Technological Forecasting & Social Change 192 (2023) 122565

Contents lists available at ScienceDirect

Technological Forecasting & Social Change

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

From technology opportunities to ideas generation via cross-cutting patent analysis: Application of generative topographic mapping and link prediction

Zhenfeng Liu [a,*], Jian Feng [b], Lorna Uden [c]

[a] School of Economics and Management, Shanghai Maritime University, Shanghai 201306, China
[b] China Institute of FTZ Supply Chain, Shanghai Maritime University, Shanghai 201306, China
[c] School of Computing, Staffordshire University, Stoke-on-Trent ST4 2DE, United Kingdom

ARTICLE INFO

Keywords:
Patent analysis
Technology opportunity analysis
Idea generation
Generative topographic mapping
Link prediction

ABSTRACT

Technology opportunity analysis (TOA) with ideas generation has been recognized an important activity to remain competitive and lead the industry in the future. However, there are several issues with existing TOA, such as an unclear path from technology opportunities to ideas generation, a fuzzy integration between automated TOA techniques and expert-based methods, and a lack of detailed schemes for technology opportunities. This study proposes a new systematic approach to show the way from technology opportunities to ideas generation via cross-cutting patent analysis. The proposed approach is comprised of three stages: 1) establishing a cross-cutting relationship between the target and reference technologies through the results of F-terms; 2) collecting and processing patents to construct patent-keyword vector matrices of the target and reference technologies, respectively; and 3) migrating corresponding ideas via cosine similarity and link prediction for the target technology opportunities that are discovered based on generative topographic mapping (GTM). The feasibility and effectiveness of the proposed approach is demonstrated by empirical research on the exploitation technology in both the natural gas hydrate (NGH) and the coal bed methane (CBM) fields. This study represents a contribution to expand the existing TOA research into generating creative ideas by providing more detailed schemes for technology opportunities.

1. Introduction

Technology opportunity analysis (TOA) and new creative ideas generated from it play a critical role in the firms' successful research and development (R&D) activities (Song et al., 2017; Zhu and Porter, 2002). According to researchers, traditional technology opportunity is defined as "the potential for technological advancement in a specific domain" (Olsson, 2005) or "a combination of monitoring and bibliometric analysis" (Porter and Detampel, 1995). Idea generation is described as the front-end part of the innovation and focuses on coming up with possible solutions to perceived or actual technology opportunities (McAdam and McClelland, 2002). According to Toubia and Netzer (2017), idea generation involves two phases: a *generative* phase, in which a set of knowledge elements are constructed, and an *exploratory* phase, in which these elements are interpreted in a meaningful way. The former phase can be addressed more effectively using existing TOA approaches, but little is known about how the two phases are related. Using link prediction, it is possible to establish a path from technology opportunities to ideas generation. Link prediction methods have been employed to forecast future links between the vacant areas and the existing patents (Yoon and Magee, 2018). But this approach is limited to existing technologies in a single domain rather than across domains but not both. In addition, existing research focuses on either technology opportunities or ideas generation among cross-domains (Ren and Zhao, 2021; Song et al., 2017). The innovation path between them is not clear. We would like to find out, how do we make a clear link relationship between technology opportunities and ideas generation among cross-domains?

Typically, the process from technology opportunities to ideas generation can use both qualitative and quantitative methods. The technology roadmap and Delphi rely heavily on various opinions of domain

* Corresponding author at: School of Economics and Management, Shanghai Maritime University, 1550 Haigang Avenue, Pudong District, Shanghai 201306, China.

E-mail address: zfliu@shmtu.edu.cn (Z. Liu).

https://doi.org/10.1016/j.techfore.2023.122565
Received 9 June 2022; Received in revised form 20 February 2023; Accepted 4 April 2023
Available online 25 April 2023

experts in the aspect of qualitative methods. Although domain experts frequently offer insights into detailed ideas for particular technology opportunities, these qualitative methods have disadvantages such as the subjectivity and restricted availability (Park and Yoon, 2017). Conversely, quantitative TOA methods can compensate expert-based methods and provide objective suggestions on the technology opportunities using highly automated techniques that allow easy and effective understanding of the complex information (Lee et al., 2015; Yoon and Magee, 2018). For this, many scholars argued that data mining and the dimensionality reduction methods, such as generative topographic mapping (GTM), can be used to find the technology vacancy for TOA in recent years (Feng et al., 2021; Son et al., 2012; Teng et al., 2021; Yoon and Magee, 2018). These systematic methods have proven useful in reducing the manual work required to analyze unstructured and long text, by identifying a few salient features in the sets, including keywords, patent classification codes, F-terms themes. In terms of TOA methods via cross-domains, technology opportunities with highly automated techniques normally discover vacant and coarse-grained information, whereas ideas generation with finer granularity often require more expert experience and are always vary from person-to-person. How to integrate automating TOA processes and interpreting ideas to form systematic methods among cross-domains are important issues to be addressed.

Furthermore, current TOA quantitative approaches have limitation because they are not sufficiently detailed to allow for generating detailed R&D ideas (Lee and Lee, 2019; Ren and Zhao, 2021; Yoon and Magee, 2018). Since the technology opportunity is defined by examining a few fragmented characteristics located in the border of the technology vacancy, its true meaning has been unspecified. Therefore, to facilitate TOA process, more precise and detailed information rather than a collection of coarse-grained knowledge elements should be provided. In addition, existing TOA only focuses on a single field of the technology opportunity. As a result, the conclusions reached using current methods are frequently inadequate in defining the details of the technology opportunity and difficult to link to practical measures for R&D planning. Therefore, further research is needed so that it can be adapted to different research contexts. According to Yoon et al. (2015) and Moaniba et al. (2018), TOA is expanding across different industries through recombinant innovation. It is no longer limited to mere potential opportunities using few salient and fragmented features. There can be many possible relationships among these features, thereby blurring the true meaning of technology opportunities. Instead, the technology opportunities are identified through the derivation of new technology ideas and defined as "the potential for technical advancement through value creation from novel ideas" (Ittipanuvat et al., 2014; Song et al., 2017). These studies extended the traditional definition of the TOA to generate creative ideas for technology opportunities. As a response from the method, link prediction began to be used to predict potential linkages based on existing unconnected nodes within a single field (Yoon and Magee, 2018). It is necessary to develop a new method for the prediction of unknown network among cross-domains. Therefore, how to provide specific guidelines for practitioners to generate ideas via cross-domains and their feasibilities as a technology opportunity is still missing.

To address these research questions, this study proposes a new systematic approach to identify the technology opportunities along with ideas generation from different technical fields. To achieve this aim, we integrate GTM and link prediction via cross-cutting patent analysis. To be specific, we first choose a target technology for which to seek new opportunities and establish a cross-cutting relationship between the target and reference technologies through collecting F-terms, a patent classification code for multiple technical viewpoints. Once the cross-cutting relationship is established, the relevant patents are collected based on certain search strategy and we process using text mining technique. Next, the target technology opportunities are discovered based on the dimensionality reduction technique (e.g., GTM). However, these opportunities are vacant areas composed of fragmented keywords and do not provide specific guidelines for practitioners to generate ideas and their feasibilities as a technology opportunity. We obtain keywords intersection between target and reference technologies to construct the patent-keyword vector matrices for the reference technology and generate GTM-based map. On this map, existing patents of the reference technology are hidden behind non-vacuums and can be used as a source of ideas generation for vacuums of the target technology. To establish the link relationship between two target and reference technologies, we use cosine similarity and link prediction. Cosine similarity is used to connect vacuums of the target technology with non-vacuums (existing patents) of the reference technology. We split the connection data into the training and test data for the link prediction. In addition, we also choose one advanced link prediction index with the highest accuracy by comparing different link prediction methods to forecast the link between GTM-based vacuums of the target technology and non-vacuums of the reference technology. Because each non-vacuum hides existing patents of the reference technology, we evaluate these patents' value and recommend high-value patents for the target technology opportunities. Finally, the established link relationships prove that keywords contained in the vacuum are highly overlapped with recommended patents which can assist experts to qualitatively generate detailed schemes and generate creative ideas for GTM-based technology opportunities. Thus, the technology migration between the target and reference technologies based on cosine similarity and link prediction is achieved.

This study differs from previous studies as follows. 1) It paves a clear path from technology opportunities to ideas generation across domains, not domain-specific schemes, by use of the link prediction. The proposed approach takes TOA one step further to generate creative ideas by cross-cutting patent analysis. Previous research was to establish connections and identify technology opportunities within a single field and an existing network. The results imply that the cosine similarity and link prediction via cross-domains can erase inconsistencies in the keywords and concepts shared between two fields. 2) Our tailored approach starts by automatically identifying technology opportunities that contain multiple keywords. To avoid random keywords combination with coarse granularity, cross-cutting patents are recommended for these keywords based on the link prediction. The results combine fine granularity of ideas generation and automatic identification of technology opportunities. 3) The suggested approach of integrating GTM and link prediction is helpful in terms of providing information that cannot be easily captured by qualitative judgment of several domain experts. These recommended patents via cross-domains can help experts to generate detailed technical schemes for the target technology.

The rest of this study is organized as follows: The literature review on TOA, ideas generation for the technology opportunities, GTM and link prediction is presented in Section 2. The research framework and the overall process are described in Section 3, and a case study is introduced in Section 4. Finally, Section 5 concludes our discussion of this work, and Section 6 lists some contributions and limitations.

2. Literature review

A literature review related to the purposes of this study is presented. Particular attention is paid to examine the patent analysis as well as techniques for TOA and the idea generation.

2.1. Patent-based technology opportunities analysis

Patents are generally selected as data sources for identifying promising technology opportunities in a quantitative manner since patents include a lot of bibliographic information about future technology (Ardito et al., 2018; Ardito et al., 2022; Kumar et al., 2021; Lai et al., 2021c). These studies are used to identify technology opportunities through technological hotspots and development trends (Albino et al., 2014; Evangelista et al., 2020; Feng and Magee, 2020) and patent citation analysis (Daim et al., 2020; Lai et al., 2021a; Lai et al., 2021b) in a single field. With the emergence of text mining and automatic algorithms, patent-based TOA is expected to reduce dependency on experts and enhance the creativity through investigating vast patent text of technological documents published worldwide (Song et al., 2017). Many researchers have attempted to identify the technology opportunities through patent analysis and examining various knowledge elements, such as keywords (Wang and Chen, 2019; Yoon and Magee, 2018), patent classification codes (Park and Yoon, 2017; Yu and Zhang, 2019), and semantic structure (Han et al., 2021; Kim et al., 2019b; Ren and Zhao, 2021). However, current studies only concentrated on TOA within a specific area of interest. To overcome this limitation, Ittipanuvat et al. (2014) introduced an approach of extracting plausible linkages between two different domains (e.g. robotics and gerontology) for leveraging existing knowledge to find potential new technology opportunities. Nakamura et al. (2015) created a knowledge combination model between two technical fields to collect breadth technology knowledge in other domains and to combine it to its own knowledge. However, there are also limitations in these two approaches because they only identified the technology opportunities among cross-domains, rather than generating ideas for them. Song et al. (2017) expanded the research on the technology opportunity cross domains to tap into the actual practice of generating technical schemes, and suggested an approach for developing novel technology ideas by combining the reference technologies with the target technology based on the F-terms. But the ideas generation process was still limited to the F-terms level and did not go deep into detailed schemes with recommending the cross-domain patents. Thus, because the true meaning of the set of F-terms has not been specified, the outcomes are frequently weak when developing the details of the target technology. In summary, TOA based on the knowledge elements generally lacks sufficient detail and have coarse granularity for identifying more refined technology opportunities, resulting in ambiguous technology opportunities. Therefore, a finer granularity with the automation of the TOA processes is necessary.

Our research aims to overcome the above stated limitations and is an extension of Song et al. (2017). The motivation of this study is similar to Song et al. (2017), but the scope of our research on technology opportunities is extended to the generation of potential technology ideas. We first identify the target technology opportunities and then recommend patents with detailed schemes of the reference technology through establishing the link relationship between the target technology and the reference technology. We provide specific guidelines on the target technology opportunities for generating new ideas and their potentials as a technology opportunity. Compared with Song et al. (2017), our research on ideas generation along with the technology opportunities will go deep into the level of fine-grained technology schemes rather than coarse-grained F-terms.

2.2. Idea generation

Typically, it is assumed that the innovation is a process with several stages ranging from idea generation to commercialization (McAdam and McClelland, 2002). R&D engineers have attempted to enrich existing knowledge elements and generate high-quality ideas in the innovation process. From a single domain perspective, the morphology analysis is an effective method to recombine current technology knowledge and generate creative ideas. Geum and Park (2016) leveraged WordNet and Kwon et al. (2018) utilized Wikipedia to extend technical morphologies as an additional tool for generating ideas. These recent studies emphasized the underlying principle that creative ideas are constructed from existing knowledge (Lee and Lee, 2019). To meet consumer's demands and solve practical technical problems, scholars have begun to combine existing knowledge with cross-cutting technical needs. For example, Yan and Li (2022) proposed a framework in which functions serve as a bridge between technologies and consumer needs. However, these studies still generated ideas through the combination of fragmented knowledge elements, whether it is WordNet, Wikipedia or functions, but lack sufficient detail for R&D engineers to use. In addition, the idea generation may look irregular, dynamic and fuzzy due to the lack of technology opportunities that can be identified by automated techniques, and therefore requires to be carried out using experts' knowledge (Christiansen and Gasparin, 2016; Geum and Park, 2016).

To address the above gaps, this study first utilizes GTM to identify the target technology opportunities, then recommends patents from the reference technology and generate ideas for these opportunities by using link prediction. Although Yoon and Magee (2018) integrated GTM and link prediction, this study differs from previous studies in that we recommend patents and detailed schemes for GTM-based keywords from the cross-domain technology using the link prediction. We strictly follow that idea generation involves a generative phase (i.e., automatic TOA) and an exploratory phase (i.e., expert-based explanation). We first identify technology opportunities including fragmented keywords for the target technology by using GTM. Then, we recommend and evaluate patents for technology opportunities from the cross-domain technology. To do this, we use the cosine similarity to establish similarity connection between the target and the reference technology. The connection data are split into the training and test data for link prediction. We choose one advanced link prediction index with the highest accuracy by comparing different link prediction methods to forecast the link between GTM-based vacuums of the target technology and non-vacuums of the reference technology. In this way, the weak meaning of the technology vacancy and ideas generation of excessive expert-based interpretation are avoided.

2.3. Link prediction

In any network analysis, predicting relationships between nodes and determining how likely links are to exist are always important research issues. Link prediction is an effective method for predicting possible future links based on existing network (Chen et al., 2020; Lü and Zhou, 2011). Originally, since the link prediction can observably lower experimental costs associated with analyzing biological interactions, it was applied in bioinformatics and biology (Lei and Ruan, 2013). It was later extended to other research fields. This method has begun to play an important role in technological innovation research (Lee, 2021), such as the technology convergence (Kim et al., 2019a) and the technology road mapping (Kim and Geum, 2021). As for TOA research, Ma et al. (2019), Han et al. (2021) and Lee et al. (2021) have combined the link prediction with other techniques to identify technology opportunities.

However, despite the popular use of the link prediction, there are limitations. It is because the link prediction only predicts potential linkages based on present nodes (Yoon and Magee, 2018). As a result, technology opportunities could only be discovered within a single field. To overcome this, in our research, we recommend patents and generate ideas for the technology opportunities from the cross-domain technology using link prediction. According to (Yoon et al., 2021), if a node represents a document in a network, keywords vector are very helpful for examining the similarity between two documents in content-based approach. For this reason, our research adopted the content-based approach, where the cosine similarity of the keyword vector that is suggested for measuring relatedness between two documents generally performs the best compared to other measures (Ittipanuvat et al., 2014), to investigate the similarity between two technologies across domains. Thus, the cosine similarity is used to establish a connection between two unconnected non-vacuums from cross-cutting technologies. Then, this study compares different link prediction methods to forecast the link between two technologies across domains and chooses one advanced link prediction index with the highest accuracy. The formula for each link prediction index used in this study is given in Appendix A.

3. Methodology

3.1. Research framework

This study proposes a new systematic approach to identify technology opportunities with ideas generation via cross-cutting patent analysis. As shown in Fig. 1, the process suggested in this study can be used from technology opportunities analysis to ideas generation. This study assumes that an emerging technology opportunity can be generated into ideas by linking to other mature fields. Firstly, a cross-cutting relationship between two technologies (the target and reference) in different areas is established through the statistics and analysis of patent classification codes, such as IPCs and F-terms. Experts' participation is essential to select one of reference technologies as the technology migration source for the target technology in this stage. We concur with Kim and Geum (2021) that expert-based method cannot be neglected despite the importance of automated techniques. Secondly, patents are collected and processed into a vector matrix for the dimensionality reduction technique. Then, the GTM algorithm is employed to discover technology opportunities for the target technology. Thirdly, we migrated creative ideas of the target technology to cater for technology opportunities from recommended reference technology patents, with the help of link prediction. The overall process is depicted in Fig. 1 with three stages: establishing a cross-cutting relationship, collecting and processing data, and migrating technology, corresponding directly to three subsections.

3.2. Research process

3.2.1. Establishing a cross-cutting relationship

In this stage, we establish the cross-cutting relationship between the target and reference technologies into four steps.

Step 2-1: Choosing a target field.

In the first step, we choose a target field for the opportunity analysis. To avoid infringement of patent rights in the reference field, the target field should be emerging and convenient for generating creative ideas.

Step 2-2: Selecting the target technology.

We select representative keywords in the target field and search for patents in at least two patent databases. A 4-digit IPC with the largest proportion in the statistical graph as the target technology is selected and extracted to generate a statistical graph which represents the technology composition of the target field. According to Zhang et al. (2016), the 4-digit IPCs are used because they not only serve as fundamental categories of the IPC classification system, but also provide a technology option for research on cross disciplines.

Step 2-3: Searching for F-terms.

We build the migration link between the reference technology and the target technology via the F-terms, which are Japan Patent Office's (JPO) original search keys. To eliminate the need for the expert opinion in the patent analysis, we use F-terms to classify patents according to specific technical attributes (such as purpose, usage, structure, material, etc.) in which technologies can be compared. By using technical attributes instead of technical content, technologies with similar inventive principles are identified and the principles applied in other fields are suggested (Kashimoto, 2016; Song et al., 2017). This facilitates the ideas generation process from different technical fields for the target technology opportunities.

We utilize the F-term corresponding to the abovementioned 4-digit IPC in Step 1-2 to determine the technologies to be referenced for the target technology. The connection between F-term and IPC is established through the file index (FI) and can be found through IPC-FI-F-term concordance list which is stored in the Patent/Utility Model Classification Search (PMGS). For example, choose "Code Inquiry" tab, enter the IPC symbol in the "Classification" box and then click "Search" button to find FI that corresponds to the inputted IPC symbol as shown in Fig. 2. We also enter FI symbol into "Classification" box and find corresponding F-term.

Step 2-4: Determining a reference technology.

Compared to IPC, the F-term provides a much wider range of technologies according to detailed technical attributes, as shown in Fig. 3. The F-term features nine digits. 2-digit F-term represents an industry (e.g., 2D: Energy mining for coal, gas or oil, etc.), while 5-digit F-term is one technology in this industry (e.g., 2D129: Earth drilling). As for 2D129, there are 19 technical attributes or dimensions (from AA to JA) and 2–48 technical shapes in each dimension (e.g., 2D129 AA01: Underwater). The 7-digit F-term (e.g., 2D129 AA: Object to be drilled) represents technical viewpoints from different technical dimensions, such as object, use, purpose, method, application, etc. Thus, the technology with 5-digit code is treated as a combination of all technical dimensions with 7-digit code. Technology configurations in different fields can be implemented by replacing other technical shape within each technical dimension. This process is consistent with morphology analysis of Kwon et al. (2018). The target technology and many reference technologies fall within these technology configurations resulting in a high potential to be good references in developing creative ideas. Lastly, we select a reference field and continue with above Step 1-2. We also check if the 4-digit IPC with the largest proportion is consistent with the target field. If not, we switch to another reference field. This enables us to verify the consistency between the target and reference technologies and establishing the cross-cutting relationship.

3.2.2. Collecting and processing data

In this stage, once we have established the cross-cutting relationship between the target and reference technologies, the relevant patents are collected based on a certain search strategy and we processed them using data mining techniques.

Step 2-1: Collecting data.

In this step, we concur with Oltra-Garcia (2012) the linear search is the most logical and simple structuring to narrow down on the target and reference technologies as precisely as possible. Firstly, we identify what to search. It is important to collect all useful information relevant to the target and reference fields in *Google Search*, identify all the related concepts that will become the object of the search. Secondly, we narrow down on the object of the search by determining how and the where to search. By using the information collected in the first step, brief keyword searches are often used separately for the target and reference technologies. This allows us to decide the best search strategy and in which patent database to search, whether to use IPC codes, keywords, date range or combinations thereof, etc. In our study, we collected patents on the target and reference technologies in Derwent innovations index (DII), and define them as Set p1 and Set p2, respectively. More than 40 patent organizations from across the world provide the vast majority of patents in DII database. For categorizing patent information, patents are rewritten into English in DII.

Step 2-2: Processing data.

Yoon and Magee (2018) argued that, although patents collected in DII are manually rewritten, it is necessary to transform unstructured patent texts into meaningful structured data and form a patent-keyword matrix by using text mining techniques. To achieve this task, firstly, we extracted keywords of the target technology (Set k1) and keywords of

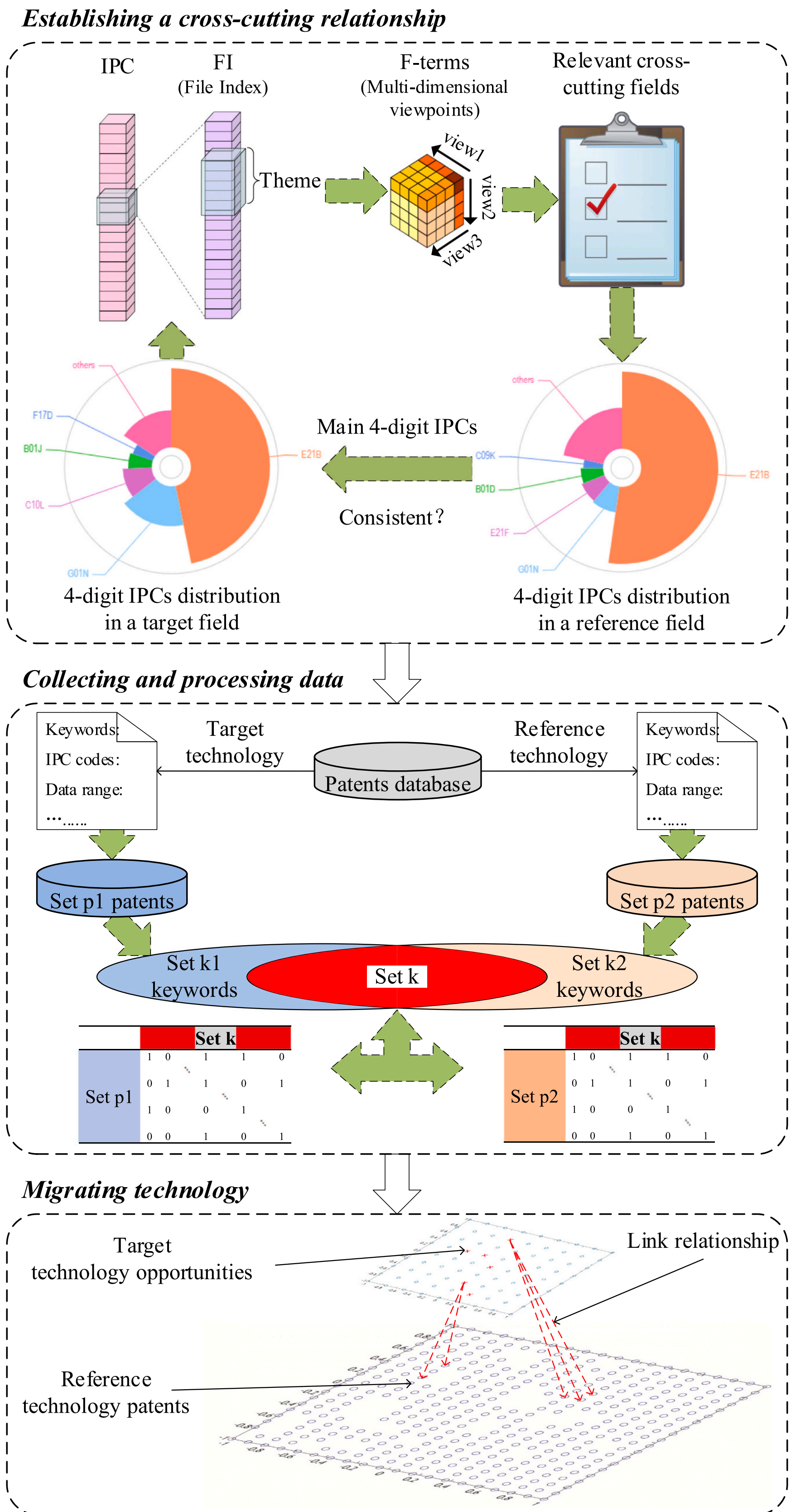


Fig. 1. Overview of the approach.

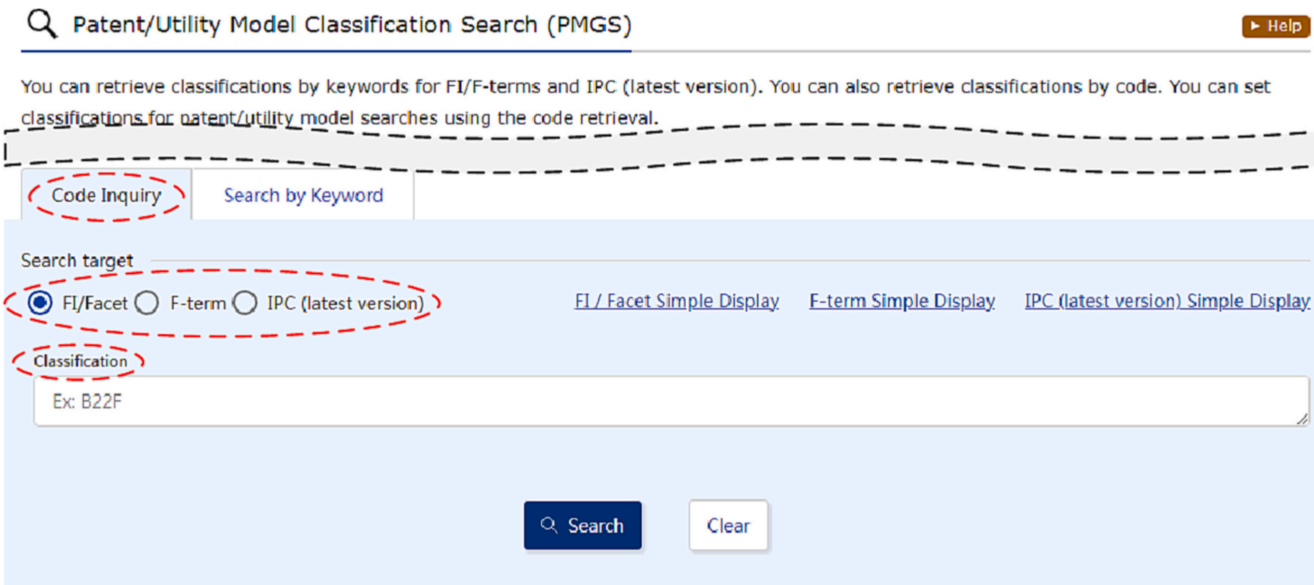


Fig. 2. Code inquiry in PMGS (https://www.j-platpat.inpit.go.jp/p1101).

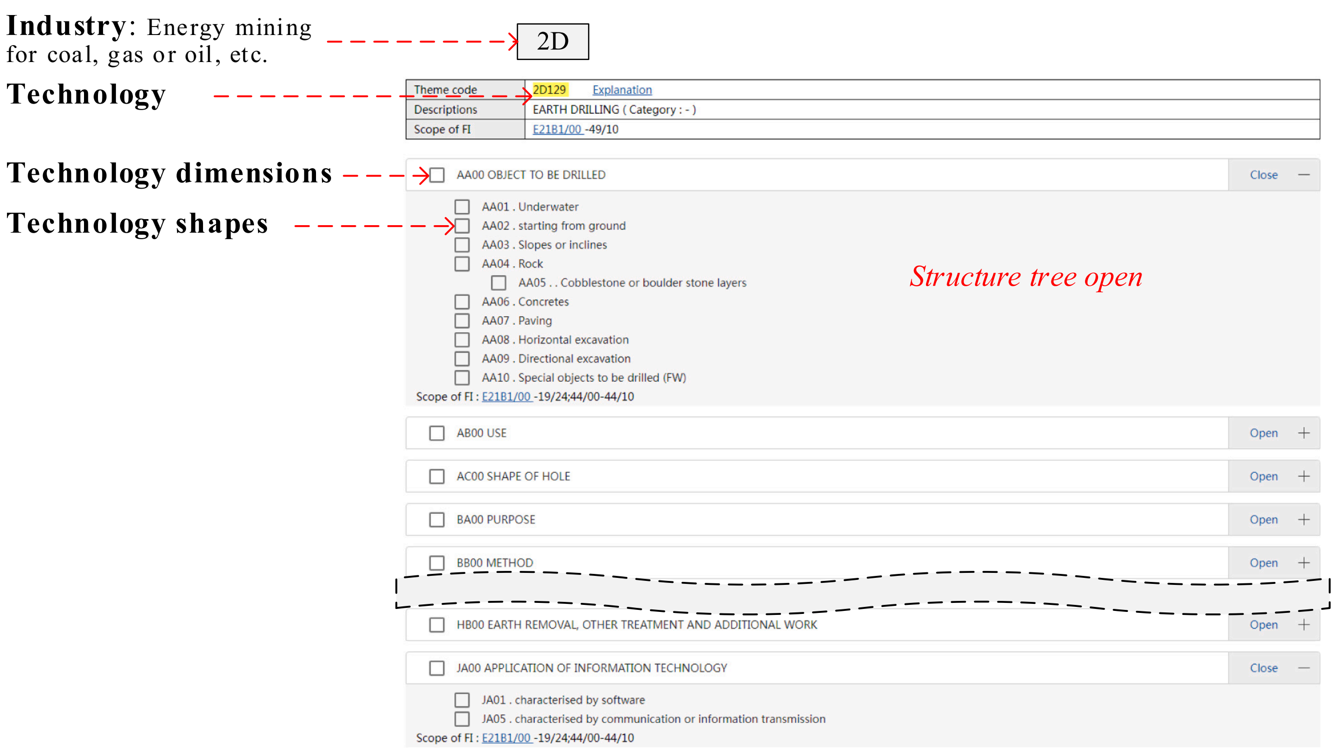


Fig. 3. Structure tree of the F-term.

the reference technology (Set k2) from the downloaded Set p1 and Set p2 patents, respectively. We then choose Abstract, which contains a summary of the patented invention, to extract keywords by stopping words, stemming, pruning, merging synonyms, keyword frequency index, and the Term Frequency — Inverse Document Frequency (TF-IDF) index. Secondly, we obtained keywords intersection (Set k) between Set k1 and Set k2. That is, keywords in Set k are chosen to reflect generic concepts of the target and reference technologies. The number of keywords is decided by domain experts. Third, the retrieved keywords in Set k for each patent of the target and reference technologies are converted into binary values, which are then used to constitute keyword vectors by means of text mining tools. Thus, we obtain the patent-keyword vector matrices of the target and reference technologies, respectively.

3.2.3. *Migrating technology*

Following the generation of the patent-keyword vector matrix which comprises many dimensions, this stage describes how the target technology opportunities are discovered based on visualization techniques and how corresponding ideas are migrated from the reference technology based on link prediction.

Step 3-1: Developing a GTM-based map.

We choose the GTM algorithm, because according to Son et al. (2012) and Feng et al. (2021), it enables objective patent mapping and detection of vacuum, by reducing the multi-dimensional vectors into a two-dimensional region. The GTM algorithm is employed to develop two patent-keyword maps corresponding to the target and reference technologies through the GTM-toolbox in the MATLAB software. By doing this, it enables us to map each patent in Set p1 and Set p2 on one grid on the two-dimensional maps. The number of grids in the GTM-based map is determined by experts' judgment and parameters adjustment in the GTM-toolbox.

Step 3-2: Identifying vacuums or non-vacuums.

Blank and occupied grids will appear on this map once the GTM-based map is generated. It is intuitive to recognize the blank grid to a patent vacuum. The occupied grid is a non-vacuum which contains several prior patents. For the target technology, to provide patent vacuums' original meaning, they are transformed into the original vectors by GTM-based inverse mapping. For the reference technology, because GTM algorithm maps these patents in the patent-keyword vector of the reference technology into corresponding occupied grids on the two-dimensional map (Bishop et al., 1998; Teng et al., 2021), each non-vacuum contains multiple existing patents. Therefore, non-vacuums are converted to the original vectors through inverse mapping. The vectors by GTM-based inverse mapping from both vacuums of the target technology and non-vacuums of the reference technology are used for the link prediction.

Step 3-3: Discovering target technology opportunities.

Based on the GTM-based inverse mapping, the original vectors of vacuums can be further converted into binary values (i.e., "1" or "0") according to the threshold. Where "1" means one keyword exists in the vacuum; otherwise, it does not exist. Thus, we derive several keywords contained in the vacuum. Previous studies have shown that it is feasible and practical to identify emerging technology opportunities based on patent vacuums on the GTM-based map (Son et al., 2012; Teng et al., 2021; Yoon and Magee, 2018). However, it is not sufficient to explain a technical opportunity with a few fragmented keywords, our study shows that the target technology opportunities can be generated into ideas by linking to existing patents and providing more detailed schemes from the reference field.

Step 3-4: Establishing the link prediction.

We used the aforementioned two GTM-based maps to recommend existing patents of the reference technology to develop creative ideas for the target technology. To achieve this goal, the link prediction is utilized by linking the two maps since link prediction can help to anticipate possible associations among current relationship. It can be achieved by calculating a link prediction index between one patent vacuum of the target technology and other patent non-vacuum of the reference technology. The higher the index value, the greater the likelihood that a link will be generated (Han et al., 2021). In this manner, the technology migration between the target and reference technologies based on the link prediction is achieved. This process consists of two parts: 1) establishing the connected relation between two non-vacuums (existing patents) belonging to two cross-cutting technologies; and 2) determining the link prediction index with the highest prediction accuracy between vacuums of the target technology and non-vacuums of the reference technology. For the first part, we use the cosine similarity to establish a connected relation between two unconnected non-vacuums (existing patents) from cross-cutting technologies. Using the inverse mapping function, the description of each grid on the GTM-based map, whether it is vacuum or not, can be expressed in the form of keyword vectors (Bishop et al., 1998). Then, the cosine similarity between keyword vectors of the patent non-vacuums is calculated. For instance, if the similarity value is greater than 0.5, two non-vacuums from the target and reference GTM-based maps are connected; otherwise, they are unconnected. Next, we split the training and test data randomly for link prediction. For the second part, we predict the link between GTM-based vacuums of the target technology and non-vacuums of the reference technology. We compare different link prediction indexes, which are explained in further detail in Appendix A, and consider using most precise link prediction methods suitable for this study. According to the most suitable link prediction index, a link between two technologies should be established on the two GTM-based maps if the prediction accuracy exceeds a predetermined threshold. The logical structure of link prediction between the target and reference technologies is shown in Fig. 4. This step can be completed in a Python environment with the cosine similarity and link prediction packages.

Step 3-5: Generating and evaluating creative ideas.

There are multiple existing patents hidden behind a linked non-vacuum. Each patent should be evaluated to narrow the recommendation according to the criteria of patent value. A high-value patent of the reference technology is evaluated as a useful potential for the target technology opportunity. Because the number of patents is not enough to reflect accurately the value of patents (Ernst and Omland, 2011), we adopt evaluation methods from Trappey et al. (2021) and Girgin Kalıp et al. (2022), which are a combination of five patent indicators: strength, quality, applicants, boundary, and ease of use. According to Li et al. (2020), the patent strength reflects the market coverage which is outside economic activities that link technical patent classification (e.g., IPC) into the industry sector. Ernst and Omland (2011) argued that the quality dimension is usually obtained from specific patent indications like citations. Trappey et al. (2021) proved that the number of both claims and applicants have positive impacts on patent values. The number of claims in a patent denotes the boundaries protected by law, while the number of applicants directly reflects the size of the company's R&D investments. As for ease of use (Yang et al., 2021), we invite two domain experts from research institutions and one patent examiner from the Intellectual Property Office of China to score this indicator according to the advantageous effects in a patent document. Finally, we therefore establish the valuation based on a radar map which contains above five indicators, such as patent citations, 4-digit IPCs, claims, applicants, and ease of use. Thus, for a technology opportunity, the most recommended patent should be the one with the largest area on the radar map, which are explained in further detail in Section 4.4.

4. Results

4.1. Background

Natural gas use has expanded rapidly during the last decade, accounting for almost one-third of total energy demand growth, more than any other fossil fuel, e.g., coal and oil. Its use in coming years is expected to keep growing rapidly since it contributes far less CO_2 and air pollutants than many of the other fuels (IEA, 2022). It is estimated that the amount of carbon in natural gas hydrate (NGH), which is widely distributed under the sea, over the world is twice the total amount of carbon in the fossil fuels on the earth (Li et al., 2016; Yang et al., 2019). Hence, NGH is considered as an alternative energy. However, till now, NGH is far from being commercially exploited due to the complicated gas production process that involves thermodynamics, kinetics, engineering technology, geology, etc. (Li et al., 2016). Since NGH is released underwater by decompression, it works essentially the same way as the exploitation of other natural gas, such as the coal bed methane (CBM), the shale gas, and the tight gas. Related studies have also shown that the

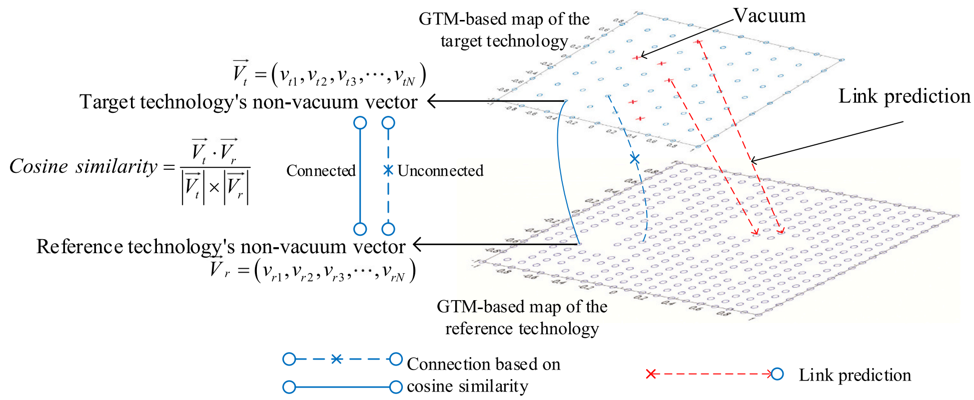


Fig. 4. Logical structure of link prediction between the target and reference technologies.

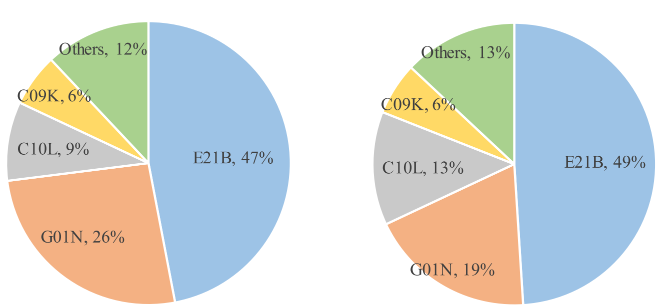
development of hydrate resources is thought to adhere the technical development roadmap of CBM or other resources (Englezos and Lee, 2005; Hyndman and Dallimore, 2001). Consequently, we select NGH as the target field to migrate technology from other reference fields.

4.2. *Cross-cutting relationship between NGH and CBM exploitation technologies*

We use search strategies of “Automatic recognition = natural gas hydrate” in China National Intellectual Property Administration (CNIPA) and “Front page = natural gas hydrate” in World Intellectual Property Organization (WIPO), respectively. The search was conducted in May 2022. We extracted 4-digit IPCs and exhibit the top five of them in Fig. 5. It shows that, in NGH field, the 4-digit IPC with the largest proportion is E21B, which means the energy exploitation technology, such as earth or rock drilling, obtaining oil, gas, water, soluble or meltable materials or a slurry of minerals from wells. Hence, we selected the exploitation technology of NGH as the target technology.

Next, we find the F-term, 2D129 (earth drilling), corresponding to E21B in JPO. As for 2D129, there are 19 technical dimensions and 2–48 technical shapes in each dimension, as shown in Table 1. Then, we can conclude that the energy exploitation technology with 2D129 is treated as the combination of 19 technical dimensions. This process can be achieved with the help of domain experts using the morphology analysis and detailed descriptions of F-terms in Table 1.

Take the natural gas industry as an example, “AA01 + AB01” means drilling underwater to extract natural gas resources, i.e., the NGH in the ocean. Whereas “AA02 + AB01” can be interpreted as drilling from ground to coal bed and extracting natural gas resources, i.e., the CBM in the coal mine. Similarly, “AA04 + AB01” is suitable for obtaining the shale or tight gas by drilling from the rock. The three fields, such as NGH, CBM and the shale gas, have high potential for being a good reference in generating creative ideas. Finally, we choose the mature CBM field as the migration source of the target technology. We measure



(a) IPCs in CNIPA. (b) IPCs in WIPO.

Fig. 5. Technologies composition of the natural gas hydrate.

Table 1
Representative technical dimensions and shapes of 2D129.

Technical dimensions	AA-Object to be drilled	AB-Use	···
Technical shapes	AA01-Underwater	AB01-Oil or gas well drilling	
	AA02-Starting from ground	AB01-Water wells or hot springs	
	AA04-Rock	⋮	
	⋮	⋮	

technologies composition of CBM by searching patents in CNIPA and WIPO. The results display that E21B accounts for 53 % and 57 %, respectively. Thus, the consistency of the key technology of CBM and NGH is verified and technology migration can be realized.

4.3. *Collecting and processing patents for NGH and CBM*

According to Step 2-1, we decide the best search strategy in DII. As for the target technology, we create one search strategy: TS = (deep-water OR sea OR ocean OR offshore OR off-shore OR “off shore” OR subsea) AND TS = (“natural gas hydrate” OR “gas hydrate” OR “methane hydrate” OR “NGH” OR “combustible ice” OR “flammable ice”) AND IP = (E21B*). In regard to the reference technology, we create the other search strategy: TS = (coalbed OR coalseam OR coal-bed OR coal-seam OR “coal bed” OR “coal seam”) AND TS = (gas OR methane) AND IP = (E21B*). The search was conducted in May 2022. Thus, we collected the patents Set p1 for NGH and Set p2 for CBM. On 31 December 2021, we obtained 409 patents in Set p1 and 3550 patents in Set p2.

Following Step 2-2, we extracted the keywords from the downloaded Set p1 and Set p2 through keyword frequency and TF-IDF weight. In the data cleaning process, except for stopping words, stemming and lemmatization, we also built a customized dictionary based on WordNet and technical terminologies. For example, we replaced technical terminologies “carbon dioxide”, “carbon-dioxide”, and “carbon di-oxide” with “co2”. Then, we extracted the top 20 keywords in each patent according the keyword frequency and TF-IDF weight. From this, we obtained the keywords Set k1 and Set k2 of the target technology and the reference technology, respectively. Thus, there are 2403 keywords related to NGH in Set k1 and 6928 keywords in Set k2. Since the keywords of two fields are not one-to-one corresponding, we extracted the intersection of two sets as Set k which contained 1968 common keywords. As a result, we selected technical 546 keywords to construct the patent-keyword vectors for the exploitation technology of NGH and CBM, as shown in Table 2. These 546 keywords included all words with the frequency greater than or equal to 3 and some technical words with the frequency less than 3. The above process required domain experts' participation to improve the accuracy and robustness of keywords extraction. This

Table 2
Partial patent-keyword vectors for the exploitation technology of NGH and CBM.

Common keywords	*Device*	*Connect*	*Pump*	⋯	*Heat*	*Waterproof*	⋯	*Vibrate*	*Evaporate*
Patent Set p1	1	1	0		0	1		0	1
	0	1	1		0	0		0	0
	⋮			⋮					⋮
	1	0	0	⋯		0	⋯	0	1
Patent Set p2	0	1	0	⋯	1	1	⋯	0	0
	1	0	0		0	1		1	1
	⋮								⋮
	0	1	1	⋯	1	0		0	1

process was performed through *python*.

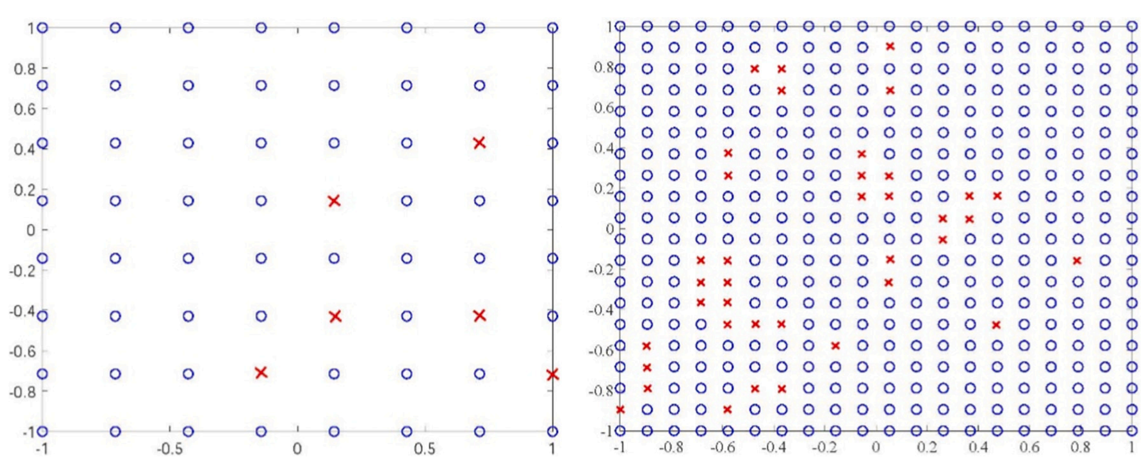
4.4. Migrating technology from CBM to NGH

According to Step 3-1, we developed two GTM-based maps corresponding to NGH and CBM in the MATLAB R2017b. To develop these maps, we set four model parameters, such as Gaussian basis functions, the width of the basis functions, the weight regularization factor, and the number of iterations. We decided the number of grids in the map based on the number of patents. Yoon and Magee (2018) have proved that the 12 * 12 GTM-based map is appropriate for visualizing medium-sized patents, but the number of patents might range from 200 to 1000. Because there was no empirical law for deciding the number of grids, we performed a sensitivity analysis from 5 * 5 to 25 * 25 grids through changing the value of model parameters. For the results, the 8 * 8 map is ideal to visualize 409 patents for NGH, whereas the 20 * 20 map fit 3550 patents for CBM. As shown in Fig. 6, two symbols ("○" and "×") represent whether a grid is occupied or blank, respectively. Thereinto, red "×" on blank grids indicate patent vacuums.

Fig. 6(a) exhibits that there exist six vacuums for the exploitation technology of NGH. We transformed six vacuums to original vectors with the help of GTM inverse mapping. The original vectors were expressed in the form of patent-keyword vectors with binary values (i.e., "1" or "0") by setting a threshold value 0.1. Consequently, we derive a set of keywords from above binary vectors to represent the description of each vacuum. Likewise, we also transform 58 non-vacuums of the GTM-based map of NGH in Fig. 6(a) and 362 non-vacuums of the GTM-based map of CBM to original vectors in Fig. 6(b). Since each grid on GTM-based map was numbered from top to bottom and then left to right, we label six vacuums as V31, V36, V38, V51, V54, and V63 in Fig. 6(a). This inverse mapping process of six vacuums are illustrated in Fig. 7. Similarly, we define 58 non-vacuums as NGH-N1, NGH-N2, etc. in Fig. 6(a) and 362 non-vacuums as CBM-N1, CBM-N2, etc. in Fig. 6(b). All original vectors are listed in Supplementary data.

In keeping with Step 3-4, we first calculate the cosine similarity between 58 NGH non-vacuums and 362 CBM non-vacuums, as shown in Supplementary data. We set the similarity value above 0.3, two non-vacuums are connected; otherwise, they are unconnected. We get 5462 links between 58 NGH non-vacuums and 362 CBM non-vacuums. Similarly, we calculate the cosine similarity between 58 NGH non-vacuums and 6 NGH vacuums and obtain 225 links. Then, we obtain total 5687 links with similarity greater than 0.3 in Supplementary data. We split the 5687 links and other unconnected relationships (58 * 362 + 58 * 6 − 5687 = 15,657) into training and test data randomly in an 8:2 ratio for link prediction. To predict the potential links between six NGH vacuums and 362 CBM non-vacuums, we compared nine link prediction indexes to provide the most appropriate method for migrating from CBM exploitation technologies to six technology opportunities in NGH field. For the comparison, we adopt AUC (area under the receiver operating characteristic curve) that are commonly used in measuring the accuracy of link prediction. Using AUC index, nine link prediction methods are compared to determine their effectiveness in predicting the potential links between six NGH vacuums and 362 CBM non-vacuums. Table 2 shows the results of prediction accuracy of nine link prediction indices. The local path (LP) has the highest AUC with 0.977127 among all link prediction indices. The most accurate index and the most suitable link prediction method are then selected for predicting the potential links between six NGH vacuums and 362 CBM non-vacuums. Consequently, we utilize the local path index to recommend CBM existing patents to predict creative ideas for NGH.

The link prediction method is used to calculate the similarity between all nodes (362 CBM non-vacuums, 58 NGH non-vacuums and 6



o: Non-vacuums (existing patents). ×: Vacuums.

(a) GTM-based map of NGH. (b) GTM-based map of CBM.

Fig. 6. GTM-based map of two exploitation technologies.

Vacuum	gas	hydrate	device	natural	system	pipe	connect	mining	layer	heat	pump
V31	0.743606	0.478097	0.043748	1.334799	0.394088	0.090587	0.491891	1.036415	-0.24625	0.243939	0.215673
V36	1.122931	0.946766	-0.06538	0.5323	0.613394	0.16068	0.066326	-0.13082	-0.00539	0.501315	0.161948
V38	0.912577	0.685969	0.199554	0.8268	0.500907	-0.16359	0.115293	0.263359	0.298512	0.589304	0.068698
V51	0.925043	0.823825	0.282453	1.016921	0.113669	-0.05187	0.015893	-0.09399	0.219539	-0.19306	-0.10337
V54	1.059596	1.016434	-0.06904	1.046454	0.446883	0.061787	0.517975	0.590763	0.846534	0.473482	0.08511
V63	0.681787	0.391729	-0.13246	1.019337	1.190852	0.308045	-0.33961	-0.04027	-0.46584	0.339492	-0.29217

. . .

biological	interconnect	pressurizat	recycling	vibrate	disassembly	croprocess	converge	waterproof	hydrophilic	terranean
-0.00077	-0.02411662	-0.00077	-0.00343	-0.02071	-0.02347641	0.000739	0.008996	0.005130131	8.68E-05	-0.02807
0.001222	0.000976399	0.001222	-0.00073	-0.01385	-0.01451316	0.001087	0.015691	-0.000334065	0.003197537	0.011553
-0.00117	0.001079167	-0.00117	0.001545	-0.02746	-0.02163297	-0.0026	0.008804	-0.003216758	-0.001038946	0.044258
-0.00091	-0.009687022	-0.00091	-0.00517	0.005339	-0.00269274	0.001354	0.000918	-0.00082257	0.010168558	0.022487
0.001507	-0.001664586	0.001507	0.004889	0.008496	0.00406117	0.001506	-0.00068	0.002798835	0.001032107	0.014514
-0.0025	0.007036089	-0.0025	0.003221	-0.00971	-0.00647156	-0.00592	-0.0022	-0.004529723	-0.001245863	-0.02627

Vacuum	gas	hydrate	device	natural	system	pipe	connect	mining	layer	heat	pump
V31	1	1	0	1	1	0	1	1	1	1	1
V36	1	1	0	1	1	1	0	1	0	1	1
V38	1	1	1	1	1	1	1	1	1	1	0
V51	1	1	1	1	1	0	0	0	1	1	1
V54	1	1	0	1	1	0	1	1	1	1	0
V63	1	1	1	1	1	1	1	0	1	1	1

. . .

biological	interconnect	pressurizati	recycling	vibrate	disassembly	nicroprocess	converge	waterproof	hydrophilic	terranean
0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0

Fig. 7. Inverse mapping process of six NGH vacuums.

NGH vacuums). The algorithm input is all nodes and existing links through the cosine similarity, while the output is the similarity matrix between nodes calculated based on the local path index. In this matrix, if the similarity between nodes is 0, the possibility of technology migration between NGH vacuums and CBM non-vacuums is extremely low. If the similarity value is high, which not only indicates that the two nodes are connected, but also gives priority to verifying the connection in the future. Therefore, these high-similarity connections are further analyzed. In this way, existing patents hidden behind CBM non-vacuum are recommended for the corresponding NGH vacuum to generate ideas. Table 4 shows the recommended CBM non-vacuums for six NGH vacuums with maximum similarity.

As shown in Table 4, we focused on V36 to analyze the potential technology opportunity for NGH and the corresponding two non-vacuums (CBM-N260 and CBM-N240) as migrating ideas from CBM to NGH. We list a set of keywords in V36 for NGH technology opportunity and existing patents hidden behind CBM-N260 and CBM-N240 for CBM, as shown in Table 5.

Table 5 shows that there are eighteen existing patents (eleven in CBM-N260 and seven in CBM-N240) of CBM for interpreting the NGH technology opportunity. This technology opportunity with a set of fragmented keywords can be used to generate ideas with the help of detailed technical schemes in existing patents. To further narrow the patent recommendation, we evaluated each patent according to the criteria of patent value. Fig. 8 shows the radar maps of the patent portfolio for both CBM-N240 and CBM-N260, where five indicators represent patent citations, 4-digit IPCs, claims, applicants, and ease of use. The largest area on the radar map is the most recommended patent portfolio. It is worth noting that some patents have never been cited before, so they are not displayed on the radar map. As a result, four of the seven recommended patents in CBM-N240 non-vacuums and six of the eleven recommended patents in CBM-N260 non-vacuums were evaluated and distributed in Fig. 8. Thereinto, patent WO2019051561 occupies the largest area on the radar map of CBM-N240 and patent CN101806228 takes up the most space on the radar map of CBM-N260.

As for the NGH technology opportunity with the vacuum V36, it is difficult to infer a clear technical solution from a few fragmented keywords in Table 5. Therefore, we recommend patent WO2019051561 titled system and method for low pressure gas lift and patent CN101806228 in CBM-N260 titled CBM radio spectrum direction-finding early warning system according to Fig. 8. Since patent WO2019051561 mainly involves the artificial lift to enhance the production of CBM to the surface, which belongs to gas transportation technology rather than gas exploitation technology of CBM. Therefore, we will focus on and analyze patent CN101806228 which aims at warning methane outburst through monitoring three signals in the CBM-rich region: 1) the signal of the gas radio stable and unstable radiation temperature based on the modulation radiation receiving system of the radio astronomy; 2) the signal of the gas radio unstable and quick changeable radiation strength and morphological based on the total power radiation receiving system of the radio astronomy; and 3) the signal of the gas radio unstable and quick change radiation frequency spectrum peak value and frequency drift based on the spectral radiation receiving system of the radio astronomy (Huang and Yuan, 2010).

In NGH field, till now, it is far from being commercially exploited. The exploitation technologies focus on the gas prospect, the simulation experiment in the lab, and the field trial. Since there were inherent disadvantages for the lab simulation away from the spot and the high cost of the field trial, the gas prospection technology was critical to verify actual gas reserves and reduce production costs. There exist three geophysics methods for the gas prospection technology, such as seismic reflection, drill hole sampling, electromagnetic detection (Wei et al., 2022). The first two methods were in contact detection. Because they were subject to the hole site and the hole depth, it was difficult to dynamically measure gas reserves in real time. The lack of prior drilling information also limited the reliability of these two methods. Although the electromagnetic detection was non-contact detection, the single channel test based on the pulse counting was difficult to distinguish from the pulse interference, which made it hard to establish model to predict gas reserves.

To fill the gap of existing gas prospection technologies in NGH field, we observed keywords of the technology opportunity in Table 5 and found that many underlined keywords most associated with CN101806228 patent were broken up by the GTM-based reverse

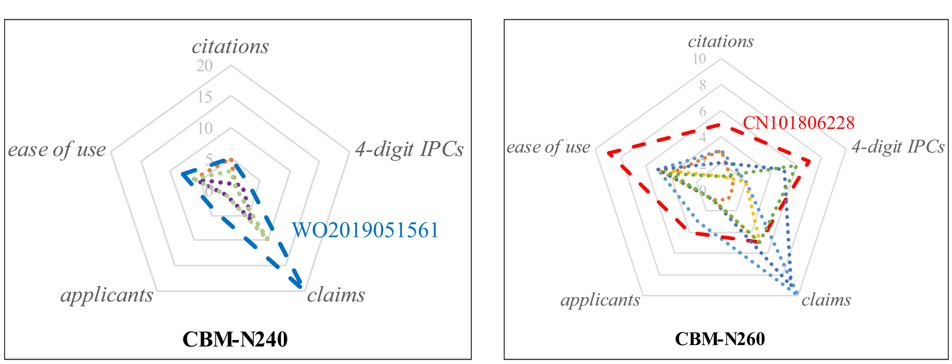


Fig. 8. The value comparisons of recommended patents in CBM-N240 and CBM-N260.

mapping. This shows although the combination of these keywords was one technology opportunity for NGH, they were difficult to be explained in detail, even with the help of domain experts. In contrast, as a solution for traditional gas prospection methods, CN101806228 patent provided a novel multi-channel and non-contact prospection method for the NGH exploitation. The descriptions in CN101806228 patent could well solve the problem of gas measurement in unexploited sea bed, such as gas orientation, gas pressure, gas concentration through the radio spectrum orientation prospection system with monitoring three signals in the NGH-rich region.

5. Discussion

The goal of TOA is to generate ideas for firms by providing more detailed schemes for problems and solutions in technical R&D. In response, this study attempts to make a clear innovation path between technology opportunities and ideas generation across domains by use of GTM and link prediction. At the TOA level, existing studies on GTM identified technology vacancies as technology opportunities including a few fragmented keywords (Feng et al., 2021; Son et al., 2012; Teng et al., 2021; Yoon and Magee, 2018). Our work is different from their studies. We established a cross-cutting relationship between the target and reference technologies through F-terms. Then, GTM maps of both target and reference technologies were generated. Target technology vacancies were composed of keywords but did not provide detailed schemes. At the idea generation level, we paved a path from technology opportunities to ideas generation across domains by use of the link prediction. Previous research on the link prediction aimed to establish relationships and discovered technology opportunities within a single field (Jeong and Yoon, 2013; Jeong et al., 2019; Yoon and Magee, 2018). Our proposed approach enables the idea generation process by providing cross-cutting patents, not domain-specific schemes. To establish the link relationship between two target and reference technologies, the cosine similarity was used to connect between non-vacuums of the target and reference technologies. The connection data was split into the training and test data for link prediction. Various link prediction indexes can be applied to predict potential links between GTM-based vacuums of the target technology and non-vacuums of the reference technology. We chose the local path index with the highest accuracy by comparing different link prediction methods in Table 3. The reasons for integrating the cosine similarity and link prediction are as follows. Since existing link prediction methods only predict potential linkages based on present nodes (Yoon and Magee, 2018). As a result, technology opportunities could only be discovered within a single field. To overcome this, this study first used the cosine similarity establish a connection between two unconnected non-vacuums from cross-cutting technologies, and then compared different link prediction methods and chose one advanced index.

(1) Clear path from technology opportunities to ideas generation across domains.

The results from our research clearly show the process from technology opportunities analysis to ideas generation. The results show that GTM was employed to identify six vacuums and discover technology opportunities for the NGH exploitation, as shown in Fig. 7. Regarding each NGH technology opportunity, CBM non-vacuums hiding many existing patents were recommended using the link relationship to generate specific ideas, as shown in Table 4. Thus, creative ideas of the NGH technology are migrated to cater for technology opportunities from recommended patents with the help of link prediction. The results in our study show that the link prediction can erase inconsistencies in the keywords and concepts shared between two fields, i.e., NGH and CBM, and provide alignment between technology opportunities and ideas generation. For example, the technology opportunity of the NGH prospection in Table 5 composed of many underlined keywords are mostly associated with CN101806228 patent. Traditional NGH prospection technologies include seismic reflection, drill hole sampling, electromagnetic detection (Wei et al., 2022). With the help of recommended patent (CN101806228) in CBM field, our research provides a novel multi-channel and non-contact prospection method for the NGH exploitation. Similarly, by providing other patents, which may not be applied to the NGH prospection technology, many other innovative options could be identified, facilitating creative innovation.

(2) Integration between automated techniques and expert-based methods among cross-domains.

Automated techniques are crucial for discovering technology opportunities with the combination of a set of knowledge elements, such as keywords and patent classification codes. Even though knowledge elements can be automatically derived and constructed with minimum experts' participation, they are generally coarse-grained in terms of generating ideas. There is a need for an approach with better and finer resolution. Although semantic features and the knowledge graph provide more detailed semantic correlation, but it is not able to generate ideas. Furthermore, semantic structures are prone to data sparsity problem because they focus on a limited number of the available words, which reduces the reliability of the technology forecasting (Han et al., 2021; Yang et al., 2017). As a contrast, our tailored approach starts by automatically identifying technology opportunities that contain multiple keywords as shown in Fig. 6(a) and Fig. 7. To avoid random keywords combination with coarse granularity, cross-cutting patents are recommended for these keywords (see Table 5 and Fig. 8). These patents can help experts to generate detailed technical schemes for the target technology. For example, compared to traditional single-channel and contact methods, CN101806228 patent provides a novel multi-channel and non-contact prospection method with the radio spectrum orientation prospection system for the NGH exploitation. This result combines fine granularity of ideas generation and automatic identification of technology opportunities. Hence, in this study, there exists a good trade-off between automating TOA processes and interpreting ideas, paving the way from technology opportunities to ideas generation.

(3) Detailed ideas for technology opportunities.

The suggested approach of integrating GTM and link prediction provides an effective means from discovering technology opportunities

Table 3
AUC values of link prediction indices.

Link prediction index	AUC
LP	0.977127
PA	0.942565
ACT	0.862431
Katz	0.8023
Cos+	0.737153
CN	0.184843
Jaccard	0.185271
Salton	0.184771
AA	0.185351

Table 4
Recommended CBM non-vacuums for six NGH vacuums with maximum similarity.

NGH vacuum	CBM non-vacuum	Vacuum	Non-vacuum	Vacuum	Non-vacuum
V31	CBM-N219 CBM-N240	V38	CBM-N260 CBM-N219	V54	CBM-N199
V36	CBM-N260 CBM-N240	V51	CBM-N199	V63	CBM-N260 CBM-N199

Table 5
Technology opportunity of NGH and existing patents of CBM.

Vacuum	Technology opportunity of NGH	Non-vacuums	Patents of CBM
V36	Gas, heat, system, drill, platform, liquid, formation, tank, bottom, seabed, discharge, valve, solid, flow, temperature, recover, supply, simulate, riser, phase, cavity, vessel, sediment, separate, excavate, locate, perform, wellhead, solar, hydrogen, mixture, path, prevent, construction, space, construct, carry, gasify, emit, floor, physics, resist, dynamic, ratio, return, explore, contact, instrument, track, potential, spraying, replace, leak, bury, automate, filling, compress, dimension, channel (excluding general terms, such as reduce, hydrate, natural, etc.)	CBM-N260	CN213360085, CN202991004, CN203081436, CN101806228, CN111927398, CN113062706, CN201851117, CN112901120, CN111396010, CN210858676, CN110857619.
		CBM-N240	CN206957684, CN107387045, WO2019051561, KR2018047764, CN107401395, CN112696177, WO2011088517

to generating novel ideas. This is helpful in terms of providing information that cannot be easily captured by qualitative judgment of several experts. Although keywords (Table 5) are identified as one technology opportunity by the GTM-based reverse mapping, they are so scattered that the technology opportunity composed of keywords are difficult to combine into a valuable technical clue. As a supplementary, the recommended patents with detailed descriptions by the link prediction provides a good reference for this technology opportunity. Except for the recommended patents WO2019051561 and CN101806228 in Fig. 8, other five NGH vacuums are corresponding to CBM non-vacuums, which can be evaluated and recommended existing patents. These existing patents in CBM field can provide many alternative ideas for NGH exploitation technology.

6. Conclusions, implications, and limitations

6.1. Conclusions

Although TOA with ideas generation have been recognized as important activities to help companies to remain competitive, there are limitations. This study proposed a new systematic approach to show the way from technology opportunities to ideas generation via cross-cutting patent analysis. For this approach, we started by selecting a target technology for which we wanted to seek novel opportunities, and established a cross-cutting relationship between the target and reference technologies through collecting F-terms is chosen. Then, we collected the relevant patents based on a certain search strategy and we processed them using text mining technique. Lastly, we discovered the target technology opportunities based on GTM and the corresponding ideas were migrated from the reference technology using the link prediction. The results from our study show that there exists a good trade-off between automating TOA processes and interpreting ideas. It is our belief that our proposed approach can be used to identify technology opportunities to ideas generation across domains. Creative ideas of the target technology are migrated to cater for technology opportunities from recommended patents in the reference field.

6.2. Implications

We belief that our approach has important implications for both academic and industries. From a theoretical perspective, this study expands the existing TOA research into generating creative ideas by providing more detailed schemes for technology opportunities. The suggested approach creates technology opportunities along with detailed ideas through cross-cutting recommended patents, in contrast to conventional TOA methods which employed only a few fragmented features to describe technology opportunities. The proposed approach also combines automatic identification of technology opportunities based on GTM and fine granularity of ideas generation based on the link prediction. The link prediction in our approach integrates the cosine similarity and traditional link prediction indexes. The cosine similarity establishes a connection between two unconnected domains and can make up for the disadvantage that existing link prediction methods only predict potential linkages based on present nodes.

In addition, there are also benefits from the practical perspective. Through TOA based on GTM and link prediction presented in this study, ideas generation by providing more detailed schemes is possible from the viewpoint of problems and solutions in technical R&D. In most of the existing studies, the technology opportunities were conducted at the keyword, topic, or patent classification code level. However, in this study, our approach takes TOA one step further to generate creative ideas by providing more detailed reference patents and provides a clear guidance path for R&D engineers. Besides, the idea generation process proposed in this study has great strength to be applied in the field because it can be implemented systematically. In previous studies, the process that is highly dependent on expert intervention has been proposed to derive technical solutions related to technology opportunities. However, in this study, a series of processes including data collection, preprocessing, and analysis are all automated processes. Especially, we established the migration relationship through F-term before GTM-based technology opportunities. Then, the connections were established between two cross-cutting technologies through the cosine similarity, and link prediction was used to recommend cross-domain existing patents for the above-mentioned technology opportunities to generating ideas. Compared to technology opportunities composed of fragmented keywords, our proposed approach helps to reduce R&D engineers' workload. Lastly, early identification of technology opportunities can narrow the scope of problem solving and is ideal for resource-constrained small and medium-sized enterprises.

6.3. Limitations and further research directions

Despite the many benefits of our approach, there are limitations. First, this study only selected the keywords intersection between two fields, ignoring some patents that do not contain these keywords in the reference field. It is our desire to establish link prediction for existing patents through patent classification codes or semantic methods in future. Second, the case study was conducted only in the exploitation technology of natural gas from cross-cutting domains (i.e., CBM and NGH). However, ideas generation for technology opportunities may cause patent infringement in the process of technology migration. This weakness can be remedied by performing the patent circumvention strategies, such as trimming and substituting, on a very large number of domains. Third, for the link prediction, we split the training and test data randomly in an 8:2 ratio. In future work, we will roll back the data according to time slice, like using data before 2015 for training and the data between 2016 and 2020 for testing. By doing it, we can investigate potential vacuums of the target technology before 2015 and how the vacuums are filled by recommended patents of the reference technology. Finally, the proposed approach focuses only on generating ideas for GTM-based technology opportunities. However, the proposed approach can also be helpful for other purposes, such as technology convergence and patent recommendation, and requires the engagement of R&D

engineers in the relevant field to validate in practice.

CRediT authorship contribution statement

Zhenfeng Liu: Conceptualization, Methodology, Project administration, Writing – original draft, Writing – review & editing. **Jian Feng:** Formal analysis, Resources. **Lorna Uden:** Writing – review & editing.

Declaration of competing interest

The authors declare that there is no conflict of interest regarding the publication of this paper.

Data availability

Data will be made available on request.

Acknowledgments

The authors gratefully acknowledge support from Shanghai Science and Technology Program (Project Nos. 22692113300, 20040501300).

Appendix A. Link prediction indices

Index name	Index definition	Index formula
LP	Local path	$(\boldsymbol{A}^2)_{xy} + \alpha \cdot (\boldsymbol{A}^3)_{xy}$
PA	Preferential attachment	$k_x \cdot k_y$
ACT	Average commute time	$\frac{1}{l^+_{xx} + l^+_{yy} - 2l^+_{xy}}$
Katz	–	$\sum_{l=1}^{\infty} \alpha^l \bullet \left\vert paths^{<l>}_{x,y} \right\vert$
Cos+	Random walk-based cosine similarity	$\frac{l^+_{xy}}{\sqrt{l^+_{xx} \bullet l^+_{yy}}}$
CN	Common neighbors	$\vert \Gamma(x) \cap \Gamma(y) \vert$
Jaccard	–	$\frac{\vert \Gamma(x) \cap \Gamma(y) \vert}{\vert \Gamma(x) \cup \Gamma(y) \vert}$
Salton (cosine similarity)	–	$\frac{\vert \Gamma(x) \cap \Gamma(y) \vert}{\sqrt{k_x \bullet k_y}}$
AA	Adamic-Adar	$\sum_{z \in \Gamma(x) \cap \Gamma(y)} \frac{1}{log k_z}$

Appendix B. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.techfore.2023.122565.

References

Albino, V., Ardito, L., Dangelico, R.M., Messeni Petruzzelli, A., 2014. Understanding the development trends of low-carbon energy technologies: a patent analysis. Appl. Energy 135, 836–854.

Ardito, L., D'Adda, D., Messeni Petruzzelli, A., 2018. Mapping innovation dynamics in the internet of things domain: evidence from patent analysis. Technol. Forecast. Soc. Chang. 136, 317–330.

Ardito, L., Petruzzelli, A.M., Albino, V., Garavelli, A.C., 2022. Unveiling the technological outcomes of microgravity research through patent analysis: implications for business and policy. IEEE Trans. Eng. Manag. 69, 3848–3859.

Bishop, C.M., Svensén, M., Williams, C.K.I., 1998. GTM: the generative topographic mapping. Neural Comput. 10, 215–234.

Chen, X., Jia, S., Xiang, Y., 2020. A review: knowledge reasoning over knowledge graph. Expert Syst. Appl. 141, 112948.

Christiansen, J.K., Gasparin, M., 2016. Managing controversies in the fuzzy front end. Creat. Innov. Manag. 25, 500–514.

Daim, T., Lai, K.K., Yalcin, H., Alsoubie, F., Kumar, V., 2020. Forecasting technological positioning through technology knowledge redundancy: patent citation analysis of IoT, cybersecurity, and blockchain. Technol. Forecast. Soc. Chang. 161, 120329.

Englezos, P., Lee, J.D., 2005. Gas hydrates: a cleaner source of energy and opportunity for innovative technologies. Korean J. Chem. Eng. 22, 671–681.

Ernst, H., Omland, N., 2011. The patent asset index – a new approach to benchmark patent portfolios. World Patent Inf. 33, 34–41.

Evangelista, A., Ardito, L., Boccaccio, A., Fiorentino, M., Messeni Petruzzelli, A., Uva, A. E., 2020. Unveiling the technological trends of augmented reality: a patent analysis. Comput. Ind. 118, 103221.

Feng, J., Liu, Z., Feng, L., 2021. Identifying opportunities for sustainable business models in manufacturing: application of patent analysis and generative topographic mapping. Sustain. Prod. Consum. 27, 509–522.

Feng, S., Magee, C.L., 2020. Technological development of key domains in electric vehicles: improvement rates, technology trajectories and key assignees. Appl. Energy 260, 114264.

Geum, Y., Park, Y., 2016. How to generate creative ideas for innovation: a hybrid approach of WordNet and morphological analysis. Technol. Forecast. Soc. Chang. 111, 176–187.

Girgin Kalıp, N., Erzurumlu, Y.Ö., Gün, N.A., 2022. Qualitative and quantitative patent valuation methods: a systematic literature review. World Patent Inf. 69, 102111.

Han, X., Zhu, D., Wang, X., Li, J., Qiao, Y., 2021. Technology opportunity analysis: combining SAO networks and link prediction. IEEE Trans. Eng. Manag. 68, 1288–1298.

Huang, F., Yuan, L., 2010. Coal seam gas radio spectrum orientation warning system. In: Nat Mine Gas Prevention & Control Eng Re (NAMI-Non-standard) Huainan Mine Group Co LTD (HUAI-C) Nat Mine Gas Prevention & Control Eng (NAMI-Non-standard) Nanjing Zihuai Mining Electronic Technology Co LTD (NANJ-Non-standard).

Hyndman, R.D., Dallimore, S.R., 2001. Natural gas hydrate studies in Canada. The Recorder 26, 11–20.

IEA, 2022. Gas. International Energy Agency. https://www.iea.org/fuels-and-technologies/gas.

Ittipanuvat, V., Fujita, K., Sakata, I., Kajikawa, Y., 2014. Finding linkage between technology and social issue: a literature based discovery approach. J. Eng. Technol. Manag. 32, 160–184.

Jeong, S., Yoon, B., 2013. A systemic approach to exploring an essential patent linking standard and patent maps: application of generative topographic mapping (GTM). Eng. Manag. J. 25, 48–57.

Jeong, Y., Park, I., Yoon, B., 2019. Identifying emerging research and business development (R&BD) areas based on topic modeling and visualization with intellectual property right data. Technol. Forecast. Soc. Chang. 146, 655–672.

Kashimoto, E., 2016. Find the most adequate FI and F-terms. World Patent Inf. 44, 24–33.

Kim, J., Geum, Y., 2021. How to develop data-driven technology roadmaps: the integration of topic modeling and link prediction. Technol. Forecast. Soc. Chang. 171, 120972.

Kim, J., Kim, S., Lee, C., 2019a. Anticipating technological convergence: link prediction using wikipedia hyperlinks. Technovation 79, 25–34.

Kim, K., Park, K., Lee, S., 2019b. Investigating technology opportunities: the use of SAOx analysis. Scientometrics 118, 45–70.
Kumar, V., Lai, K.-K., Chang, Y.-H., Bhatt, P.C., Su, F.-P., 2021. A structural analysis approach to identify technology innovation and evolution path: a case of m-payment technology ecosystem. J. Knowl. Manag. 25, 477–499.
Kwon, H., Park, Y., Geum, Y., 2018. Toward data-driven idea generation: application of wikipedia to morphological analysis. Technol. Forecast. Soc. Chang. 132, 56–80.
Lai, K.-K., Bhatt, P.C., Kumar, V., Chen, H.-C., Chang, Y.-H., Su, F.-P., 2021a. Identifying the impact of patent family on the patent trajectory: a case of thin film solar cells technological trajectories. J. Informetrics 15, 101143.
Lai, K.-K., Chang, Y.-H., Kumar, V., Wei, T.-Y., Su, F.-P., Mittal, A., 2021b. The position and role on patent citation network of the parking lot guidance system. Tech. Anal. Strat. Manag. 1–17.
Lai, K.-K., Chen, H.-C., Chang, Y.-H., Kumar, V., Bhatt, P.C., 2021c. A structured MPA approach to explore technological core competence, knowledge flow, and technology development through social network patentometrics. J. Knowl. Manag. 25, 402–432.
Lee, C., 2021. A review of data analytics in technological forecasting. Technol. Forecast. Soc. Chang. 166, 120646.
Lee, C., Kang, B., Shin, J., 2015. Novelty-focused patent mapping for technology opportunity analysis. Technol. Forecast. Soc. Chang. 90, 355–365.
Lee, C., Lee, G., 2019. Technology opportunity analysis based on recombinant search: patent landscape analysis for idea generation. Scientometrics 121, 603–632.
Lee, J., Ko, N., Yoon, J., Son, C., 2021. An approach for discovering firm-specific technology opportunities: application of link prediction to F-term networks. Technol. Forecast. Soc. Chang. 168, 120746.
Lei, C., Ruan, J., 2013. A novel link prediction algorithm for reconstructing protein–protein interaction networks by topological similarity. Bioinformatics 29, 355–364.
Li, S., Zhang, X., Xu, H., Fang, S., Garces, E., Daim, T., 2020. Measuring strategic technological strength : patent portfolio model. Technol. Forecast. Soc. Chang. 157, 120119.
Li, X.-S., Xu, C.-G., Zhang, Y., Ruan, X.-K., Li, G., Wang, Y., 2016. Investigation into gas production from natural gas hydrate: a review. Appl. Energy 172, 286–322.
Lü, L., Zhou, T., 2011. Link prediction in complex networks: a survey. Physica A 390, 1150–1170.
Ma, J., Abrams, N.F., Porter, A.L., Zhu, D., Farrell, D., 2019. Identifying translational indicators and technology opportunities for nanomedical research using tech mining: the case of gold nanostructures. Technol. Forecast. Soc. Chang. 146, 767–775.
McAdam, R., McClelland, J., 2002. Sources of new product ideas and creativity practices in the UK textile industry. Technovation 22, 113–121.
Moaniba, I.M., Su, H.-N., Lee, P.-C., 2018. Knowledge recombination and technological innovation: the important role of cross-disciplinary knowledge. Innovation 20, 326–352.
Nakamura, H., Suzuki, S., Sakata, I., Kajikawa, Y., 2015. Knowledge combination modeling: the measurement of knowledge similarity between different technological domains. Technol. Forecast. Soc. Chang. 94, 187–201.
Olsson, O., 2005. Technological opportunity and growth. J. Econ. Growth 10, 31–53.
Oltra-Garcia, R., 2012. Efficient situation specific and adaptive search strategies: training material for new patent searchers. World Patent Inf. 34, 54–61.
Park, Y., Yoon, J., 2017. Application technology opportunity discovery from technology portfolios: use of patent classification and collaborative filtering. Technol. Forecast. Soc. Chang. 118, 170–183.
Porter, A.L., Detampel, M.J., 1995. Technology opportunities analysis. Technol. Forecast. Soc. Chang. 49, 237–255.
Ren, H., Zhao, Y., 2021. Technology opportunity discovery based on constructing, evaluating, and searching knowledge networks. Technovation 101, 102196.
Son, C., Suh, Y., Jeon, J., Park, Y., 2012. Development of a GTM-based patent map for identifying patent vacuums. Expert Syst. Appl. 39, 2489–2500.
Song, K., Kim, K.S., Lee, S., 2017. Discovering new technology opportunities based on patents: text-mining and F-term analysis. Technovation 60–61, 1–14.
Teng, F., Sun, Y., Chen, F., Qin, A., Zhang, Q., 2021. Technology opportunity discovery of proton exchange membrane fuel cells based on generative topographic mapping. Technol. Forecast. Soc. Chang. 169, 120859.
Toubia, O., Netzer, O., 2017. Idea generation, creativity, and prototypicality. Mark. Sci. 36, 1–20.
Trappey, A.J.C., Trappey, C.V., Govindarajan, U.H., Sun, J.J.H., 2021. Patent value analysis using deep learning Models—The case of IoT Technology Mining for the Manufacturing Industry. IEEE Trans. Eng. Manag. 68, 1334–1346.
Wang, J., Chen, Y.-J., 2019. A novelty detection patent mining approach for analyzing technological opportunities. Adv. Eng. Inform. 42, 100941.
Wei, N., Bai, R., Zhou, S., Luo, P., Zhao, J., Zhang, Y., Xue, J., 2022. China's Deepwater gas hydrate development strategies under the goal of carbon peak. Nat. Gas Ind. 42, 156–165.
Yan, H.-B., Li, M., 2022. Consumer demand based recombinant search for idea generation. Technol. Forecast. Soc. Chang. 175, 121385.
Yang, C., Zhu, D., Wang, X., Zhang, Y., Zhang, G., Lu, J., 2017. Requirement-oriented core technological components' identification based on SAO analysis. Scientometrics 112, 1229–1248.
Yang, L., Liu, Y., Zhang, H., Xiao, B., Guo, X., Wei, R., Xu, L., Sun, L., Yu, B., Leng, S., Li, Y., 2019. The status of exploitation techniques of natural gas hydrate. Chin. J. Chem. Eng. 27, 2133–2147.
Yang, W., Cao, G., Peng, Q., Sun, Y., 2021. Effective radical innovations using integrated QFD and TRIZ. Comput. Ind. Eng. 162, 107716.
Yoon, B., Kim, S., Kim, S., Seol, H., 2021. Doc2vec-based link prediction approach using SAO structures: application to patent network. Scientometrics 127, 5385–5414.
Yoon, B., Magee, C.L., 2018. Exploring technology opportunities by visualizing patent information based on generative topographic mapping and link prediction. Technol. Forecast. Soc. Chang. 132, 105–117.
Yoon, J., Park, H., Seo, W., Lee, J.-M., Coh, B.-Y., Kim, J., 2015. Technology opportunity discovery (TOD) from existing technologies and products: a function-based TOD framework. Technol. Forecast. Soc. Chang. 100, 153–167.
Yu, X., Zhang, B., 2019. Obtaining advantages from technology revolution: a patent roadmap for competition analysis and strategy planning. Technol. Forecast. Soc. Chang. 145, 273–283.
Zhang, Y., Shang, L., Huang, L., Porter, A.L., Zhang, G., Lu, J., Zhu, D., 2016. A hybrid similarity measure method for patent portfolio analysis. J. Informetrics 10, 1108–1130.
Zhu, D., Porter, A.L., 2002. Automated extraction and visualization of information for technological intelligence and forecasting. Technol. Forecast. Soc. Chang. 69, 495–506.

Zhenfeng Liu is an Associate Professor of Technology Innovation and Management at the Shanghai Maritime University. Before completing his PhD in 2015, his research focuses on the new product development and R&D management with a special focus on innovation method and disruptive innovation. His work has appeared in technology opportunity analysis, patent analysis, and creativity and innovation management.

Jian Feng is a Lecturer in China Institute of FTZ Supply Chain at Shanghai Maritime University. Before completing his PhD in 2019, her research focuses on the supply chain innovation with a special focus on data analysis. Her research interests are machine learning and industrial engineering.

Lorna Uden is a Professor in Faculty of Computing and Digital Technologies of School of Computing at Staffordshire University. Her recent research focuses on knowledge engineering and knowledge graph.