional Scaling method for developing the maps; it means that distances between the Areas of Technology are interpretable. The Areas of Technology that are close to each other are very similar in their geodesic distances. This angle of view is in contrary to last section; in this section we spotted the Areas of Technology one by one, while on the last section we assume the relations as a system for better meaning.

Table 9- Example of near Areas of Technology

Time interval	Near Areas of Technology
1900-1988	Nacelle Assembly and Mounting Structure, Inlet Guide Vanes Fan Casing and Fan Blade Loading/Unloading
	Fan Casing and Fan Blade Loading/Unloading Rotor and Shaft Assembly, Bearing Structure
1989-1999	Noise Reduction Variable Blade Pitch, Rugged, Staggered
	Microturbomachinery Improving Cooling Systems and Methods
2000-2010	Variable Geometry Guide Vane System, Variable Exhaust Nozzle Variable Torque (Low Pressure and Booster aft of Counter Rotating fans)
	Variable Torque (Low Pressure and Booster aft of Counter Rotating fans) Inlet Guide Vanes Controlling Methods

For example as you see in table 9, according to figure 5, “Nacelle Assembly and Mounting Structure, Inlet Guide Vanes”, “Fan Casing and Fan Blade Loading/Unloading” and “Rotor and Shaft Assembly, Bearing Structure” has similar geodesic distances. According figure 6, “Noise Reduction” and “Variable Blade Pitch, Rugged, Staggered” has similar geodesic distances and “Microturbomachinery” with “Improving Cooling Systems and Methods” too. And according figure 7, ” Variable Geometry Guide Vane System, Variable Exhaust Nozzle”, “Inlet Guide Vanes Controlling Methods” and “Variable Torque (Low Pressure and Booster aft of Counter Rotating fans)” are in similar geodesic distances.

Discussions

We applied a method of text mining with patent analysis along with expert supervision to attain evolution or trend of Turbofan Engine technologies. This is a fairly new activity in the field of air breathing engines technology management. For its economical and business importance, the field is considered as a classified field, and access to information is basically different from other fields like phone industry or an electric device, and for this reason, patents of this field are less transparent. Certainly this presented algorithm and results are not perfect, but it shows a promising future for technology analysis.

The suggested approach is based on two-dimensional patent mapping, which has several advantages over the basic patent statistics [38]. Firstly, it is easy to understand patent data when it has been visualized in two-dimensional space rather than has been summarized just in tables. Secondly, the overall structure of the data can be investigated in short time and the results have been shown in the patent map are easy to remember. And finally, the latent meanings in the data are easy to be explored by elimination any noise from raw data. In general, a large scale of data inevitably has been containing noises, which is needed to remove for revealing the significant meanings from the data. The technology evolution map has been enabled to distinguish only the necessary information from the unnecessary data.

Therefore, it may be possible to reduce role of experts by developing a supporting system, and yet the analysis and interpretation of expert shaving domain knowledge is indispensable not only during the selection process of keywords or labeling but also throughout the whole process of this research. In most data mining applications, data mining techniques do not eliminate the need for human input[39]. They indeed require experts to set the boundaries of the analysis (e.g.to provide query to gather patent documents of concern), a process known as feature selection (e.g. to determine a keyword list), and interpret the results of the analysis. These are particularly important when applying data mining to patent analysis, due to the complex linguistics embedded in patent files. Though large parts of patent analysis can be automated, the role of experts must be emphasized.

CONCLUSION

This article presents a new approach for development and illustration of patent analysis using text mining, to attain technology evolution of an industry. Compared to conventional patent analysis methods, the keyword-based patent analysis and illustration has considerable advantages in terms of information extraction, visualization and analysis. Operational efficiency is also enhanced, as most tasks are performed automatically. We cluster patents and explore those emerging Areas of Technology, which are not as clear as other Areas. This also helps us to find the future opportunities, i.e. the potential Areas of Technology that may push edges of technologies. The scheme also clarifies the relation of Areas in one temporal section, with previous Areas of Technology.

Above all, the focus of this research is not limited to the development of the patent map. Rather, this research emphasizes on how to analyze, interpret and utilize the patents to discover evolution of technologies, new technology opportunities and the relation and priority of Areas of Technology. We have automated most activities, which saves considerable time and effort in the process. It simplifies the procedure and permits you to enter fields of science and technologies that are codified based on a special literature, as we applied it to Turbofan Engines.

By its nature, this study is an exploratory one, and needs more extension and/or elaboration in terms of methodology and application. To complete this work, further research in the following issues is essential. First, further automation is helpful. Although we have developed an expert system and automated some elements, there is still considerable scope for operational efficiency enhancement. Second, the validity of this approach requires employing more patent documents from a wider range of technologies, which is indispensable to gain external validity. Most importantly, the outcome of the current research may generate a set of Areas of Technology that may contribute to identification or exploration of existing areas of technologies, and we hope to use this data to generate or develop the Technology Roadmap for a product or an industry. We are currently developing these ideas.

REFRENCES

1. Guan, J.C. and Y. He, 2007. "*Patent-bibliometric analysis on the Chinese science-technology linkages"*. Scientometrics, **72**(3): p. 403-425.
2. Lee, S., et al., 2009. "*Business planning based on technological capabilities: Patent analysis for technology-driven roadmapping"*. Technological Forecasting and Social Change, **76**(6): p. 769-786.
3. Verbeek, A., K. Debackere, and M. Luwel, 2003. "*Science cited in patents: A geographic "flow" analysis of bibliographic citation patterns in patents"*. Scientometrics, **58**(2): p. 241-263.
4. Moed, H.F., W. Glänzel, and U. Schmoch, 2004. "*Handbook of Quantitative Science and Technology Research.The Use of Publication and Patent Statistics in Studies of S&T Systems"*. Dordrecht: Kluwer Academic Publishers.
5. Lee, S., et al., 2008. "*Using patent information for designing new product and technology: keyword*

based technology roadmapping". R & D Management, **38**(2): p. 169-188.
6. Geisler, E., 2005. "*The measurement of scientific activity: Research directions in linking philosophy of science and metrics of science and technology outputs"*. Scientometrics, **62**(2): p. 269-284.
7. Yoon, B. and Y. Park, 2004. "*A text-mining-based patent network: Analytical tool for high-technology trend"*. Journal of High Technology Management Research, **Volume 15**: p. 37-50.
8. Ernst, H., B. Fabry, and J.H. Soll, 2004. "*Enhancing market-oriented R&D planning by integrated market and patent portfolios"*. Journal of Business Chemistry, **Vol.1**(1).
9. Ernst, H., 2003. "*Patent information for strategic technology management"*. World Patent Information, **Vol.25**: p. 233–242.
10. Yoon, B. and Y. Park, 2007. "*Developing of New Technology Forecasting Algorithm: Hybrid Approach for Morphology Analysis and Conjoint Analysis of Patent Information"*. IEEE Transactions on Engineering Management, **Volume 54**(3).
11. Noyons, E., 2001. "*Bibliometric mapping of science in a science policy context"*. Scientometrics, **Volume 50**(1): p. 83-98.
12. W.Boyack, K., R. Klavans, and K. Borner, 2005. "*Mapping the backbone of science"*. Scientometrics, **Volume 64**(3): p. 351-374.
13. Tseng, Y.H., C.J. Lin, and Y.I. Lin, 2007. "*Text mining techniques for patent analysis"*. Information Processing & Management, **43**(5): p. 1216-1247.
14. Meyer, M., 2002. "*Tracing knowledge flows in innovation systems"*. Scientometrics, **54**(2): p. 193-212.
15. Geisler, E., 2002. "*What do we know about R&D metrics in technology-driven organizations?"*. Illinois Institute of Technology: Chicago.
16. Tseng, Y.-H., et al., 2007. "*Patent surrogate extraction and evaluation in the context of patent mapping"*. Journal of Information Science, **33**(6): p. 718-736.
17. Griliches, Z., 1990. "*Patent Statistics as Economic Indicators - A Survey"*. Journal of Economic Literature, **28**(4): p. 1661-1707.
18. Jaffe, A.B., M. Trajtenberg, and P. M.Romer, 2002. "*Patents, citations, and innovations: a window on the knowledge economy"*. Cambridge: MIT PRESS,.
19. Bhattacharya, S., 2004. "*Mapping inventive activity and technological change through patent analysis: A case study of India and China"*. Scientometrics, **Volume 61**(3): p. 361-381.
20. Huang, Z., et al., 2003. "*Longitudinal patent analysis for nanoscale science and engineering: Country, institution and technology field"*. Journal of Nanoparticle Research, **Volume 5**: p. 333-363.
21. Daim, T.U., et al., 2006. "*Forecasting emerging technologies: Use of bibliometrics and patent analysis"*. Technological Forecasting & Social Change, **Volume 73**: p. 981-1012.
22. Fabry, B., et al., 2006. "*Patent portfolio analysis as a useful tool for identifying R&D and business opportunities—an empirical application in the nutrition and health industry"*. World Patent Information, **Vol.28**: p. 215–225.
23. Ernst, H., 1998. "*Patent portfolios for strategic R&D planning"*. Journal of Engineering and Technology Management, **15**(4): p. 279-308.
24. Narin,F.,1994. "*Patent bibliometrics"*. Scientometrics, **Volume 30**(1): p. 147-155.
25. Chen, C. and D. Hicks, 2004. "*Tracing knowledge diffusion"*. Scientometrics, **59**(2): p. 199-211.
26. D.Adams, J., J.R. Clemmons, and P. E.Stephan, 2006. "*How Rapidly does science leak out?"*. National Bureau Of Economic Research, **Working Paper 11997**.
27. Nomaler, Ö. and B. Verspagen, 2007. "*Knowledge Flows, Patent Citations and the Impact of Science on Technology"*. Maastricht Economic and social Research and training centre on Innovation and Technology.
28. Lee, S., B. Yoon, and Y. Park, 2009. "*An approach to discovering new technology opportunities: Keywords-based patent map approach"*. Technovation, **Volume 29**: p. 481-497.
29. Li, X., et al., 2007. "*Patent citation network in nanotechnology (1976-2004)"*. Journal of Nanoparticle Research, **9**(3): p. 337-352.
30. Kim, Y.G., J.H. Suh, and S.C. Park, 2008. "*Visualization of patent analysis for emerging technology"*. Expert Systems with Applications, **Volume 34**.
31. *Scopus Database;* Available from http://www.scopus.com.
32. *Microsoft SQL Server 2008 Enterprise*, 2008. Available from http://www.microsoft.com/downloads/details.aspx?FamilyID=265f08bc-1874-4c81-83d8-0d48dbce6297&displaylang=en
33. Microsoft SQL Server Full-Text Services; 2008. Available from: http://msdn.microsoft.com/en-us/library/cc879306.aspx.
34. Janssens, F., 2007. "*Clustering of Scientific Fields by Integrating Text Mining and Bibliometrics"*. In *Faculteit Ingenieurswetenschappen*, Katholike University Leuven: Leuven. p. 261.
35. Rasmussen, M. and G. Karypis, 2004. "*gCLUTO – An Interactive Clustering, Visualization, and Analysis System*". University of Minnesota, Department of Computer Science and Engineering: Minneapolis.
36. Rasmussen, M. and M. Newman, 2003. "*gCLUTO Documentation*". University of Minnesotas.

37. Koff, B.L., 2004. "*Gas turbine technology evolution: A designer's perspective"*. Journal of Propulsion and Power, **20**(4): p. 577-595.
38. Engelsman, E.C. and A.F.J. Vanraan, 1994. "*A Patent-Based Cartography of Technology"*. Research Policy, **23**(1): p. 1-26.
39. Piramuthu, S., 2004. "*Evaluating feature selection methods for learning in data mining applications"*. European Journal of Operational Research, **156**(2): p. 483-494.

