Technological Forecasting & Social Change 132 (2018) 105–117

Contents lists available at ScienceDirect

Technological Forecasting & Social Change

journal homepage: www.elsevier.com/locate/techfore

Exploring technology opportunities by visualizing patent information based on generative topographic mapping and link prediction[☆]

Byungun Yoon[a,*], Christopher L. Magee[b]

[a] Department of Industrial & Systems Engineering, School of Engineering, Dongguk University, Seoul, Republic of Korea
[b] Institute for Data, Systems and Society, Massachusetts Institute of Technology, Cambridge, MA, USA

ARTICLE INFO

Keywords:
Technology opportunity analysis
Visualization
Patent information
GTM
Link prediction

ABSTRACT

The shortening lifetime of technology requires companies to make intensive efforts to continuously explore new technology. Although many researchers have proposed visualization methods to find technology opportunities, little attention has been paid to present detailed directions of technology development with specified characteristics of technology. Thus, this research aims to suggest a systematic approach to conducting technology opportunity analysis by visualizing patent information, such as patent documents and citation relationships. First, keywords that explain core concepts, functions, and so on are extracted from collected patent documents by text mining. Second, patents are visualized in a two-dimensional space, and vacant cells are identified with their estimated keyword vectors by generative topographic mapping (GTM). Third, since many vacant cells will be potential candidates for developing new technologies, link prediction tools can choose promising vacant cells to connect existing cells with potential, but not yet existent, cells. Finally, the results of prediction are tested by comparing the predicted cells with the actual developed cells. The research reported in this paper is based in three technologies that have emerging, stable, and declining patterns, in order to illustrate the proposed approach, and investigate in which types it is relevant. It is found that the proposed approach provided a good prediction performance in the case of a technology that has a stable pattern. In addition, among link prediction methods, a semantic similarity-based approach showed better prediction results than a machine learning technique due to modest data availability for training. Thus, the results of this research can help R&D managers plan and evaluate R&D projects for technology development.

1. Introduction

In an uncertain economy, the survival of companies has been more challenging than in the past, due to rapidly changing customer needs and technology. In particular, it is pertinent to note the importance of technology, because since the mid-20th century, it has been considered as the key driver of innovation (Dismukes et al., 2005; Schumpeter, 1939). Moreover, disruptive innovation, in restructuring the current industries or markets, can make existing paradigms obsolete. Although many factors can influence the successful management and competition of companies, technology has been regarded as a critical trigger to competitive edges. Emerging technologies have the potential to disrupt the status quo, by changing the patterns of resource utilization, and rearranging value pools (Cheng et al., 2017). Thus, most companies invest a considerable portion of their budget in supporting the activities of technology innovation. Among the subjects of technology innovation management, the identification of technology opportunities is a starting point for managing the subsequent activities, such as technology acquisition and exploitation. New opportunities for disruptive, as well as sustaining technology should be explored to generate considerable profits for companies (Cozzens et al., 2010, 2010; Porter and Newman, 2011).

The concept of technology forecasting has become popular, as it includes systematic activities to predict and comprehend the direction, rate, characteristics, and effects of technological change (Coates et al., 2001). Thus, many researchers have been actively interested in technology opportunities, because the research theme contains academic, as well as practical value (Savioz and Blum, 2002). Since exploring technology opportunities is related to forecasting and must consider the unique characteristics of a technology, both theoretical and methodological approaches need to be applied to address the prediction problems in a substantial way. At the same time, from a practical

☆ It is confirmed that this item has neither been published nor is currently being submitted elsewhere.

* Corresponding author at: Department of Industrial & Systems Engineering, School of Engineering, Dongguk University, 3-26, Pil-dong 3ga, Chung-gu, Seoul 100-715, Republic of Korea.

E-mail address: postman3@dongguk.edu (B. Yoon).

https://doi.org/10.1016/j.techfore.2018.01.019
Received 27 July 2017; Received in revised form 25 December 2017; Accepted 18 January 2018
Available online 16 February 2018

perspective, technology intelligence that can support decision makers for technology planning by identifying risks and opportunities of the technological environment has been implemented in many global companies (Kerr et al., 2006). However, the existing literature on technology opportunity has some limitations, in terms of analytic methods, performance, and detailed specification. First, most of the existing approaches are based on qualitative processes, and dependent on expert opinion. Thus, reliable outputs are hardly provided for decision making; and moreover, their validation is very difficult, rendering the decisions for technology opportunity still uncertain. Second, while many researchers utilized visualization in order to explore vacant areas for technology development as technology opportunities, the true meaning of a technology vacancy has not been specified. In addition, the results are often weak in designing the details of a technology, because they focus on the provision of the areas of technology opportunities. Third, in even quantitative approaches, the forecasting of technology opportunities is not systematic, because most of the methods concentrate on finding rough vacant areas of technology maps or promising categories, rather than predicting details of technology opportunities. Thus, the results that are derived by existing methodologies are at best weakly connected to practical actions for R&D planning.

In order to address these limitations, this research aims to propose a systematic approach to the exploration of technology opportunities. Basically, patent information, such as patent documents and citations, is utilized to discover promising technology opportunity in a quantitative manner. Since the visualization of patent information can provide an effective locus to find technology opportunity, a process to visualize a large amount of patent information is employed in the form of patent maps. For this, text mining techniques are utilized to extract valuable keywords from unstructured texts that are recognized as a basis for visualization. In addition, and more importantly, the meaning of vacant areas in a patent map needs to be clarified, in order to support a decision making process to plan technology development. Thus, generative topology mapping (GTM) is utilized to identify the meaning of technology opportunities, by estimating their keywords in the map. Since there are many vacant areas in a map that have no existing patent, the promising vacant areas should be forecasted by applying a systematic approach. In this research, a patent network is additionally visualized on a patent map with patent citation information, and link prediction methods are then employed to forecast the future links, among pairs of cells in the map that previously had no links. Thus, this research combines the patent map and patent network as visualization methods, in order to explore vacant areas, and predict missing links in a patent map.

This paper is structured as follows: Section 2 reviews the theoretical background on the main research ideas and methodologies, including technology opportunity analysis, visualization of patent information, and link prediction. Section 3 explains the basic concept and overall process of the proposed approach for exploring technology opportunities. Section 4 presents real technologies for illustration, while Section 5 discusses the managerial and methodological implications. Finally, Section 6 reviews the limitations, and suggests future research following up on this paper.

2. Theoretical background

2.1. Technology opportunity analysis

Technological opportunity can be defined as a set of possibilities for technology advance to enhance either the functions of products, or their production (Olsson, 2005). Many researchers have conducted academic research to identify innovative opportunities to forecast emerging technologies (Savioz and Blum, 2002). Technology opportunity analysis (TOA) that explores and evaluates the risks, as well as the opportunities of technology development, is a type of technology forecasting, because it aims to predict the future scenarios of technology evolution. However, several experts must spend a lot of time to understand, evaluate and specify technology opportunities in practical R&D planning (Lee et al., 2017). Thus, qualitative and quantitative approaches can be applied to TOA, like technology forecasting. In terms of qualitative methodologies, the Delphi approach is effective to synthesize the various opinions of domain experts. In contrast, data analysis can quantitatively support the process to explore promising technology opportunities, by utilizing systematic techniques, such as text mining, data mining, and machine learning.

In terms of data sources for TOA, technological information can be classified into four categories, including patents, scientific and technical publications, people, and products and processes (Granstand, 1999). Since patents comprise a large amount of data as direct outputs of R&D activities, and are evaluated and generated to an international standard, they are often utilized to analyze technology opportunities. In particular, patent analysis has been regarded as a core analytical tool for technology opportunity analysis by employing computerized tools, such as text mining and bibliometric analysis (Lee et al., 2011). Most patent analyses utilize visual forms, such as charts, graphs and networks, to facilitate the exploration of technology opportunities. Thus, various datamining techniques have been integrated with patent analysis, adding more systematic processes to the conventional statistical approaches. Among these, text mining is intensively employed in TOA, because it can analyze the contents of documents that are very critical for understanding the characteristics of technology in technological texts. The text mining techniques are frequently combined with other techniques to investigate technological trends. For example, text mining was used with the Theory of Inventive Problem Solving (TRIZ) to analyze technological evolutionary trends (Wang et al., 2010). Elsewhere, network analysis and citation analysis were combined with text mining to visualize technological and business opportunities (Lee et al., 2009).

Although patent analysis is a representative method for TOA, the combination of scientific and technological knowledge nourishes the process of exploring new technology opportunities (Wang et al., 2015). Scientific knowledge provides the foundation for technological knowledge, and feedback stimulus from technological knowledge can catalyze continuous search of scientific knowledge (Glänzel and Meyer, 2003). It is controversial as to which type of knowledge precedes the other, because patents are often issued before scientific papers on the topic are published. In general, most people think that scientific knowledge is applied to develop patents, because research based on science is close to basic research, and technological knowledge can be obtained by applied research. Since the two types of knowledge are complementary, the data on scientific discovery can improve technological development and commercialization (Hellmann, 2007). Although many researchers have used patent data to concentrate on the utilization of technological knowledge, some studies have tried to integrate the two main data sources, of academic papers and patents, in order to explore technology opportunities. Shibata et al. (2010) applied social network analysis to develop citation networks of science and technology based on citation information in the references of papers, as well as patents. Technology opportunities can then be identified by determining the gaps between clusters of patents and papers.

2.2. Visualization of patent information

Information visualization can help analysts gain an overview of objects in a two-dimensional or three-dimensional space (Kwakkel et al., 2014). The main advantage of information visualization is to extract a critical insight from a large amount of information by reducing the dimensions of data. In the recent patent analysis, various visualization tools, such as maps and networks, have been utilized to investigate patent information. Since the patent database has a very large number of patent documents, as well as potential patent infringement or overlap, visualization tools have been actively employed to analyze the characteristics of patents. For example, patent maps and patent networks can be developed to examine the technology trends in a specific technology domain, or the similarities among technology domains (Chen, 2009).

The main data sources for visualization in patent analysis are twofold, and include bibliometric information and patent language content. The bibliometric information refers to the application year, applicants, cited/citing patents, and other mostly structured data that have quantitative values in a specific data field. Chang et al. (2009) proposed a framework to identify the technological trends by conducting bibliometric patent analysis based on network analysis. In addition, Tang et al. (2012) suggested an inventor–company–topic (ICT) model that integrates information of inventors/companies and topics of technology, mining information from heterogeneous patent networks. In contrast, patent documents are unstructured data that consist of natural language, and thus require an additional process to deduce core features of the patents based on text mining. Text mining approaches using natural language processing can be classified into two groups: (a) subject-action-object (SAO)-based approaches, and (b) keyword-based approaches. While the keyword-based approach just utilizes the extracted keywords to investigate the characteristics of documents, the former focuses on the formalized structure of sentences. In particular, patents should have a required format for making a patent document, and the main objective of patents is clearly to describe and claim a new idea or approach to solve a technological problem. Therefore, the SAO information will be very critical for technology opportunity analysis. Thus, many researchers have applied text mining approaches to visualize patent information. Yoon et al. (2013) introduced a method to dynamically construct patent maps by analyzing the SAO-based contents to identify the technological competition trends. They extract the multiple SAO structures from patent documents, and generate patent maps. Choi et al. (2012) proposed an SAO-based approach for text mining that develops a technology tree by mining and examining patent information. NLP and text mining techniques are employed to extract the SAO structures, and a similarity matrix is constructed by calculating the similarities among the SAO structures to draw a technology tree. In particular, Lee et al. (2015, 2015) proposed a novelty-focused patent mapping for technology opportunity analysis by applying text mining to extract the patterns of word usage and outlier factor. Recently, advanced text mining approaches such as word2vec and latent Dirichlet allocation (LDA) have been applied to improve the information extraction from documents. Word2vec has been very popular because it can represent words that are semantically close based on the vectors of relationships between words and their context (Enríquez et al., 2016). In addition, LDA is a key tool to discover latent semantic structure within a variety of document collections as a probabilistic model (Blei et al., 2003). Therefore, text mining techniques are often combined with other visualization techniques, such as neural network, multidimensional scaling, and principal component analysis.

2.3. Link prediction

Many social, information and technological systems can be explained in terms of a network that consists of nodes and links. Enormous research efforts on such networks have continued to analyze the features of networks (Costa et al., 2007), the dynamic evolution of networks (Albert and Barabási, 2002), and relations between functions and topologies (Boccaletti et al., 2006; Newman, 2003). In particular, an important research issue in network analysis is the prediction of relationships between nodes, estimating the likelihood of the existence of links. The seminal work of Liben-Nowell and Kleinberg (2003) proposed the earliest model for link prediction, and provided some classical prediction measures based on topological information. Since the data in generating biological or social networks has incomplete or inaccurate information, the network might contain spurious links. The link prediction algorithms can identify the spurious links, and predict links that may appear in future networks. For example, in the social network services, highly likely links that do not yet exist can be recommended as potential friends, to help users find new friends. Leskovec et al. (2010) applied a logistic regression model to predict links in signed social networks, connecting the model to balance and stabalize theory. In addition, Hansen et al. (2006) showed that the link prediction problem could be dealt with by a concept of classification, and used to predict co-authorship in BIOBASE. In a technological system, the patterns of technology convergence were predicted by jointly utilizing the link prediction and association rule (Lee et al., 2015, 2015). In terms of linking product and technology areas, Kim et al. (2017) suggested an approach to identifying potential areas for concentric diversification at a product level. Since link prediction is a long-standing challenge in information science, various algorithms, such as statistical models and machine learning, have been suggested to predict accurate potential relationships among the nodes in networks. However, most of the approaches do not provide important insights, and fail to consider the structural characteristics of networks.

Many link prediction approaches have been proposed to forecast potential links in networks but it is important that most testing of such models has been performed on social networks rather than on networks where the relationhip is technical. If the link prediction problem is regarded as a supervised classification, popular tools, such as naïve Bayes, neural networks, and support vector machines (SVM), can be used to predict future links in test data by training the learning model. Since the performance of various classification tools can be compared, some tools work better than others for a specific data set or domain. Hasan et al. (2006) found that bagging and SVM have the marginal competitive edge for a co-authorship network- a social network. A second general method that has been used is the graph topological feature-based approaches utilizing natural features, such as the similarity based on node neighborhoods, or paths between a pair of nodes for link prediction. Node neighborhood-based features include common neighbors, the Jaccard coefficient, and Adamic/Adar. As the number of common neighbors increases, the probability that two nodes will have a link, increases. Newman (2001) showed that in collaboration networks, the number of common neighbors and the probability that two nodes will collaborate in the future have a positive correlation. In terms of path feature-based methods, shortest path distance and Katz are the most frequently utilized approaches. In particular, the underlying concept of the shortest path distance is that the path distance between nodes in a network can affect the formation of links. Two nodes that have shorter path distance have higher probability that they will be connected in the future. A third approach is the Bayesian probability model which can be employed for link prediction by deriving *a posteriori* probability for co-occurrence of vertex pairs in a network. The probability is utilized as a feature in the classification of the potential existence of links. Wang et al. (2007) used the output from the probability model as a feature in the subsequent step that applies other features, such as Katz and vertex similarity, to predict a binary value for the classification of links.

3. The proposed approach

3.1. Research concept

This paper aims to propose a new approach to exploring promising opportunities for new technology development. Basically, this approach utilizes a visualization technique to facilitate finding vacant areas that existing technologies have not dealt with. A visual form that can show the features of technology or the relations between technologies has been regarded as a key tool that enables the investigation of current technology, and the anticipation of the future direction of technology

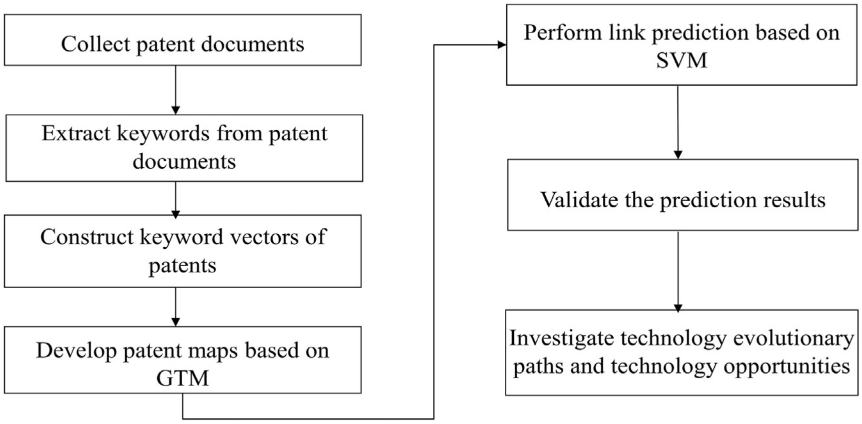


Fig. 1. Overall research process.

development. Thus, patent maps or patent networks are drawn to visualize the information relationships of the technology of interest. However, the two typical types of technology visualization are separately developed, because they have different objectives, so to speak – locations, and relations. This paper combines patent maps and patent networks, in order to show the relations between patents in a patent map that is drawn by using the similarity between patents. Although a patent may be similar to another patent, and they are closely positioned in a patent map, they may have no technological relation. While the similarity can be measured by the keywords that each patent has in the patent document, the technological relation can be analyzed by the citation information - a new patent cites previous patents- capturing a second knowledge relationhip beyond keyword similarity. Thus, in this paper, both characteristics of patents, and knowledge flows between patents, are considered in a forecast of new technology opportunities.

Although the technology opportunity can be incremental or radical, this paper focuses on incremental innovation for technology development. Since radical (sometimes similar events are termed disruptive) technology often involve change of complex environments (Gans, 2016), the link prediction technique in specific technological areas- our approach- is inappropriate for such kinds of technological change. Therefore, the proposed approach deals with incremental innovation, by using the information on historical patent applications. The landscape that is created by collected existing patents has lower relevance to forecasting radical innovation, because the identified patent vacancy is strongly related to the patents around it. More importantly, radical innovation generally creates a new paradigm for technology development and tends to fail to meet the needs of current markets. Thus, this research, utilizing patent information without other information on the change of markets or industries, concentrates on technology forecasting for incremental innovation. In addition, patent vacancy in a patent map can be considered a technology opportunity, whereas already occupied areas in the map cannot provide promising opportunities. Although a new technology can be devised in areas of existing patents, technology opportunity can be defined as a new promising area among patent vacancies in a patent map. Thus, technology opportunity in this paper refers to new technology by incremental innovation, rather than simple derivatives of previous patents.

This research applies both retrospective and prospective approaches to technology evolution. First, the changing patterns of past patents are analyzed to understand the trends of technology development. In a patent map, the sequence of patent applications is drawn to present the technology path. Second, in terms of prospective approaches, the future evolutionary path is investigated to forecast promising technology opportunity. Link prediction methods are employed to anticipate possible links between existing patents and potential vacancies. A patent vacancy that is linked to recent patents has a high potential for technology development, because its links with recent patents mean the reliability and practicality of the opportunity.

3.2. Framework

The overall process of this approach has the following steps: First, patents of interest are collected from patent databases, and generate structured data that have critical information on titles, abstracts, citations, application number, application year, and so on. Second, keywords are extracted from patent documents by a text mining technique. Due to the unstructured format of texts, documents cannot be visualized *per se*, important keywords that can represent the whole contents of documents are extracted, and utilized for visualization. Third, to visualize the characteristics of patents, keyword vectors of patent documents are constructed. The occurrence frequency of keywords that is the number of times a keyword appears in a patent is filled in a data field of keyword vectors. Fourth, a patent map is developed by utilizing GTM and keyword vectors of patents. Since the GTM generates a map based on cells that divide a whole map into many parts, it is able to show both dense areas and vacant areas in a patent map. The vacant areas can be regarded as new technology opportunities, because there has been no attempt to tackle the technology opportunity. Fifth, to derive promising technology opportunities, link prediction between existing patents and vacant cells is performed. Several methods for link prediction, such as semantic similarity and SVM, are employed and compared, to obtain an accurate forecasting approach. Consequently, the anticipated cells can be derived as promising opportunities, and the characteristics of the cells are analyzed by the GTM, presenting the estimated keyword vectors. Sixth, the prediction results are tested by calculating the forecasting errors, and comparing the results of the proposed approach with those of other approaches. Finally, the dynamic patterns of technology development and potential future technology are investigated to help build a technology strategy. Fig. 1 shows a flow chart of the overall process.

3.3. Process

3.3.1. Data collection

Since the proposed approach explores technology opportunities with collected patents in specific technological areas, relevant data must be retrieved to derive accurate results for technology forecasting. First, this paper selects the USPTO database among the various data sources of many countries, such as Europe, Korea, and Japan. The USPTO database has many patents, and the rigorous process of patent citation can provide valuable information to the proposed approach. Although all patents from various worldwide patent databases can be collected and analyzed, each database has unique characteristics in terms of main patent classifications and diversity of applicants, and the integration of patent databases might make the results of patent analysis unclear. The selection of a single patent database enables the clear exploration of new technology opportunities, investigating the dynamic patterns of technology development. Kim and Lee (2015) concluded

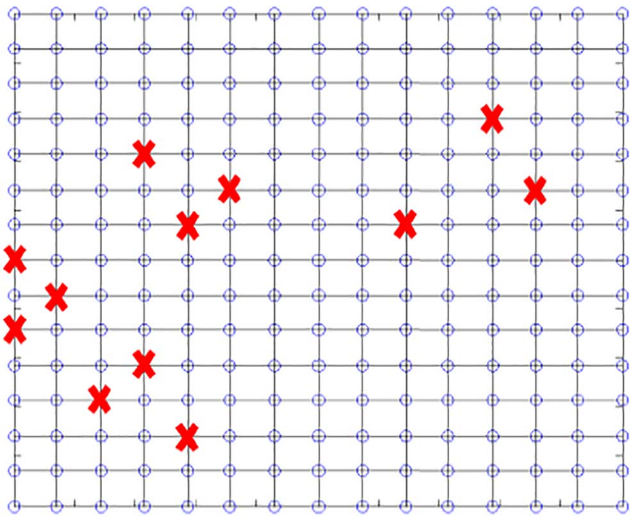


Fig. 2. Vacant cells in a patent map.

that the USPTO database has a large number of patents, the higher growth rate of patent application, broad coverage in patent classification and more diverse nationality of applicants in comparison with other patent databases. Second, in terms of a retrieval method, this research basically combines the keyword search and classification constraint, because a patent that is retrieved by specific keywords might be included in a totally different technology classification. The searching process proposed by Benson and Magee (2013, 2015) is employed to collect appropriate patents of interest. The first step is to pre-search using a set of keywords to find the most representative patent classes in both US and international patent systems. The pre-search is completed by searching for the two-word query in the title or abstract of patents. In the second step, the set of patents that are collected from the pre-search is used to determine the US patent classes (UPC) and international patent classes (IPC) that are representative of the specific technology. Finally, the patents that are contained within both UPC and IPC that are identified in the previous step are retrieved to find the final set. If a patent is included in the most representative patent class in both patent systems, the dual membership can guarantee that it has higher relevance (Criscuolo, 2006). Thus in this way, patents of interest are collected for the TOA.

3.3.2. Construction of keyword vectors

Since the patent documents are written in natural language, they can be structured to quantitatively analyze the relations between patents. The most efficient method to perform the structuring of patent documents is to extract meaningful keywords by text mining techniques, and establish the keyword vectors of patents. In general, the data fields of keyword vectors have specific values, such as the frequency with which a keyword appears in a patent document. Although a binary value (0 or 1) that just considers the existence of a word can be utilized to build the keyword vector, the occurrence frequency can play a role similar to a weighting parameter. In this paper, the occurrence frequency is applied to analyze the relations between patents as singular use of a word may not serve to signal content differences but potentially only writing style. In terms of screening general or meaningless words from the set of extracted words, several steps are employed to generate the list of keywords. First, stop words that do not provide meaningful information, *i.e.* articles and prepositions, are removed from the documents. Since such words make a document heavier and less focused for analysis, removing them can reduce the dimensionality of document analysis. Second, a stemming method is applied to identify the stem of a word, removing various suffixes. This process aims to reduce the number of words, enabling time and memory space to be saved. Third, term frequency-inverse document frequency (TF-IDF) that reveals how important a word is to a document in a collection of documents is utilized, to exclude common words that appear in many documents. Since the index is often used as a weighting factor in information retrieval, specific words that can significantly explain the characteristics of documents are chosen by calculating the value of the TF-IDF. Consequently, the keyword vectors of each patent document are constructed by excluding stop words and common words, and using stemming words.

3.3.3. Development of a patent map

Patent maps that visualize the position of patents in a two-dimensional space by analyzing the keyword vectors of patent documents are drawn to find vacant areas, as well as dense areas. GTM is employed to develop a patent map, because it enables the visualization of data, as well as the identification of characteristics of vacant areas. The inverse function of GTM allows analysts to find the estimated keyword vectors of vacant cells in a map. GTM is the most relevant method in this research, because the proposed approach should visualize the collected patents, and then anticipate technology opportunities by identifying patent vacancies.

In the GTM-based patent map, each patent is mapped to a rectangular planar surface, in order to explore vacant cells and occupied cells. Two types of symbol ("o" and "x") in the map represent whether a cell is vacant or occupied, as shown in Fig. 2, where x cells indicate an area of undeveloped technology. The keyword vectors of patent documents are located on a cell of the two-dimensional map by GTM with several parameters, such as the number of latent points, and the number of basis functions. The reverse keyword vectors of the vacant cells are constructed using Eq. (1). In the equation, W is an initial weighting matrix, and φ(x) is the activation of basis functions when fed a latent variable sample x. The function y(x, W) is a transformation function, which maps points x in the latent space into corresponding points y(x; W) in the data space (Bishop et al., 1998). Thus, we can derive a list of keywords that can provide the characteristics of each vacant cell.

$$y(x; \mathrm{W}) = \mathrm{W}\varphi(x) \tag{1}$$

3.3.4. Technology opportunity analysis by link prediction

Not all vacant cells represent viable opportunities for technology development, because some of them might be unlinked due to low usefulness. Thus, identified vacant cells in a patent map should be evaluated to discover promising technology opportunities, separating out useless options. Although various methods can be employed to perform the evaluation of technology opportunities, this research utilizes the link prediction approach, in order to explore promising vacant cells by linking recent patents with vacant cells. From the perspective of sustaining innovation, if a vacant cell is linked with recent patents that were lately filed in patent offices, it can have a higher probability of being developed for a new technology opportunity. In this paper, the links between recent patents and vacant cells is interpreted as a path of technology development, because past patents are connected with the next opportunities for technology.

For the link prediction, this research utilizes two data sources; citation information, and keyword vectors. First, since the citation information can signal both the quality of patents and the knowledge trajectory includes the link, it is relevant information to analyze previous links and predict potential links. Second, keywords are generally utilized in performing the content analysis of documents, and keyword vectors can be critical data, when the relations between documents are examined from content analysis. In terms of methods for link prediction, this research applies two methods: SVM, and the semantic similarity approach. As mentioned before, there are several methods for link prediction, such as the node neighborhood-based, path-based, and classification approaches. However, not all methods can be employed for the proposed approach, because in the context of this paper, potential links between existing cells and vacant cells should be predicted.

Since neighborhood-based and path-based methods are applied to predict possible links between existing nodes, a node should have at least one link with other nodes. However, in this paper, the vacant cells have no links with existing cells. Thus, this research uses a classification approach, in which potential links between unseen data and existing data can be predicted. If a pair of vacant cell and existing cell has a strong relation by the classification analysis, they are classified as “linked”, connected in the patent network. By a link prediction process based on classification, we can derive predicted links among patents that were not connected in the previous patent network, using SVM with keyword vectors. Since the SVM conducts a learning process with past data (citation information and keyword vectors), it can find latent relationships between patent vacant cells and existing patents. SVM is a representative method for supervised learning that infers a function from labeled training data. In this research, the training data have independent variables (the differences between keyword frequencies of two patents with keyword vectors), and a dependent variable (binary value of “connected” and “unconnected”). The classification of new pairs of vacant cells and existing patents are then predicted by using the results from learning and estimated keyword vectors of vacant cells through GTM. Fig. 3 shows the logical structure of SVM for link prediction used in this paper.

The second method for link prediction is the semantic similarity approach that is newly proposed in this paper. Since identified cells and patents have keyword vectors, the semantic similarity can be a strong signal in investigating the relation between existing patents and vacant cells. For this, the semantic similarity between recent patents and vacant cells is calculated by Euclidean distance. Then, the closest vacant cell of which semantic similarity with a recent patent is the highest is selected as a potential link. Consequently, a small number of vacant cells can be identified as potential technology opportunities.

3.3.5. Comparison and testing

This study intends to suggest relevant types of technology, a link prediction method, and keywords that the proposed approach can be successfully applied to or with. Thus first, keywords that can predict future technology are selected, by comparing the results of all keywords and recently increasing keywords. In addition, the prediction performance of keywords is compared by considering the TF-IDF index, investigating whether keywords of high TF-IDF index are effective or not. Second, in terms of link prediction, two types of methods (the semantic similarity approach and SVM) are compared, to provide more appropriate methods for predicting new technology opportunities. For the comparison, this research utilizes three indices (recall, precision, and lift) that are usually applied in the evaluation of prediction. While precision can be defined as the proportion of positive cases that were correctly identified, recall is the proportion of actual positive cases that are correctly identified. In contrast, the lift can be understood as a ratio of two percentages: the percentage of correct positive classifications made by the model to the percentage of actual positive classifications in the test data. Thus, the lift is a measure that compares the relative performance of a classifier and another control group (usually by random selection). Using these three indices, the two link prediction methods are compared to determine their effectiveness in predicting new technology opportunities. Third, three types of technology domain that include increasing, stable, and decreasing patterns of patent application are compared in terms of their prediction performance. The newly proposed approach might have better performance within specific types of technology, due to the characteristics and constraints of methodology. Thus, this paper selected three technology domains by analyzing the trends of patent applications. While the domain of the 3D printer has an increasing pattern, nuclear fusion is included in the decreasing pattern type. In addition, patent applications in the water purification domain have been stable over the past 10 years. Therefore, we identify which technology domains are relevant to the application of the proposed approach by comparing the prediction accuracy. Chi-square analysis is employed to statistically test the prediction performance in the technology domains.

3.3.6. Dynamic analysis of technology development

The location of the technology in the life cycle is identified by investigating the dynamic patterns of patent application, enabling the selection of more promising opportunities (patent vacancy). This research splits all data into three periods, and then investigates potential vacant cells, and how the vacant cells are filled by new patents. For example, the whole time period from 2006 to 2014 is broken down into three periods (2006–2008, 2009–2011, and 2012–2014). In general, while development projects aim to generate significant profits within three years, research projects target a long-term profit more than five years later. Aboody and Lev (2001) insist that most of the operating income benefits are generated in 3 years from the R&D investment. In addition, the National Research Council classified three types of R&D projects: short-term (3 years), mid-term (3–10 years), long-term (plus 10 years) in terms of expected time to the first commercial application (National Research Council, 1986). Thus, the 3 year time period we use appears an appropriate choice to investigate the dynamic trends of technology development and the accumulation of short-term development projects can create R&D trends. Dynamic analysis in patent maps considers two perspectives of technology development. First, a cell that is always occupied in all three time periods represents a technology area where technology has been steadily developing for a long time. Emerging cells and declining cells can be included in this type of cell,

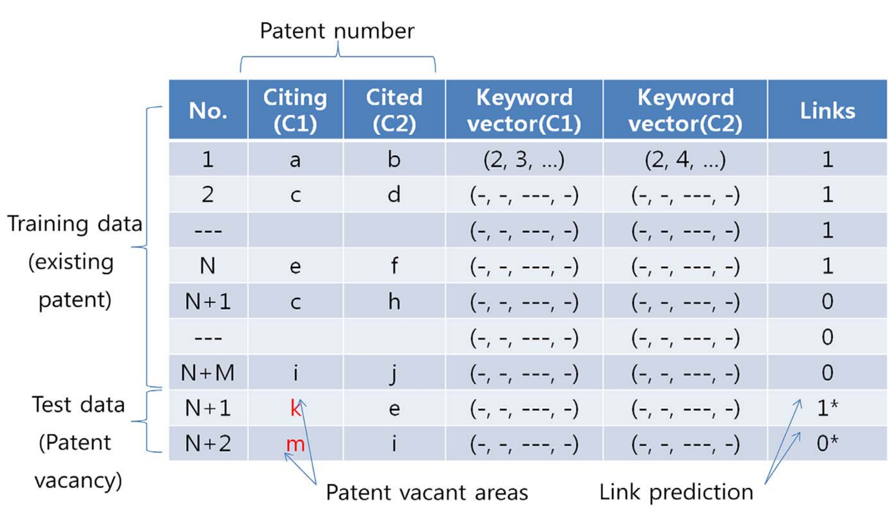


Fig. 3. Logic of link prediction through SVM.

according to the growth rate of patent applications. While the number of patent applications in emerging cells increases over time, declining cells have a decreasing number of patents. From the analysis, we can identify main patent areas that have been intensively developed, separating emerging and declining areas of technology development. Second, newly occupied cells refer to technology areas where a new patent that did not appear in the previous time period is filed. Trend analysis on which new patents are developed can provide a clear view of the changing pattern of technology development. The information on two types of technology trends from the dynamic analysis can give an important insight on technology strategy. If a company considers a technology portfolio, various types such as emerging, declining and newly occupied cells can be utilized to provide the information on the life cycle of main technologies. A company can plan a technology strategy by constructing a technology portfolio, integrating the information with technology capability and business importance.

4. Application of the approach to three technoligies

4.1. Data collection

This paper selects three technology domains to compare the applicability of the proposed approach. First, considerable attention has recently been paid to t 3D printing, because the technology and products are critical to what is often viewed as the next generation industrial revolution. Since the 3D printer enables making a prototype of a new product in a customer's home, the production orientation from factories might rapidly transform into self-production. Thus, the number of patent applications has increased for the past 5 years, indicating that this technology can be classified into a group of increasing patterns. Second, water purification is involved in a group of steady patent applications. When the number of patent applications is plotted over time, this technology domain shows little fluctuation in the graph. In addition, since the technology has already passed the growth in the technology life cycle, it cannot show an explosive increase in the number of patent applications. However, significant efforts have been made to develop various techniques, approaches, materials and so on. Technology development in water purification is still promising, due to the importance of clean water to human health and economic growth, and the serious scarcity of water in many developing countries. Thus, the activities in developing new water purification technology have been stable. Third, the technology domain of nuclear fusion is chosen as a representative technology of decreasing patterns, because its graph of patent applications has constantly decreased over time. Nuclear fusion has long been considered as a nearly limitless source of energy that is safe, clean, and self-sustaining. Even though many scientists are working to perfect such energy sources, the status of technology development still remains in a stage of unclear solutions for the practical application of the technology. There is no clear dominant design in nuclear fusion technology, continuously looking at a realistic configuration for a gap between theoretical soundness and applicability. Thus, the efforts of technology development have been declining, meaning that the number of patent applications in this field is gradually reducing.

In order to compare the analysis results of the three technologies, patents in the three technology domains were collected by constraining the time range from 2006 to 2014. Although there are many patent databases in many countries, this paper selected USPTO patent databases, because USPTO has the biggest database of patents, and many companies tend to apply their inventions to the database for competition. The data retrieval process suggested by Benson and Magee (2013,2015)) was employed to collect relevant patents without data noise. Consequently, the numbers of patents for the 3D printer, water purification, and nuclear fusion are 251, 745, and 690 respectively. Fig. 4 shows the changing patterns of the three domains, confirming that each technology is included in a different pattern of the three.

4.2. Data pre-processing

In order to visualize the collected patents in a two-dimensional space and explore promising cells in the map, the keyword vector of each patent is built from the patent documents by utilizing text mining. For this, keywords are first extracted, excluding stop-words, and deriving the root of words. After obtaining a whole set of keywords, general keywords that frequently appear in most patent documents, and have little ability to differentiate the characteristics of patent documents, are removed from the list of keywords for visualization. The final numbers of keywords that are used to construct the keyword vectors are 198, 245, and 182 for the 3D printer, water purification, and nuclear fusion, respectively. With the set of keywords, the keyword vector of each patent is generated by filling the occurrence frequency of keywords in the field of the vector. For example, in the case of the nuclear fusion, 690 patents have their own keyword vectors that consist of 182 data fields. Thus, Table 1 shows a small portion of the 690 by 182 matrix that is constructed in the database for visualization and demonstrates that the vectors have significant differences.

4.3. Visualization

In exploring new opportunities for technology development, the first step is to draw the patent map that can provide a locus of patent vacancy. The preprocessed data is utilized to visualize the patent information of the collected patents. GTM generates a two-dimensional map by analyzing the relationships between the patent data. Although several parameters should be set with a value, the size of map is very important to investigate vacant cells on a map. Since there is no rule of thumb in deciding the map size, we performed a sensitivity analysis that makes analysts evaluate the relevance of selected parameters, by changing the value of parameters. We fixed on three options of $10 * 10$, $12 * 12$, and $15 * 15$ size. As a result, the $12 * 12$ map is proper to visualize the medium size of patents (200–1000 patents), because the $10 * 10$ map and $15 * 15$ map have few vacant cells or many vacant cells, providing an insignificant landscape for predicting technology opportunities. We chose three cases of technology to compare the feasibility of the proposed approach. Thus, we drew each patent map of the 3D printer, water purification, and nuclear fusion, using GTM and patent data. Fig. 5 shows the final maps of the three technology domains by using all data covering all time periods from 2006 to 2014.

However, since dynamic analysis can present a detailed investigation of the predicted promising technology, the whole set of data is split into three subsets that respectively cover three years. Fig. 6 presents the evolutionary maps of the water purification case. The first map visualizes the technological information of 407 patents that are issued until 2008, and selected among all 524 patents. In the map, grey-colored nodes represent patents that are recently issued from 2006 to 2008. Recent patents are indicated in the same way in the subsequent maps (maps that are patents from 2009 to 2011 and from 2012 to 2014 are added to the data set of the previous period).

4.4. Link prediction

Since a variety of link prediction methodologies can be applied to forecast the potential relationship among nodes, we need to determine a proper methodology for predicting future promising patents in a patent map. Although many different techniques can be considered to offer the best methodology, two popular techniques are selected, and their prediction performances are compared. First, the distance-based approach is simple and straightforward to utilize, because it enables the exploration of nodes that have not been connected, despite strong relationship, using the distance (or similarity) among nodes. In this paper, the distances among nodes in a patent map can be calculated by analyzing the similarity between keyword vectors of nodes. Second, SVM, a machine learning approach that has shown good performance in link

Fig. 4. Trends of patent application in three technological domains.

Table 1
Keyword matrix of nuclear fusion.

Patents	Keyword vectors							
	Accelerator	Acid	Air	Alkali	Alloy	Aluminum	Ammonia	…
US4017163	0	0	0	0	0	0	0	…
US4047068	2	0	0	0	0	0	0	…
US4046630	0	0	0	0	0	0	0	…
US4065351	0	0	0	0	0	0	0	…
US4069457	10	0	0	0	0	0	0	…
…	…	…	…	…	…	…	…	…

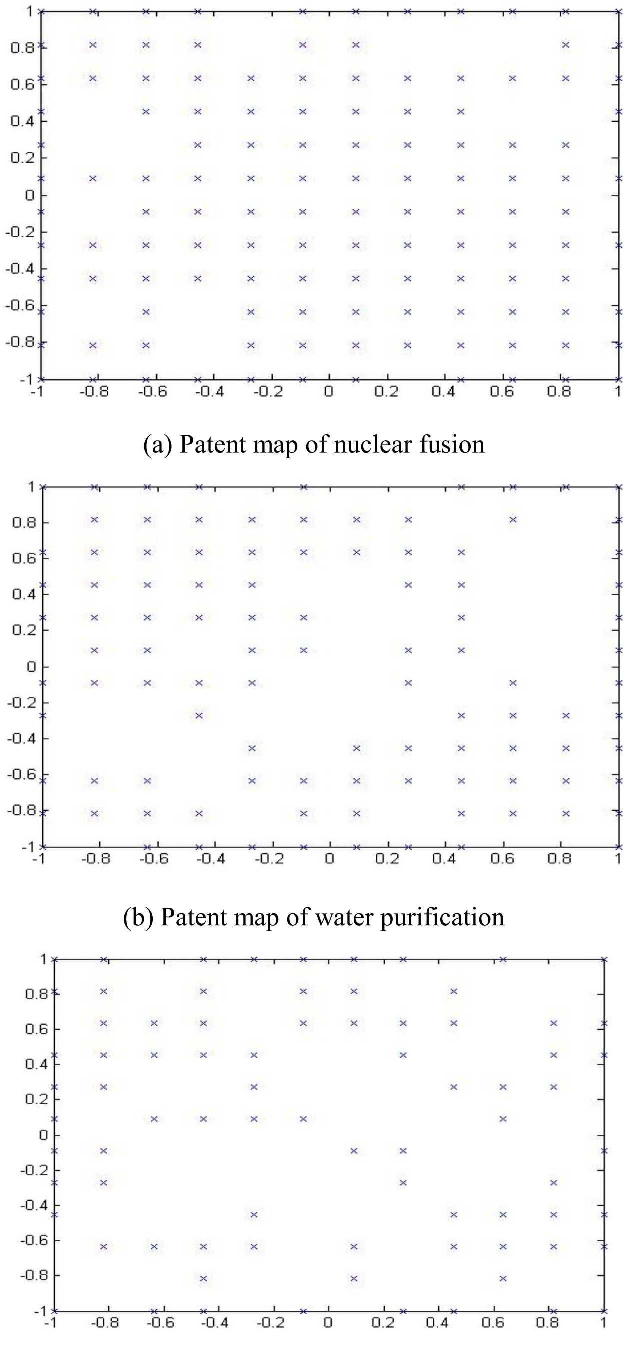
prediction, is chosen as a candidate for one of the proper methodologies. Since our database has similarity between the keyword vectors of patents and the existence of links, the relationship between the similarity and link can be established through the machine learning process.

In order to compare the prediction power of the two link prediction approaches, three indices (precision, recall, and lift) are considered. First, precision is the fraction of predicted cases that are correctly predicted. In link prediction, the index can be calculated, dividing the number of correct results by the number of all predicted cases. Second, recall is the fraction of correctly predicted cases to actual cases that should be correctly predicted. Since this research aims to predict future technology in a patent map, it is given by the number of correct results divided by the number of actual technology developments. The third index is the lift, which means the performance of a targeting model for predicting cases measured against a random choice targeting model. Lift is simply the ratio of target response to average response. If the value of this index is greater than 1, this means that the prediction is significantly greater than random choice. In addition, in order to present the statistical validity of link prediction, we performed chi-square analysis on the hit ratio of prediction.

In this research, the water purification case was chosen to compare the prediction performance of the two link prediction methods. Table 2 shows the results of prediction accuracy based on the aforementioned indices. The precision, recall, and lift of the semantic similarity-based approach are higher than those of the SVM. In the prediction using semantic similarity, 13 vacant cells in the patent map of water purification were predicted as potential cells where new technology can be developed, analyzing 407 patents until 2008. As a result, 5 cells among the 14 predicted cells were realized in the first period from 2009 to 2011, whereas 4 cells that newly appeared could not be predicted by this approach. In addition, the result in the second period from 2012 to 2014 shows that 2 predicted cells out of 4 newly occupied cells are correct. In the patent map of water purification, 57 vacant cells were identified, and 7 cells out of 13 newly occupied cells came from the 14 predicted cells. Thus, the values of precision, recall, and lift are 0.5, 0.53, and 2.19 respectively. In contrast, SVM predicted 18 cells, among which 6 cells were correct, missing 8 cells out of 14 newly emerging cells. Thus, the precision, recall, and lift of SVM become 1.46, 0.33, and 0.46. Consequently, since all the indices of the semantic similarity-based approach are better than those of SVM, we utilize semantic similarity to compare the prediction results of the three technology domains.

4.5. Prediction evaluation according to the types of technology

The three domains represent the main types of technology, because water purification technology shows a stable pattern in the curve of patent applications, and the 3D printer and nuclear fusion are in the increasing and decreasing phase, respectively. The comparison among types of technology can determine which technology will be proper for the application of the proposed approach. The first analysis for this is performed to use the aforementioned indices. Table 3 presents the values of the three indices in the three technology domains. The prediction performance of water purification was highest among the three technologies, because its values of precision, recall, and lift were conspicuously higher than those of the other technology domains. The gap between water purification and the other technology domains needs to be statistically validated. This paper applies the chi-square test to analyze the significance of the proposed approach compared to random selection. Although its prediction performance was significantly better than random selection in the water purification field, the results in the other domains were not significant. Thus, the proposed approach can



(a) Patent map of nuclear fusion

(b) Patent map of water purification

(c) Patent map of 3D printer

Fig. 5. Final patent maps of three technologies.

show high performance in predicting technology opportunities but only did so in our tests for the water purification domain.

4.6. Forecasting technology opportunities

From the results of visualization, 14 cells were predicted as new technology opportunities in the case of water purification as shown in Fig. 7. Black-colored nodes in the first map were forecasted as new technology opportunities by link prediction. Among these, 7 cells (dotted, grey-colored nodes in the second and third maps) were correctly predicted, which means that new patents were filed in the subsequent time period. In contrast, 7 cells among the predicted cells were not realized, and new patents in 6 cells (solid, grey-colored nodes in the second and third maps) that were not predicted by the proposed approach were developed. In Fig. 7, white-colored nodes mean that a patent or multiple patents have been developed again in the cells that were previously occupied by existing patents. Table 4 presents the list of patents that were correctly or incorrectly predicted, adding patents that were not predicted. New cells, such as Nos. 53, 94, 96, 97, 109, 119, and 141 are correctly predicted, because new patents filled the vacant areas. The cell numbers are sequentially given from the top, left-hand side to the bottom, right-hand side. For example, the coordinate of No. 53 cell is the fifth line from the left-hand side and the fifth line from the top. Those patents are mostly technology-related, with processes, devices, and systems for water purification. In contrast, the proposed approach failed to correctly predict several cells, such as Nos. 13, 23, 55, 81, 83, and 91 that mostly deal with new materials or methods. Thus, the results indicate that this approach is more successful in predicting incremental technology rather than radical technology, because new processes, devices, and systems can be generally developed by the improvement of previous technology, and new materials, and the methods have relatively more disruptive features.

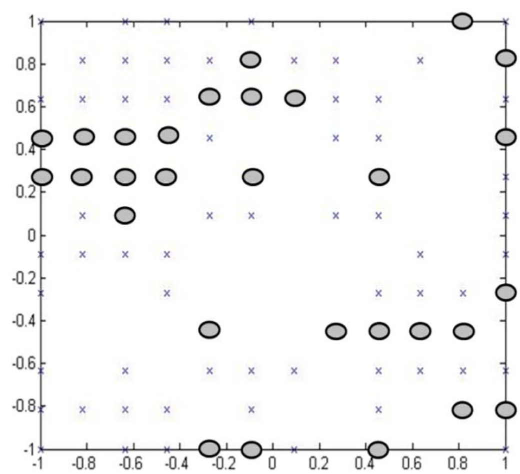
Dynamic analysis can identify the changing patterns of technology development over time. This paper posed two perspectives of technological changes: incremental and radical changes. From the developed maps, we can find two types of cells in which new technologies were added in comparison with the previous period in Fig. 7. While some cells can have patents in all periods for analysis (black-colored nodes in the third map), other cells that have no patent in the previous periods have newly developed patents (grey-colored nodes) in a corresponding period. In addition, the first type of cell can be divided into emerging cells and declining cells, because the changing patterns of always occupied cells are different in terms of the number of patents. In the aforementioned case of water purification, an emerging cell (No. 51) and a declining cell (No. 117) were derived among 8 always-occupied cells by investigating the number of patents over three time periods (2006–2008, 2009–2011, and 2012–2014). The emerging cells that can be identified as wastewater conversion methods have 1, 1, and 5 patents in each period. In contrast, the declining cells have patents related to systems and methods for water treatment, showing the decreasing patterns of patent applications, like 11, 14, and 1. Table 5 shows the change of patent applications in emerging and declining cells over time.

The second type of cell can be regarded as a relatively radical one, because the cells were newly occupied in each period. Table 6 presents the list of newly occupied cells and patents in two periods. First, in the period from 2009 to 2011, the newly occupied cells are nos. 23, 53, 55, 83, 91, 94, 97, 109, and 141 in the patent map of water purification. The cells are mostly associated with membrane modules and systems, or water treatment processes and equipment. Second, in the next period from 2012 to 2014, nos. 13, 81, 96, and 119 are newly occupied cells that were not occupied in the previous period from 2006 to 2011. The patents that are included in the cells mostly deal with filtrate monitoring and separation (removal) processes. While always occupied cells provide the information on changing patterns of existing technology, the newly occupied cells can be utilized to monitor the technology trends by investigating new technologies.

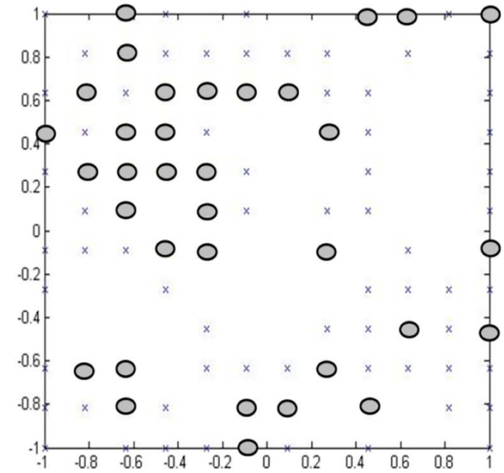
5. Discussions

5.1. Methodological implications

Since technology forecasting deals with the uncertainty of technology development and reflects complex information, relevant methodologies and data need to be applied for successful forecasting. This research utilizes patent documents as well as patent citation information, combining two main forms of patent information – contents, and bibliometric data. In addition, patent maps and patent networks are drawn together into a single canvas, by using the two types of information. Although they are a kind of visualization tool, their purposes



(a) 1st Patent map (2006-2008)



(b) 2nd Patent map (2009-2011)

Fig. 6. Evolutionary maps of water purification.

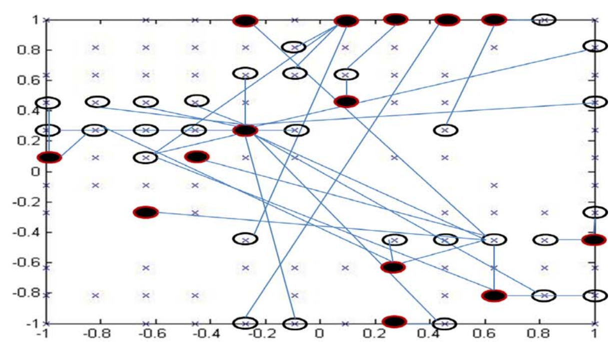
Table 2
Success and collaboration performance factors.

Performance factors	Approaches	
	Semantic similarity	Classification (based on SVM)
Lift	$\frac{7/13}{14/57} = 2.19$	$\frac{6/13}{18/57} = 1.46$
Precision	$\frac{7}{14} = 0.5$	$\frac{6}{18} = 0.33$
Recall	$\frac{7}{13} = 0.53$	$\frac{6}{13} = 0.46$

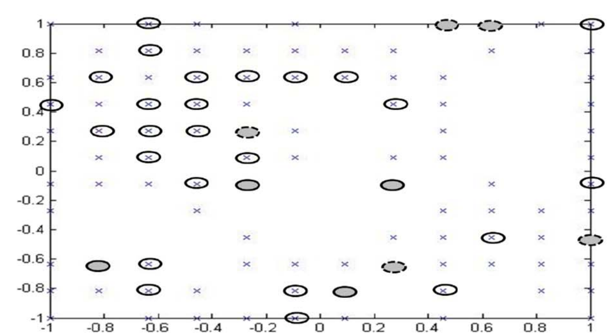
Table 3
Results of multiple regression analysis in the IT service industry.

Index	Technology		
	Water purification	3D printer	Nuclear fusion
Lift	$\frac{7/13}{14/57} = 2.19$	$\frac{3/11}{14/77} = 1.5$	$\frac{1/2}{8/20} = 1.25$
Precision	$\frac{7}{14} = 0.5$	$\frac{3}{14} = 0.21$	$\frac{1}{8} = 0.125$
Recall	$\frac{7}{13} = 0.53$	$\frac{3}{11} = 0.27$	$\frac{1}{2} = 0.5$
χ^2 value	4.107	0.464	0.037
p value	0.043*	0.496	0.847

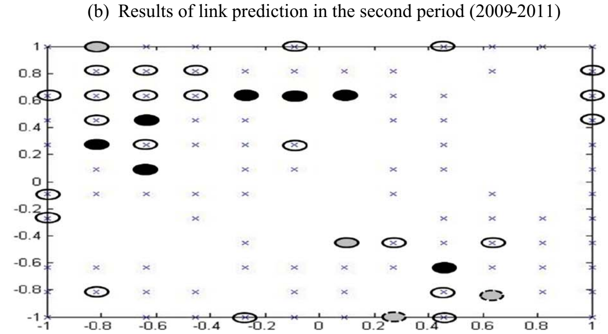
Note:
* Indicates the statistical significance under the 95% significant level.



(a) Predicted cells in the first period (2006-2008)



(b) Results of link prediction in the second period (2009-2011)



(c) Results of link prediction in the third period (2012-2014)

Fig. 7. Results of link prediction in the water purification.

and applications are totally different. While the critical results of maps are the location or distribution of data points, the important thing in a network is a link or relation among nodes. However, a combination of the two methods can provide ample information for technology opportunity analysis, because it enables the simultaneous exploration of vacant areas and link prediction in a single visualization space. Thus, the proposed approach can reflect complex information, using the two different but related methodologies of mapping and network analysis.

Although there are many visualization tools, in this research GTM is appropriate for developing a patent map. Since the aim of visualization in the proposed approach is to find vacant areas in a map where past patents are positioned, and moreover, provide the technological meaning of the vacant areas, the application of GTM is critical to grasp the estimated keyword vectors of each predicted cell. In order to

Table 4
Patents in the correctly and incorrectly predicted cells.

Predicted cells		Patents	
Correct prediction	53, 94, 96, 97, 109, 119, 141	US7601755(53)	Process for treating water
		US7686955(94)	Submerged hollow fiber membrane module
		US8142660(96)	Filtrate monitoring device, and filtrate monitoring system
		US8075776(97)	Apparatus for withdrawing permeate using an immersed vertical skein of hollow fiber membranes
		US7713418(109)	Process for recovering organic compounds from aqueous streams containing same
		US8449774(119)	Separation process
		US7767077(141)	Membrane filtration system
Incorrect prediction	13, 23, 55, 81, 83, 91	US8652333(13)	Solvent removal
		US7695628(23)	Polyarylether membranes
		US7922910(55)	Cyclic aeration system for submerged membrane modules
		US8414767(81)	Module for purifying a fluid, in particular water
		US7691268(83)	Waste gas/wastewater treatment equipment and method of treating waste gas/wastewater
		US7713419(91)	Method for treatment of sludge from waterworks and wastewater treatment plants

Table 5
Change of patent applications in always occupied cells.

Always occupied cells	Number of patents		
	2006–2008	2009–2011	2012–2014
117	11	14	1
17	2	1	3
28	3	1	2
29	3	3	1
30	3	1	1
51	1	1	5
63	1	2	2
75	3	2	3

Table 6
Patents in newly occupied cells.

Period (year)	Patents in newly occupied cells	
2009–2011	US7601755(53)	Process for treating water
	US7686955(94)	Submerged hollow fiber membrane module
	US7922910(55)	Cyclic aeration system for submerged membrane modules
	US7695628(23)	Polyarylether membranes
	US7691268(83)	Waste gas/wastewater treatment equipment and method of treating waste gas/wastewater
	US8075776(97)	Apparatus for withdrawing permeate using an immersed vertical skein of hollow fiber membranes
	US7713418(109)	Process for recovering organic compounds from aqueous streams containing same
	US7713419(91)	Method for treatment of sludge from waterworks and wastewater treatment plants
	US7767077(141)	Membrane filtration system
2012–2014	US8652333(13)	Solvent removal
	US8142660(96)	Filtrate monitoring device, and filtrate monitoring system
	US8449774(119)	Separation process
	US8414767(81)	Module for purifying a fluid, in particular water

Note: the numbers of parentheses indicate the cell number of patent map.

support the process of technology planning, the decision makers require the details of technology opportunities, rather than rough information on the location of vacant areas. The GTM can provide a possible set of keywords for each vacant cell by analyzing the keyword vectors of existing patents.

Various link prediction techniques can be applied to predict potential links in patent networks. Since the techniques show different prediction performance in a specific circumstance of analysis, the characteristics of techniques and data should be deeply considered to obtain the successful results of forecasting. This paper compared the prediction performance of two main techniques - SVM and semantic similarity. Although the neighborhood-based approach and path-based approach are frequently employed in link prediction, they are not applicable in this research, because they aim to predict missing links among existing nodes. The proposed approach finds the vacant cells, generating new nodes in the patent network, before exploring missing links. Thus, we cannot utilize the information on the links that existing nodes have, because new nodes have no links with existing nodes. As a result, the performance of the semantic similarity-based approach is higher than that of SVM. Two possible reasons can be discussed to analyze the result. First, the size of data that is utilized in this research is insufficient to get strong results in applying the machine learning approach. Although SVM can achieve high performance in a prediction problem, the number of the collected patent data is less than 1000 patents in all three cases. Second, the amount of information that is utilized in the two analyses is not equal, resulting in the difference of performance. While the semantic similarity-based approach utilizes just keyword vectors to analyze the similarity among nodes, SVM applies keyword vectors and citation information. The learning process to relate the similarity between keyword vectors with citation might make the patterns to predict the missing links complex and inaccurate, in the situation that there are a limited number of learning data. Consequently, the semantic similarity-based approach is a proper link prediction technique in the proposed process, showing higher performance.

5.2. Managerial implications

All forecasting methods show lower prediction power over time. In this research, since the predicted technology opportunities might appear in the far future, the time bucket for forecasting is divided into two periods. As a result, the general assumption on the prediction power was validated in the analysis of technology opportunity. However, even if the predicted technology opportunities scarcely appear in the long-term period, an influential technology that has a great impact on the products and markets might appear in the far future. Therefore, a manager who is in charge of technology planning should be aware of the possibility of the long-term realization of technology.

The result of technology forecasting in this paper does not determine the probability of technology appearance, but the list of potential technology. Thus, the analysts should focus on the opportunities rather than the forecasting errors, paying great attention to monitoring the change of technology. Since the proposed approach investigates abundant information based on systematic methodologies, its results should be regarded as supporting information for making a decision on technology planning under a complex situation of technology, markets, and competition. Thus, analysts need to combine other information, such as markets, products, and competitors, with the results derived from the proposed approach, in order to successfully identify and implement technology opportunities.

Since many types of technology can be considered to find proper categories for which a forecasting technique can show good performance, analysts should be careful to apply the proposed approach to real cases. From the results in the comparison of various technology types, the technologies that have a stable pattern of patent application are appropriate for this approach. Other types that show emerging and declining patterns do not provide significant results, because there are too many or too few vacant areas for new technology opportunities. If a technology has many vacancies in the map, the hit ratio of prediction is relatively low. In contrast, in the case of declining pattern, the proposed approach cannot recommend considerable technology opportunities, because most areas in the map are already occupied, meaning that there are too few opportunities. Thus, in such types of technology, other approaches should be applied to explore new technology opportunities.

6. Conclusions

The exploration of new technology opportunities promotes companies obtaining a competitive edge, and connection to the growth engines of the economy. Although many researchers have tried to propose various methods to perform technology opportunity analysis, in the main they focused little on presenting the details of technology opportunities, showing instead the rough technology areas or classifications. Thus, this paper aims to propose a systematic approach to find technology opportunities, presenting the detailed features of technology. For this, GTM and link prediction are employed as the main methodologies of mapping and forecasting. Keywords that are extracted by text mining techniques are utilized to visualize patents of interests, and the characteristics of vacant cells in a patent map are explained by keywords that are estimated from the inverse function of GTM. In addition, the link prediction process plays a critical role in forecasting new technology opportunities. This paper compared the prediction power of two popular link prediction methods to select the best tool for technology opportunity analysis. The proposed approach has several advantages from the perspectives of data, methodologies, and insights. Since patent documents and patent citation information are important in technology analysis, they need to be combined to perform sophisticated technology forecasting. This research utilizes the two main data sources of technology analysis for visualization and prediction, respectively. In terms of methodologies, the integration of maps and networks can provide a single landscape for technology opportunity analysis. Although the existing literature deals separately with the two methodologies, a visual form that includes the locations of patents, as well as the relations among patents, can help analysts effectively identify new technology opportunities. Finally, from the perspective of insights, this research proposes a process to find technology types where the suggested approach can successfully be applied. In addition, the results enable the specific features of technology opportunities to show not just areas, but the characteristics of technology, providing critical insights for technology planning.

Despite the aforementioned advantages, this research has some limitations. First, although two link prediction tools are selected and compared, more methods can be included to find the most relevant method for link prediction. Since technology opportunity analysis should consider totally new technology areas, conventional link prediction tools that utilize existing nodes in a network seem to be inappropriate. In particular, SVM that is actively utilized to solve a prediction problem provided a low prediction performance due to the data issue in this paper. Second, three types of technology were chosen in this paper, in order to compare the prediction performance according to types. However, since three representative technologies were utilized to investigate the performance of the proposed approach in each category, the results might be unable to provide a generalized insight. This weakness is best remedied by performing our methods on a very large number of domains. Third, the subjective opinions of analyzers intervene in interpreting the meanings of potential vacant cells with the estimated keywords. Although a set of multiple keywords is used to understand the characteristics of cells in a patent map, more systematic process should be proposed to objectively perform the interpretation of technology opportunities. Finally, since the proposed approach can be applied to forecast an incremental improvement of existing technologies, potential radical technology cannot be explored at the present time. It appears possible that along with studying more domains that a broader data-based approach can enable finding radical technology opportunities by analyzing various information on markets and industries in addition to technology/patents. Although seminal papers try to investigate outliers that are distant from a group of patents in a patent map to derive the opportunities of radical technology development, more appropriate information such as technology breakthrough and new social/market trends need to be analyzed. However, this paper concentrates on suggesting a new approach that can utilize patent-specific information to reflect the patterns of patent development. Thus, future research can improve the proposed approach, by trying to solve the limitations of this research. More advanced link prediction methodology can be proposed to conduct the TOA, modifying existing methodologies such as SVM and network analysis. In addition, a greater data set of technology can be analyzed to clearly derive the characteristics of technology where the proposed approach is successfully workable. Moreover, a study that applies various approaches, such as SAO and LDA, can improve the process of interpreting the meaning of technology opportunities.

Acknowledgments

This work was supported by the National Research Foundation of Korea Grant funded by the Korean Government (NRF-2017R1D1A1 A09000758).

References

Aboody, D., Lev, B., 2001. R&D Productivity in the Chemical Industry, Working Paper. New York University, New York, NY.
Albert, R., Barabási, A.L., 2002. Statistical mechanics of complex networks. Rev. Mod. Phys. 74, 47.
Benson, C.L., Magee, C.L., 2013. A hybrid keyword and patent class methodology for selecting relevant sets of patents for a technological field. Scientometrics 96 (1), 69–82.
Benson, C.L., Magee, C.L., 2015. Quantitative determination of technological improvement from patent data. PLoS One 10 (4), e0121635.
Bishop, C.M., Svensen, M., Williams, C.K.I., 1998. GTM: The generative topographic mapping. Neural Comput. 10 (1), 215–234.
Blei, D.M., Ng, A.Y., Jordan, M.I., 2003. Latent dirichlet allocation. J. Mach. Learn. Res. 3, 993–1022.
Boccaletti, S., Latora, V., Moreno, Y., Chavez, M., Huang, D.U., 2006. Complex networks: structure and dynamics. Phys. Rep. 424, 175.
Chang, S.B., Lai, K.K., Chang, S.M., 2009. Exploring technology diffusion and classification of business methods: using the patent citation network. Technol. Forecast. Soc. Chang. 76 (1), 107–117.
Chen, R., 2009. Design patent map visualization display. Expert Syst. Appl. 36 (10), 12362–12374.
Cheng, Y., Huang, L., Ramlogan, R., Li, X., 2017. Forecasting of potential impacts of disruptive technology in promising technological areas: elaborating the SIRS epidemic model in RFID technology. Technol. Forecast. Soc. Chang. 117, 170–183.
Choi, S., Park, H., Kang, D., Lee, J.Y., Kim, K., 2012. An SAO-based text mining approach to building a technology tree for technology planning. Expert Syst. Appl. 39 (13), 11443–11455.
Coates, V., Farooque, M., Klavans, R., Lapid, K., Linstone, H., Carl, P., Porter, A., 2001. On the future of technological forecasting. Technol. Forecast. Soc. Chang. 67, 1–17.
Costa, L., Rodrigues, F.A., Travieso, G., Boas, P.R.U., 2007. Characterization of complex networks: a survey of measurements. Adv. Phys. 56, 167–242.
Cozzens, S., Gatchair, S., Kang, J., Kim, K.S., Lee, H.J., Ordóñez, G., Porter, A., 2010. Emerging technologies: quantitative identification and measurement. Tech. Anal. Strat. Manag. 22, 361–376.
Criscuolo, P., 2006. The 'home advantage' effect and patent families. A comparison of OECD triadic patents, the USPTO and the EPO. Scientometrics 66 (1), 23–41.
Dismukes, J.P., Bers, J.A., Miller, L.K., McCreary, W.N., 2005. Technologies of thinking seen key to accelerated radical innovation. Res. Technol. Manag. 48, 2–4.
Enríquez, F., Troyano, J.A., López-Solaz, T., 2016. An approach to the use of word embeddings in an opinion classification task. Expert Syst. Appl. 66, 1–6.
Gans, J., 2016. The Disruption Dilemma. MIT Press, Cambridge, MA.
Glänzel, W., Meyer, M., 2003. Patents cited in the scientific literature: an exploratory

study of ‘reverse’ citation relations. Scientometrics 58, 415–428.
Granstand, O., 1999. The Economics and Management of Intellectual Property: Toward Intellectual Capitalism. Edward Elgar Publishing, Cheltenham.
Hellmann, T., 2007. The role of patents for bridging the science to market gap. J. Econ. Behav. Organ. 63 (4), 624–647.
Kerr, C.I.V., Mortara, L., Phaal, R., Probert, D.R., 2006. A conceptual model for technology intelligence. Int. J. Technol. Intell. Plan. 2 (1), 73–93.
Kim, H., Hong, S., Kwon, O., Lee, C., 2017. Concentric diversification based on technological capabilities: link analysis of products and technologies. Technol. Forecast. Soc. Chang. 118, 246–257.
Kim, J., Lee, S., 2015. Patent databases for innovation studies: a comparative analysis of USPTO, EPO, JPO and KIPO. Technol. Forecast. Soc. Chang. 92, 332–345.
Kwakkel, J., Carley, S., Chase, J., Cunningham, S.W., 2014. Visualizing geo-spatial data in science, technology and innovation. Technol. Forecast. Soc. Chang. 81, 67–81.
Lee, C., Jeon, J., Park, Y., 2011. Monitoring trends of technological changes based on the dynamic patent lattice: a modified formal concept analysis approach. Tech. Forcasting Soc. Chang. 78 (4), 690–702.
Lee, C., Kang, B., Shin, J., 2015. Novelty-focused patent mapping for technology opportunity analysis. Technol. Forecast. Soc. Chang. 90, 355–365.
Lee, J., Kim, C., Shin, J., 2017. Technology opportunity discovery to R&D planning: key technological performance analysis. Technol. Forecast. Soc. Chang. 119, 53–63.
Lee, S., Yoon, B., Park, Y., 2009. An approach to discovering new technology opportunities: keyword-based patent map approach. Technovation 29 (6/7), 481–497.
Lee, W.S., Han, E.J., Sohn, S.Y., 2015. Predicting the pattern of technology convergence using big-data technology on large-scale triadic patents. Tech. Forcasting Soc. Chang. 100, 317–329.
Leskovec, J., Huttenlocher, D., Kleinberg, J., 2010. Predicting positive and negative links in online social networks. In: Proceedings of the 19th International Conference on World Wide Web, Raleigh, NC, USA, 26–30 April 2010.
Liben-Nowell, D., Kleinberg, J., 2003. The link prediction problem for social networks. In: International Conference on Information and Knowledge Management, New York, NY, USA.
National Research Council, 1986. Planning for Energy Conservation R&D: A Review of the DOE's Planning Process. The National Academies Press, Washington, DC.
Newman, M.E.J., 2001. Clustering and preferential attachment in growing networks. Phys. Rev. E 64, 025102.
Newman, M.E.J., 2003. The structure and function of complex networks. SIAM Rev. 45, 167.
Olsson, O., 2005. Technological opportunity and growth. J. Econ. Growth 10, 35–57.
Porter, A.L., Newman, N.C., 2011. Mining external R&D. Technovation 31, 171–176.
Savioz, P., Blum, M., 2002. Strategic forecast tool for SMEs: how the opportunity landscape interacts with business strategy to anticipate technological trends. Technovation 22, 91–100.
Schumpeter, J.A., 1939. Business Cycles: A Theoretical, Historical and Statistical Analysis of the Capitalistic Process. McGraw-Hill, New York, NY.
Shibata, N., Kajikawa, Y., Sakata, I., 2010. Extracting the commercialization gap between science and technology—case study of a solar cell. Technol. Forecast. Soc. Chang. 77, 1147–1155.
Tang, J., Wang, B., Yang, Y., Hu, P., Zhao, Y., Yan, X., 2012. Patentminer: topic-driven patent analysis and mining. In: Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining 2012, pp. 1366–1374.
Wang, C., Satuluri, V., Parthasarathy, S., 2007. Local probabilistic models for link prediction. ICDM'07. In: Proceedings of International Conference on Data Mining, October 28–31, Washington, DC, USA.
Wang, M.Y., Chang, D.S., Kao, C.H., 2010. Identifying technology trends for R&D planning using TRIZ and text mining. R D Manag. 40, 491–509.
Wang, M.Y., Fang, S.C., Chang, Y.H., 2015. Exploring technological opportunities by mining the gaps between science and technology: microalgal biofuels. Tech. Forcasting Soc. Chang. 92, 182–195.
Yoon, J., Park, H., Kim, K., 2013. Identifying technological competition trends for R&D planning using dynamic patent maps: SAO-based content analysis. Scientometrics 94 (1), 313–331.

Byungun Yoon is a Professor in Department of Industrial & Systems Engineering of the Dongguk University. His theme of study has involved patent analysis, new technology development methodology and visualization algorithms. His current interest is in enhancing technology road mapping and product designing with data mining techniques.

Professor Christopher L. Magee is a member of the National Academy of Engineering, a fellow of ASM and SAE and a participant on major National Research Council Studies. He is a native of Pittsburgh, PA and received his B.S. and Ph.D. from Carnegie-Mellon University in that city. He later received an MBA from Michigan State University. He is professor of the Practice, Engineering Systems Division and Mechanical Engineering, and co-director of Engineering Design and Advanced Manufacturing Focus Area of the MIT-Portugal Program, Massachusetts Institute of Technology.