ARTICLE IN PRESS

Technological Forecasting & Social Change xxx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Technological Forecasting & Social Change

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

Forecasting technology trends using text mining of the gaps between science and technology: The case of perovskite solar cell technology

Xin Li[a], Qianqian Xie[a], Tugrul Daim[b,c,d,*], Lucheng Huang[a]

[a] College of Economics and Management, Beijing University of Technology, Beijing, China
[b] Department of Engineering and Technology Management, Portland State University, Portland, OR, USA
[c] National Research University Higher School of Economics, Moscow, Russia
[d] Chaoyang University of Technology, Taichung, Taiwan

ARTICLE INFO

Keywords:
Technology trend
Technology forecasting
Text mining
Perovskite solar cell technology

ABSTRACT

How to detect and identify the future trends of emerging technologies as early as possible is crucial for government R&D strategic planning and enterprises' practices. To avoid the weakness of using only scientific papers or patents to study the development trends of emerging technologies, this paper proposes a framework that uses scientific papers and patents as data resources and integrates the text mining and expert judgment approaches to identify technology evolution paths and forecast technology development trends within the short term. The perovskite solar cell technology is selected as a case study. In this case, the text mining and expert judgment methods are applied to analyze the technology evolution path, and gaps analysis between science and technology is used to forecast the technology development trend. This paper will contribute to the technology forecasting and foresight methodology, and will be of interest to solar photovoltaic technology R&D experts.

1. Introduction

In recent years, we have witnessed the emergence of major advancements in new technologies that have disruptive characteristics, such as information technology, nanotechnology, biotechnology, and new material technology. The emergence and development of these new technologies not only changed the global competitive structure, but also created new industries, changed the lives of people, and changed the socio-economic production mode (Rifkin, 2011). Identifying the future trends of these new technologies as early as possible is crucial for governments' and enterprises' research and development (R&D) strategic planning to gain a first-mover advantage in future global competition. Many decision makers are aware of the significance of understanding of the emergence path, and identifying the future development trends of these new technologies for an organization's competitiveness and sustainable development when facing the wave of revolutionary technological changes (Li et al., 2015). Therefore, it becomes a strategic concern for public sectors and enterprises to identify and grasp the opportunity to develop their new technologies, which will ultimately contribute to their international competitiveness and sustainable development. This strategic issue raises one question: How can one detect and forecast the future development trend of these emerging technologies given the better understanding of their emergence? In response to this question, this paper develops a framework for detecting and forecasting the future development trend of these emerging technologies, based on an understanding of their existing evolution path and the identification of the gaps between science and technology.

A technology trend is considered as a continuously growing technology area with a certain pattern, and the pattern as a trend should have existed for a certain period of time (Ena et al., 2016). Many methods have been developed to identify and forecast the pattern. The traditional method to identify and forecast technology trends is usually based on the experience of experts, along with a long and costly procedure affected by subjective factors (Wang et al., 2015b). However, with the advancements in information and computer technology, the body of public technical literature including scientific papers and patents has grown, and researchers have begun to use this information to analyze and study technology trends. Thus, scientific papers (Daim et al., 2006; Dotsika and Watkins, 2017; Jaewoo and Woonsun, 2014; Kajikawa et al., 2008; Kostoff and Schaller, 2001; Rezaeian et al., 2017; Tsai, 2012) and patents (Chen et al., 2017; Golembiewski et al., 2015; Noh et al., 2016; Yoon and Kim, 2012) are applied as data resources for technology trends analysis. The amount of this technical information makes it more difficult to forecast technology trends solely based on expert knowledge (Kostoff, 1998). Therefore, based on public technical literature, technology trends forecasting activities usually use

* Corresponding author at: Department of Engineering and Technology Management, Portland State University, Portland, OR, USA.
E-mail addresses: ji2td@pdx.edu, xieqq@emails.bjut.edu.cn (T. Daim).

https://doi.org/10.1016/j.techfore.2019.01.012
Received 20 November 2017; Received in revised form 6 January 2019; Accepted 11 January 2019
0040-1625/

quantitative approaches to explore trends and provide early indications of potential changes and developments for anticipatory policy and strategy making. These approaches include bibliometric analysis (Daim et al., 2012; Dereli and Durmusoglu, 2009; Kajikawa et al., 2008; Lee et al., 2010), and text mining method (Chen et al., 2017; Choi and Hwang, 2014; Hao et al., 2014; Kostoff et al., 2008; Yoon and Kim, 2012). These quantitative methods can process massive raw data, reduce the size of data for further manual process, and mining the intelligence information of technology from technical literature. While experts' knowledge can provide powerful creditability to take responsibility for analyzing the intelligence information when we forecasting technology trends, so integrating the quantitative methods with experts' knowledge to study the development trends of technology and the potential changes of technology has gradually become a focus of researchers' attention.

The relationships between science and technology have long been debated in academia. Much of the empirical evidence since the mid-1980s supports the idea that science and technology co-evolve and interact in complex ways, and has replaced the old linear model in which the progress of science was essentially exogenous and technological advances were the outcomes of applied R&D efforts (Breschi and Catalini, 2010). Many researchers have noted in different empirical studies that the interdependencies and interactions between science and technology have been increasing. Despite this interdependence, gaps also exist between them (Wang et al., 2015a). Through the analysis and comparison of scientific papers and patents can be used to determine the gaps and identify technological opportunities (Shibata et al., 2010; Wang et al., 2015a). Many empirical studies have shown that an enormous amount of scientific information is organized and codified in patents, which shows that the development of an increasing number of technologies depends on science (Breschi and Catalini, 2010; McMillan et al., 2000; Narin and Noma, 1985; Tijssen, 2001). In knowledge-intensive science technology, basic research in science provides a fundamental basis for technology development, and the science is considered as seeds of technology and innovation in the linear innovation model (Shibata et al., 2010). Much information of technological development can be obtained from analysis of the development of science. Scientific papers and patents record modern and advanced knowledge in scientific discovery and technological development (Wang et al., 2015a). Therefore, how to discover the gaps between them through the mining of technical information contained in scientific papers and patents, and to combine experts' knowledge to predict the future development trend of technology is an important issue that scholars need to pay attention to. Some previous studies exploring technology trends are based on scientific papers or patents alone, whereas few scholars have combined information from both scientific papers and patents to study technology trends. Some scholars noted that different types of information sources provide diverse knowledge about the evolution path of technological development, and integrated use of the data sources will certainly give a more complete picture of the technology trends (Ena et al., 2016). Thus, the analysis of only scientific papers or patents is not sufficient to fully understand the evolution path of technological development and forecast technology trends.

To avoid the weakness of using only scientific papers or patents to study the development trends of emerging technologies, therefore, this paper proposes a framework that integrates scientific papers and patents as data resources. In the framework, topics-based text mining and experts' knowledge approaches are applied to mining the technical knowledge and information contained in scientific papers and patents respectively, and to identify the technology evolution path, and a gaps analysis between science and technology is used to forecast the technology future development trend within the short term. It takes the perovskite solar cell technology as a case study against the background of the rapid development of the photovoltaic technology field.

The rest of this paper is organized as follows. Section 2 briefly presents the literature review. Section 3 provides the proposed methodology. Section 4 analyzes the case study of perovskite solar cell technology. Section 5 discusses the paper, and Section 6 concludes the paper.

2. Literature review

Scientific papers represent a branch of literature on the research and description of scientific results in various academic fields, and offer not only a means of exploring the problems of academic research but also an instrument of academic exchange. As the latest technological information is frequently discussed and shared in scientific papers, the scientific paper is an important carrier of information about technology and an important data resource to study the development and change of technology (Behkami and Daim, 2012; Watts and Porter, 1997). Meanwhile, patents provide an up-to-date and reliable information source for reflecting technological development. Thus, patent documents are considered as a fruitful source of data for understanding technological trends (Noh et al., 2015). Thus, scientific papers (Daim et al., 2006; Dotsika and Watkins, 2017; Jaewoo and Woonsun, 2014; Kajikawa et al., 2008; Kostoff and Schaller, 2001; Rezaeian et al., 2017; Tsai, 2012) and patents (Chen et al., 2017; Golembiewski et al., 2015; Noh et al., 2016; Yoon and Kim, 2012) are applied as data resources for technology trends analysis. With the rapid increase in the number of scientific papers and patents, relying only on expert knowledge to quickly analyze technological trends is not effective (Kostoff, 1998). As the advancement of computing technologies, various quantitative and qualitative methods are developed to study technology development trends, such as citation analysis, co-word analysis, topic analysis, and technology roadmapping.

2.1. Citation analysis

Citation illustrates the exchange of key ideas within specific science field (Garcíalillo et al., 2016), it provides insight into the relationship between scientific papers or patents, the transfer, and interplay of knowledge (Teufel et al., 2006). Therefore, citation analysis is widely used for understanding the technological development trends.

Citation relations between scientific papers represent the knowledge-related relationships between them, and citation clustering can be used to cluster papers that share knowledge-related relationships. Through the analysis of citation clustering with time changes, the evolution path of technology and technology development trend can be clearly identified. Therefore, many scholars use citation analysis to study the evolution path of technology and technology development trend (Kajikawa and Takeda, 2008; Kajikawa and Takeda, 2009; Marzi et al., 2017; Shibata et al., 2008). Patent citation analysis assumes that patents cited by many later patents have a strong possibility of containing important ideas that later inventors build on (Yoon and Kim, 2011). A patent citation provides general relationships among patents that are related to each other, and it is suited to tracing the historical development of the main path in a given technology field (Chang et al., 2010). By analyzing the citation links among various patents over time, we can discover the similar and related technological topics in the cited-citing patents. Therefore, patent citation analysis has been widely used not only to track the history of technology development (Chen et al., 2012; Choi and Park, 2009; Gibson et al., 2017), but also to identify and monitor the new trends of technological development (Chang et al., 2010; Lee et al., 2010; Wu and Leu, 2013).

Scientific papers and patents contain much technical knowledge and information, such that solely relying on citation analysis to carry out technology trend analysis is not comprehensive (Madani and Weber, 2016). With the development of text mining methods and computer technology, scholars apply the text mining method to extract valuable knowledge and information from the textual content of the scientific papers for technology trend analysis (Callon et al., 1991; Dotsika and

Watkins, 2017; Jaewoo and Woonsun, 2014; Lee et al., 2008; Liu et al., 2013; Wu and Leu, 2014).

2.2. Co-word analysis

Text mining can give researchers access to technical information in patent documents (Madani and Weber, 2016), and offer a more complete picture of the technological development trend by integrating experts' knowledge. Co-word analysis is a text mining method that can extract technical information from the textual content, and reveal the deep meaning of documents (Dotsika and Watkins, 2017). The high-frequency keywords in the patent documents can reflect the documents' technical themes. The co-word analysis method can be used to explore themes and relations among these keywords (Ravikumar et al., 2015). Analyzing the characteristics of these themes over time is helpful to detect and forecast technological development trends. With the development of text mining methods and computer technology, many scholars used text mining method to extract the keywords in the publications, and the co-word analysis method was used to predict technological trends by analyzing the evolution of technical themes over time (Callon et al., 1991; Dotsika and Watkins, 2017; Jaewoo and Woonsun, 2014; Ravikumar et al., 2015).

Although co-word analysis has been widely applied to analyze the technology trends, keywords are usually too ambiguous to indicate a concept (Tseng et al., 2007), and *co*-word analysis cannot handle synonym and polysemy terms very well (Ding and Chen, 2014). Thus, the use of co-word analysis may not be sufficient to reveal the changes of the technical topics and technology trends. With the development of information and computer technology, topic analysis based text mining has rapidly developed. Compared with co-word analysis, topic analysis based text mining enables to handle synonym and polysemy terms, which can consider more textual information, thus it could significantly improve the performance of text mining.

2.3. Topic analysis

Topic analysis is an interesting text mining method that based on semantic clusters. Topic modeling method has become a popular approach for topic analysis, as it can handle synonym and polysemy terms very well, and identity hidden topics from documents. So many topic modeling methods have been developed for topic analysis. Latent Dirichlet Allocation (LDA) is one of topic analysis model for detecting latent topics from large size dataset (Wang et al., 2014). Chen et al. (2017) applied LDA to uncover the latent topics underlying massive patent claims, and identified the future trends of technology. While using LDA topic model, number of topics is usually given by researchers according to their subjective judgment (Wang et al., 2014), and it is hard to determine the number of topics without understanding the input data. Therefore, Teh et al. (2006) proposed the Hierarchical Dirichlet process (HDP) to address these problems. Ding and Chen (2014) compared the performance of HDP, co-citation analysis and co-word analysis, shows that HDP is more sensitive and reliable than other two methods in Topic Detection and Tracking. Furthermore, they demonstrated the topic evolution trends in the literature of terrorism research with the HDP method.

Technological topic analysis is one of the most important tasks for technology competitive intelligence analysis, since detecting technology trends, hot spots, and core technology are all based on it (Wang et al., 2014). Topic analysis plays an important role in exploring the development trend of technology. Some scholars began to use topic analysis based text mining to uncover the technical topics that are implicit in the documents to study technology trends (Wang et al., 2015b; Zhang et al., 2016). Lingo algorithm, a type of topics clustering algorithm, combines common phrase discovery and latent semantic indexing techniques to separate search results into meaningful groups, and it can identify topic labels automatically (Osinski and Weiss, 2005). Therefore, in this paper, we used the Lingo algorithm to generate technical topics from scientific papers and patents, and to identify the path of technological evolution of emerging technology.

2.4. Technology roadmapping

Technology roadmapping (TRM) provides a structured approach to map the evolution and development of technology (Phaal et al., 2004), and is recognized as an effective, comprehensive tool for planning the future development of technology (Phaal et al., 2011). Roadmapping provides a mechanism to help experts understand the emergence of science and technology, and forecast science and technology developments trend within targeted areas (Li et al., 2015). TRM has been employed to map historical technological evolution and emergence, and thus improve our understanding of its dynamics characteristics and gain insight into its future development trends (Zhang et al., 2015). The majority of traditional TRM studies applied qualitative approaches as research methods (Carvalho et al., 2013). With the development of text mining and informatics techniques, quantitative approaches are increasingly applied to TRM. Quantitative approaches are viewed as more valid and less biased as they are supported by objective data, and they can augment and amplify the capabilities of the expert by providing insights to the database structure and contents (Kostoff and Schaller, 2001). Therefore, researchers blend qualitative and quantitative methods to TRM (Hussain et al., 2017; Lee et al., 2009; Zhang et al., 2013; Zhang et al., 2015; Zhang et al., 2017), which can reduce knowledge gaps and improve strategic decision. The hybrid TRM that blends qualitative and quantitative methodologies is a trend in terms of Science, Technology & Innovation studies (Zhang et al., 2015). To better understand the detailed evolution path of emerging technology, in this paper, we applied roadmapping approach and domain experts' knowledge to construct a map of its evolution based on the topic clustering results obtained from scientific papers and patents.

2.5. Comparative analysis of related work

Through the above literature analysis, scholars usually take scientific papers or patents as a data source and apply citation analysis, co-word analysis, topic analysis and TRM to analyze the technology trends. However, these analysis methods have some shortcomings and limitations in the analysis of technology trends. Citation analysis uses only the structural information but do not involve textual content information. In addition, the use of citation analysis to analyze technology trends has other shortcomings, such as a degree of time delay. Although the co-word analysis method involves the textual content information, the *co*-word analysis method cannot handle synonym and polysemy terms very well thus it isn't an ideal tool for subsequent clustering tasks (Ding and Chen, 2014). Therefore, the co-word analysis method cannot effectively reflect the technical topics of documents. The topic analysis overcomes the limitation, and the topic analysis method has been used by scholars to analyze technology trends. Though quantitative approaches can automatically process massive raw data, experts' knowledge plays important roles in understanding of technology process. The proper use of quantitative method is to augment and amplify the capabilities of experts not to replace the experts (Kostoff and Schaller, 2001). So many researchers are working to combine quantitative methods with experts' knowledge to study technology trends. The combination of topics-based text mining and expert judgment has been considered a good way to predict technology trends. Although some scholars have combined information from both scientific papers and patents to determine the gap and identify technological opportunities (Shibata et al., 2010; Wang et al., 2015a), few scholars integrate text mining and expert judgment approach to forecast technology trends by identifying the gaps between science and technology. Therefore, this paper, taking scientific papers and patents as data sources, will use the topics-based text mining and expert judgment approaches to analyze the detailed evolution path of technology, and predict the technology trends by identifying the gaps between science and technology.

3. Methodology

Text mining can find implicit, previously unknown, and potentially useful patterns from large text documents, and such patterns can become important intelligence for decision-making after being supplemented with additional information and interpreted by experts' knowledge (Tseng et al., 2007). Thus, text mining has been considered as a practical technique to help the forecaster detects early signs of technological change. Many researchers have developed text mining approaches to mine technology intelligence in scientific papers and patents, and to catch weak signs of technology trends. Although existing approaches can reveal rough technology trends in the technical area, the detailed contents of the trends remain hidden (Chen et al., 2017). Topic-based clustering is an interesting text mining method that has been applied to analyze technology trends. Topics can be generated from scientific papers and patents using a clustering algorithm; however, most of the clustering algorithms used to generate topics are too broad to identify the refined technology trends. Thus, there is a need to adapt improved clustering algorithms that may go deeper into the text and generate more refined technological topics. Lingo algorithm, a type of clustering algorithm, combines common phrase discovery and latent semantic indexing techniques to separate search results into meaningful groups, and it has been shown to generate more refined topics compared with other clustering algorithms (Ena et al., 2016) and can identify topic labels automatically (Osinski and Weiss, 2005). Therefore, in this paper, we will use the Lingo algorithm to generate technical topics from scientific papers and patents.

This paper proposes a framework that uses scientific papers and patents as data resources and integrates topic-based text mining and expert judgment approaches to forecast technology trends by identifying the gaps between science and technology. The overall process of the framework is illustrated in Fig. 1. The detailed analysis steps of the framework are as follows:

Step 1. Retrieving and collecting the data. We take the Web of Science (WOS) and Derwent Innovations Index (DII) databases as the data sources for collecting data, and use different search queries related to the subject of the study (namely, perovskite solar cell technology) to download the relevant scientific papers and patents. The purpose of this step is to obtain a collection of scientific papers and patents related to perovskite solar cell technology.

Step 2. Preprocessing the obtained data. The scientific papers and patents obtained from Step 1 need to be divided by year, and the noise data needs to be removed. The cleaned scientific papers and patents are converted into text format compatible for text mining, and then the scientific papers dataset and the patents dataset related to perovskite solar cell technology are obtained.

Step 3. Clustering topics with the Lingo algorithm. In this step, we use the Lingo algorithm to generate technology topics from scientific papers and patents. Lingo algorithm uses the vector space model (VSM) and singular value decomposition (SVD) to find conceptually varied cluster labels and then assigns documents to the labels to form groups (Osinski and Weiss, 2005). The step is conducted using four specific tasks which are applied one after the other. The four detailed tasks are as follows:

(1) Preprocessing input data. We cleaned the preprocessed data obtained from step 2 by using a combination of three common text-preprocessing methods. These methods involve tokenization, stemming and marking stop words.
(2) Phrase extraction. The preprocessed data is used as an input for a phrase-extraction algorithm. A modified version of semantic hierarchical clustering algorithm was employed to discover phrases and single terms that could explain the verbal meaning of abstract concepts.
(3) Cluster-label induction. The phrases and single terms are utilized to construct term-document matrices which are able to show the conjoint appearance relationships between phrases and documents. To eliminate bias towards the query words, the logarithm term frequency-inverse document frequency (log TF-IDF) term-weighting scheme was used to calculate values in term-document matrix and then normalized each column vector to obtain matrix A. Then the SVD was utilized to decompose the matrix A to discover abstract concepts. Finally, the classic cosine distance was applied to calculate the cosine similarity between every abstract concept and phrase or term, and the maximum value indicates the term that best approximate the corresponding abstract concept (Osinski and Weiss, 2005). The maximum term was treated as cluster label.
(4) Cluster-content allocation. From what has been discussed above, the contents and cluster-labels have been determined, then we matched the input contents against cluster-labels. The classic cosine distance was applied to calculate the cosine similarity of the clustering label and the term-document matrix. If the similarity exceeds a predefined threshold, the corresponding documents are allocated to the corresponding clusters sets to generate the topics based on scientific papers and patents. The threshold ranges from 0 to 1, higher threshold values would result in more documents being put in one cluster, lowering threshold value leads to few documents in one group. In this paper, the threshold is 0.2 which verified by empirically exercise. The clustering results are sensitive to the parameter setting of the clustering algorithm. To get better clustering results, the selection of the control parameters for clustering algorithms was carried out by variation and subsequent expert validation of the results in relation to the following control parameters (Ena et al., 2016): Cluster size, cluster merging threshold, and size-score sorting ratio.

Step 4. Generating the hierarchical structure of the technology. After we obtained the clustering results from Step 3, we treated the clustering results as objective evidence for decision-making, and combined these results and domain experts' knowledge for topic analysis and the construction of the hierarchical structure of perovskite solar cell technology. The detailed steps are as follows. Firstly, perovskite solar cell technology domain experts who at least have focused on perovskite solar cell technology studies for more than 5 years were invited to screen the clustering results, and they divided the technical topics into three categories: material, structure and fabrication. The three categories form the first level of the hierarchical structure of the technology. Secondly, based on domain experts' knowledge, we found that there are interconnections between some topics in the clustering results. In order to refine the interconnections of the topics, the topics were rendered with a visualization tool which integrates into the Carrot 2 workbench. The visualization tool shows the connections among topics, and the topics were clustered based on their connections and harmonized by domain experts. The domain experts, comprising one senior researcher and two research assistants who has focused on solar cell studies, were invited to serve as our panel to construct the hierarchical structure of perovskite solar cell technology. The results of the topic harmonization form the second level and the third level of the hierarchical structure of the technology respectively. Finally, the hierarchical structure of technology was generated.

Step 5. Constructing the maps of the technological evolution based on scientific papers and patents. We invited the domain experts and one associated professor who has focused on TRM studies for more than 10 years to organize our TRM panel to construct the map of the evolution of perovskite solar cell technology. The vertical structure of the map is directly obtained from the first level of the hierarchical structure of the technology. The horizontal axis of the map indicates time. The detailed steps to construct the map are as follows. Firstly, based on the results of topics clustering obtained from Step 3, we obtained the

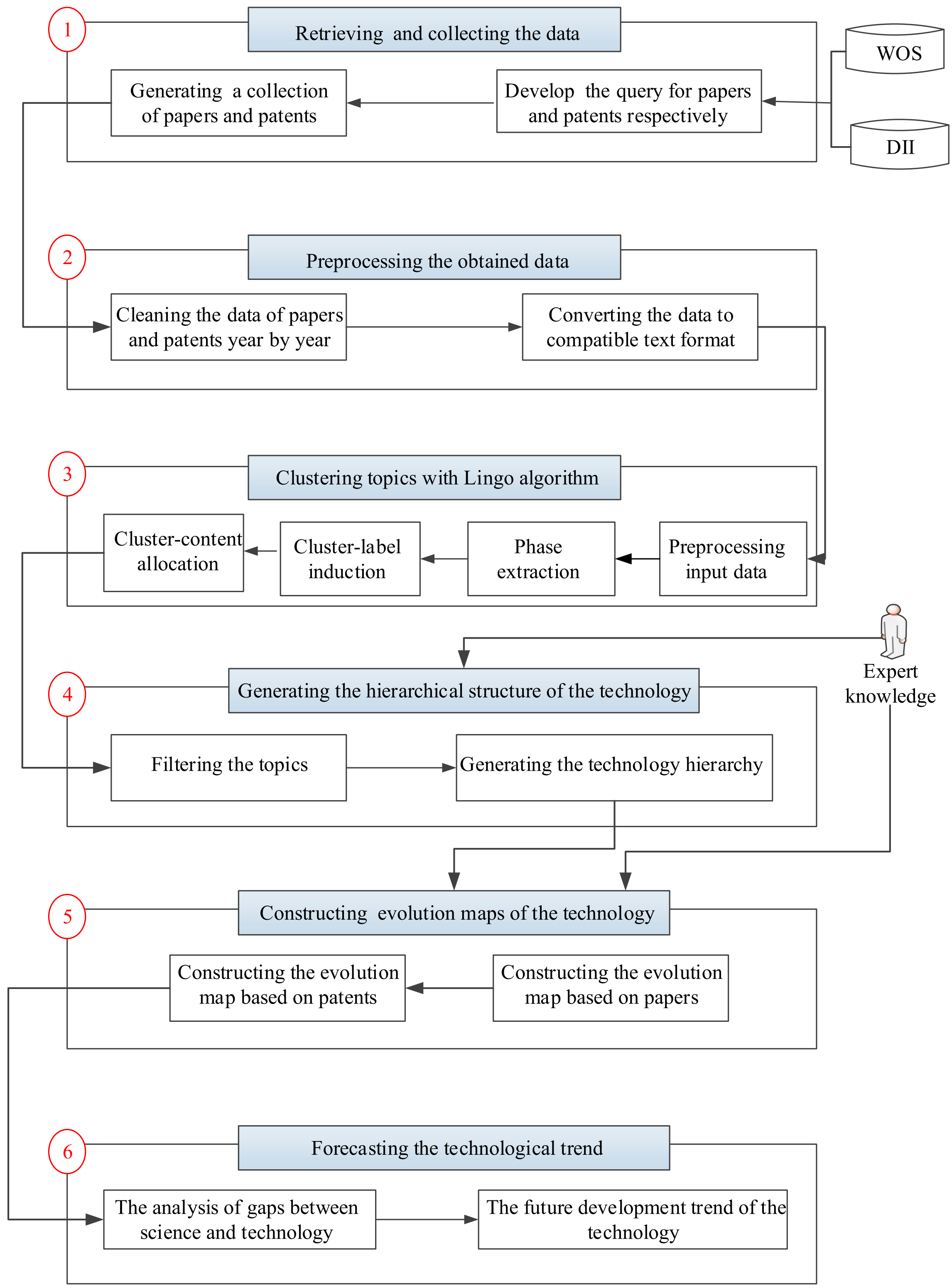


Fig. 1. Framework for forecasting the technology trends based on scientific papers and patents.

annual technical topics. Secondly, based on the hierarchical structure of the perovskite solar cell technology, we classified topics into appropriate layers and phases of TRM, and located them on the map. For example, we placed the annual topics representing different materials into the annual material layer, and the annual topics representing different structure into the annual structural layer. At the same time, the topics having the same attribute feature are merged in each layer according to the hierarchical structure of the perovskite solar cell technology. Finally, we constructed the evolution maps of perovskite solar cell technology based on the scientific papers and patents.

Step 6. Forecasting the technology trends. After we obtained the technical topics from scientific papers and patents, we carried out the differences analysis of the first appearance of the technical topics and the differences analysis of the technical topics with high growth rates between scientific papers and patents. We compared these two gaps and combined them with the detailed path of the technology evolution based on scientific papers and patents to forecast the future development trend of the technology.

4. Case study

To illustrate the proposed methodology, we take perovskite solar cell technology as a case study. Perovskite solar cell technology is one type of emerging solar cell technology in the solar photovoltaic technology field. With the rapid increase of photoelectric conversion efficiency, perovskite solar cell technology was selected as one of the ten most important scientific and technological advances in 2013 by the journal *Science*, and was assessed as one of the most anticipated breakthroughs in science and technology in 2014 by *Nature* magazine. In 2016, the World Economic Forum (WEF) listed the top ten emerging technologies that will change human life and included perovskite solar cell technology. The WEF believes that these emerging technologies are not out of reach and that they will have a significant impact on our lives in the near future.[1] Therefore, forecasting the future development of perovskite solar cell technology is of great significance for managers and policy makers' R&D strategic planning.

4.1. Data collection

In this paper, we take the Web of Science (WOS) and Derwent Innovations Index (DII) databases as the data sources for collecting data. This paper used the term "(Perovskite solar cell*) or (Perovskite based solar cell*) or (Perovskite* AND solar cell*) or (Perovskite* AND photovoltaic cell*)" as the query to search the scientific papers from WOS, and 3908 papers were retrieved from the database from 2009 to 2016. The term "((perovskite*) and ((solar cell*) or (solar cells*) or (photovoltaic cell*) or (photovoltaic cells*)))" was used as the query to search the patent data from DII, and 634 issued patents were retrieved from the database from 2009 to 2016. The search was done on March 20th, 2017. The annual numbers of scientific papers and patents related to perovskite solar cell technology are shown in Fig. 2.

As shown in Fig. 2, scientific papers and patents related to perovskite solar technology first appeared in 2009, and the number of scientific papers and patents grew rapidly each year as the technology's development entered a phase of sharp increment. In particular, the number of scientific papers and patents increased exponentially in the period from 2014 to 2016 (352% increase for papers and 487% increase for patents). These high rates show that the research and development of perovskite solar cell technology has been very active in the last three years. Furthermore, Fig. 2 indicates that the growth rate of patents has been higher than the growth rate of papers since 2015. This shows that the process of transforming scientific results into technological applications has been accelerating in the last two years.

4.2. Data analysis

This data analysis consists of four parts: topic clustering, construction of the hierarchical structure of the technology, analysis of the map of the technology's evolution based on scientific papers and patents, and forecasting of technology development trends.

4.2.1. Topic clustering

The annual data of the scientific papers and patents was processed individually, excluding the keywords "research progress", "development history" and "development status" contained in titles, and the reprocessed data were converted into XML format compatible with Carrot2 software. It is important to set the control parameters of the Lingo algorithm in the Carrot2 software to get better clustering results. The parameters were set to Minimum cluster size = 2, cluster merging threshold = 0.7, size-score sorting ratio = 1.0. The selection of the control parameters for clustering algorithms was carried out by experimental variation and subsequent expert validation of the results. After we got the clustering results, we selected the top 50% of topics according to the number of documents treating the topic, and removed multi meaning topics and meaningless topics, such as method and used, Layer for high, improvement of the power conversion. Then two domain experts, one of them who has focused on perovskite solar cell technology studies for more than 5 years from the School of Materials Science and Engineering, Tsinghua University, China, and the other one who has focused on perovskite solar cell technology studies for more than 7 years from College of Materials Science and Engineering, Beijing University of Technology, China, were invited to screen the rest topics in the clustering results based on their domain knowledge. Finally, we obtained annual topics based on scientific papers and patents with the help of two domain experts. Fig. 3 shows a part of the clustering results in 2015; the red circle indicates the candidate topics. Because the related data in the papers from 2009 to 2011 is too little, we only made extraction results from 2012 to 2016. The annual extraction results are shown in Tables 1 and 2. In Tables 1 and 2, the values in the brackets represent the number of documents treating the topic.

4.2.2. Hierarchical structure of the technology generation

A technology tree is a branching diagram that represents relationships among technologies (Choi et al., 2012). It provides a picture of the technology (Bildosola et al., 2017) to represent the relationships among product components, technologies, or functions of a technology in a specific technology area (Choi et al., 2012). The technology hierarchy can be utilized in selecting an interested technology area for in-depth analysis (Yoon and Park, 2005). Thus, it is important to construct the hierarchical structure of the technology to more fully understand the technology. In this paper, the purpose of constructing a hierarchical structure of perovskite solar cell technology is to provide the grounds for the classification of the topic clustering results, and to gain a more systematic and comprehensive understanding of the evolution path of perovskite solar cell technology. As mentioned in the Methodology section, we treated our cluster topics as objective evidence for decision-making, and combined quantitative and qualitative methods for topic analysis and the construction of the hierarchical structure of perovskite solar cell. Hence, we engaged experts on perovskite solar cell-related subjects for topic confirmation and classification, and we constructed the hierarchical structure of perovskite solar cell based on the cluster topics obtained from Section 4.2.1 with the help of domain experts. The two domain experts, comprising one senior researcher who has focused on solar cells studies for more than 10 years and two research assistants, from the College of Economics and Management, Beijing University of Technology, China, were invited to serve as our panel to construct the hierarchical structure of perovskite solar cell. First, the two domain experts previewed the topic clustering results, and based on their research experience and domain knowledge, they divided the technology topics into three categories, and named them. The three categories also fully reflect the current R&D directions of the perovskite solar cell technology. The three categories of the technology are considered: "Material", "Structure", and "Fabrication". The "Material" category addresses the material used in the technology; "Structure" addresses the technical structure; and "Fabrication" is related to the key technology used to solve challenges in producing the technology. This information makes it possible to identify the topic clustering results belonging to which category, and is useful for generating the roadmap, as it directly defines the vertical structure of the map of technology evolution. Then, on the basis of the three categories that have been divided, the two domain experts consolidated interrelated topics and named them by means of the most representative elements, and subdivided them to the

[1] The World Economic Forum's Meta-Council on Emerging Technologies. Top 10 emerging technologies of 2016. World Economic Forum, 2016.

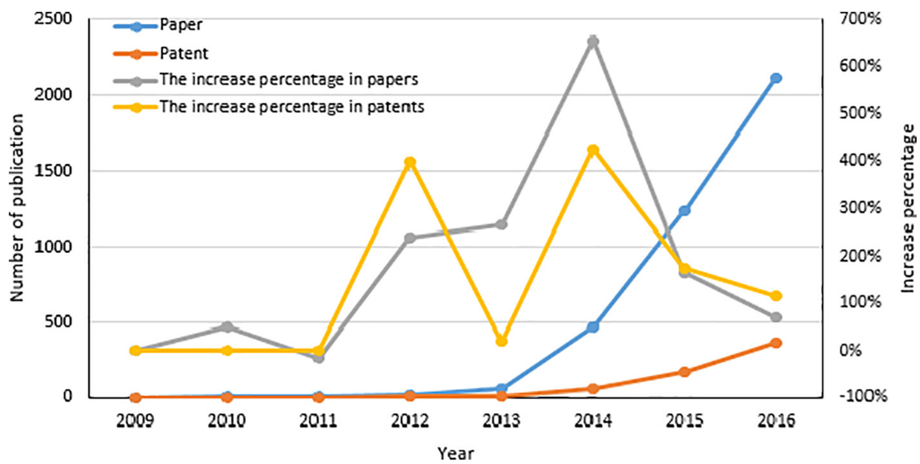


Fig. 2. Statistical results of scientific papers and patents related to perovskite solar cell technology.

corresponding categories and layers. Finally, the hierarchical structure of the perovskite solar cell technology (technology tree) was generated, as shown in Fig. 4.

In Fig. 4, the short name of the topic clustering is shown in Table 1 of Appendix I. Fig. 4 shows the hierarchical structure of perovskite solar cell technology. It can be observed that the first layer is made of material, structure, and fabrication. The second layer is a further detailed description of the first layer: The materials layer includes perovskite material, ETM, HTM, and perovskite layer material; the structural layer mainly includes perovskite crystals structure, mesoporous structure, perovskite heterojunction structure, and hole-conductor-free structure; the fabrication layer includes the film fabrication, crystal fabrication, and material fabrication. The third layer is a further description of the second layer, which includes the results of the topic clustering analysis and domain experts' naming process.

4.2.3. Analysis of the path of the technology evolution based on scientific papers and patents

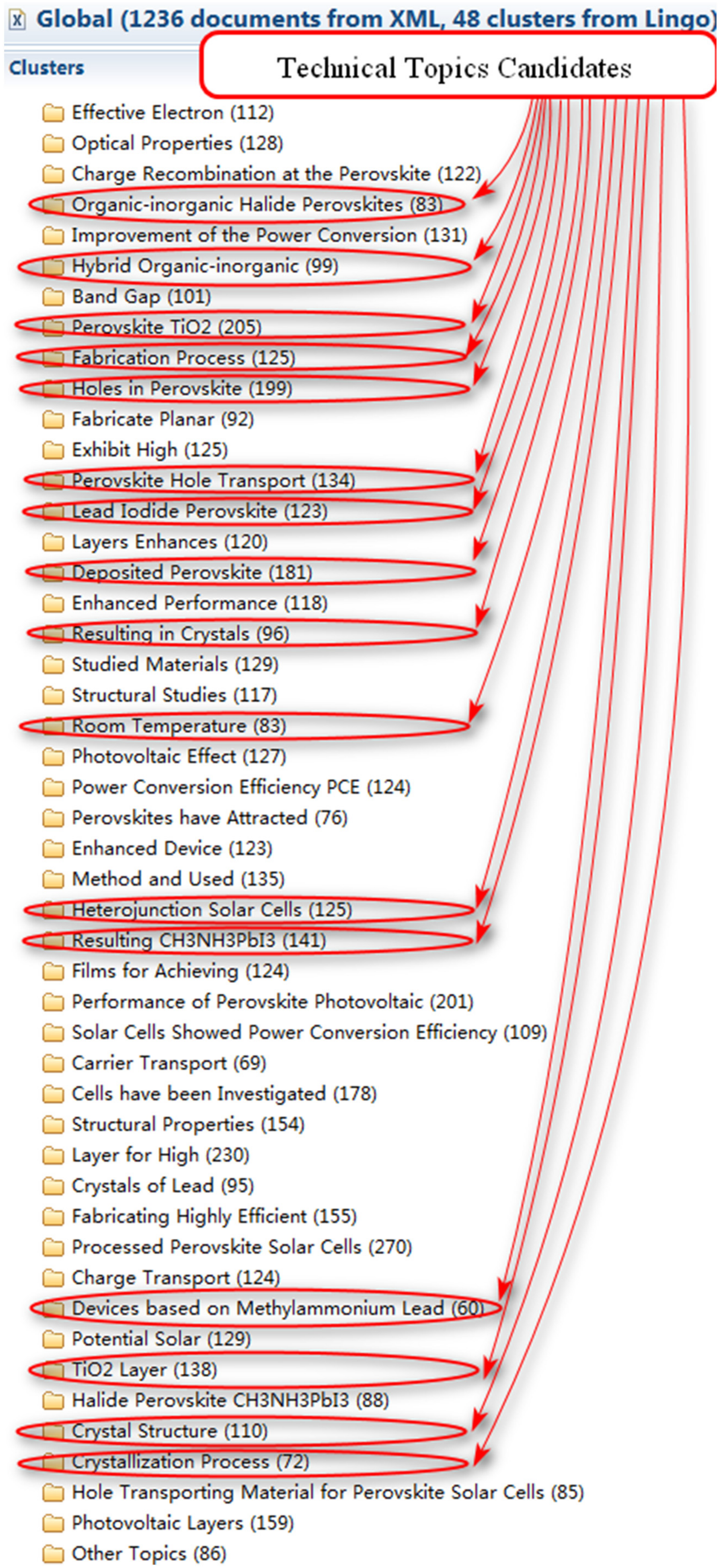
Forecasting technology development trends usually depend on fully understanding the path of the technology evolution. To understand the evolution path of perovskite solar cell technology and forecast its development trends for the short term based on the results of topic clustering, this paper constructed one map of the technology evolution based on the scientific papers and another based on patents. The clustering tools used in this paper can produce more refined clustering results than other clustering tools, and the selection of clustering topics has been assessed by experts' knowledge. In the case study, the topic clustering results obtained from scientific papers and patents are better able to respond to the technical topic, which we may use to build the map of the technology evolution. First, we invited the two domain experts to classify the topics in Tables 1 and 2 according to the hierarchical structure of the technology. The classification results are shown in Table 2 of Appendix I. Then, the maps of the technology evolution are constructed based on topic clustering results obtained from scientific papers and from patents. Finally, we explore the development trend of perovskite solar cell technology by analyzing the gaps between the map of the technology evolution based on scientific papers and the map based on patents.

4.2.3.1. Analysis of the path of the technology evolution based on scientific papers. With the help of the two domain experts and the one senior researcher of our expert panel, we extracted the candidate topics of the scientific papers based on the classification results (as shown in Table 2 in the Appendix I), and obtained the technical topics from the period 2012–2016, as shown in Table 3. In Table 3, the blue five-pointed star (★) represents the technical topics' first appearance. The number in the first brackets represents the number of documents containing the topic, and the number in the second brackets represents the percentage increase of the topic with respect to the previous year. The numbers in the second brackets are obtained by calculating the documents increment compared with the previous year in the first brackets.

As shown in Table 3, the technical topics in the scientific papers related to perovskite solar cell technology are perovskite materials and mesoporous structure in 2012. In 2013, many new technical topics appeared for the first time, such as ETM, HTM, perovskite layer material, absorbent layer structure, HTL structure, ETL structure, and film fabrication. In 2014, the technical topics which appeared for the first time are perovskite crystal structure, perovskite heterojunction structure, hole-conductor-free structure, crystal fabrication, and material fabrication. The emergence of the new technical topics show that the basic research of perovskite solar cells is no longer limited to study the perovskite material level but is also related to the structural and fabrication levels. It also shows that the basic research development and complexity of the technology are increasing. There are no new technical topics appeared in 2015 and 2016.

The percentage increase of the technical topics represents a strong growth potential. As shown in Table 3, the basic research related to perovskite solar cell technology at the material, structural, and fabrication levels is increasing from 2012 to 2016. Among them, at the material level, perovskite material showed a high growth percentage in 2013 and in 2014; ETM showed a high growth percentage in both 2014 and 2015; HTM showed a high growth percentage in 2014 and 2016, and perovskite layer material showed a high growth percentage in both 2014 and 2016. At the structural level, perovskite crystals structure showed a high growth percentage in both 2015 and 2016, and hole-conductor-free structure showed a high growth percentage in 2016. At the fabrication level, the film fabrication showed a high growth percentage in 2014, and the crystal fabrication showed a high growth percentage in 2015. A higher growth rate of technology topics is an important indicator for forecasting the future development potential of the technology (Bildosola et al., 2017). Their higher growth rates indicate that ETM, HTM, perovskite crystal structure, hole-conductor-free structure, film fabrication, and crystal fabrication have great development potential. They will be future research focuses of perovskite solar cell technology.

To better understand the detailed evolution path of perovskite solar cell technology, we constructed a map of its evolution based on the topic clustering results obtained from scientific papers and domain experts' knowledge. The two domain experts, comprising one associated professor who has focused on TRM studies for more than 10 years, from the School of Public Policy and Management, Tsinghua University, China, were invited to serve as our TRM panel to construct the map of



(caption on next page)

Fig. 3. Lingo algorithm clustering results (a part of clustering results of the scientific papers in 2015).

Table 1
Obtained results of the topics extracted from scientific papers.

2012	2013	2014	2015	2016
Mesoscopic TiO2 film (2)	Hole transport materials (9)	Deposited perovskite (76)	Perovskite Tio2 (205)	CH3NH3PbI3 perovskite films (269)
Oxides (2)	Absorber layers (8)	Perovskite crystal (67)	Holes in perovskite (199)	Perovskite Tio2 (267)
PbI3(2)	Mesoporous Tio2 (7)	Organic-inorganic perovskites (66)	Deposited perovskite (181)	Electron transport in perovskite (252)
	$CH_3NH_3PbI_3$ (4)	Perovskite planar (64)	Resulting $CH_3NH_3PbI_3$ (141)	Halide perovskite films (252)
	Electrode (3)	$CH_3NH_3PbI_3$ perovskite films (52)	Tio2 layer (138)	Crystal perovskite films (238)
	Flexible (3)	Perovskite heterojunction (59)	Perovskite hole transport (134)	Effective electron (237)
	Nanowires (3)	Methylammonium lead (53)	Fabrication process (125)	Transport material in perovskite solar cells (231)
	Solid-state mesoscopic solar cells (3)	Inorganic organic (37)	Heterojunction solar cells (125)	Charge transport (213)
	Atomic layer deposition (2)	Mesoporous Tio2 (32)	Crystal structure (110)	Crystal structure (211)
	Fullerene (2)	Halide perovskites (33)	Hybrid organic-inorganic (99)	Films prepared (210)
	Organolead halide perovskites (2)	Light harvester (30)	Resulting in crystals (96)	Hybrid structure (210)
	Spirobifluorene spiro-OMeTAD (2)	Room temperature (24)	Organic-inorganic halide perovskites (83)	Halide perovskite photovoltaic (203)
	Two-step deposition technique (2)	Inorganic hybrid perovskites (23)	Room temperature (83)	Crystallization leads (191)
	ZnO nanorod (2)	ZnO nanoparticles (5)	Crystallization process (72)	Material CH3NH3PbI3 (188)
		Mesoporous scaffold (15)	Lead iodide perovskite (123)	Processing temperature (160)
		Hole transport materials (50)	Devices based on methylammonium (60)	Halide CH3NH3PbI3 (153)
		Perovskite hole transport (100)		Tio2 as electron (150)
				Single crystals (15)
				CH3NH3PbI3 perovskite (213)

Table 2
Obtained results of the topics extracted from patents.

2012	2013	2014	2015	2016
Metal oxides (2)	Porous dielectric scaffold material (2)	Organic inorganic (8)	Film of solution (20)	Coating perovskite (75)
	Solid (2)	Light absorption material (7)	Hole or electron transport (20)	Perovskite absorber (61)
		Solid junction-type photoelectric (7)	Hybrid solar cell (20)	Metal oxide (59)
		Titanium oxide (7)	Organic-inorganic perovskite (16)	Inorganic solar cell (58)
		Precursor solution (6)	Zinc oxide (13)	Depositing perovskite (52)
		Carbon nanotube (5)	Titanium dioxide (12)	Halide perovskite (50)
		Scaffold (3)	Active material (11)	Perovskite crystal (46)
		Hybrid solar cell electronic (7)	Light absorber of solar cell (19)	Titanium dioxide (45)
			Derivative used (8)	Comprises hole transport material (41)
			P-type semiconductor (8)	Hybrid perovskite (40)
			Solid state (7)	Perovskite solar cells (32)
			Organic hole-transporting material (6)	Photoelectric conversion element (30)
			Planar heterojunction (6)	Comprising organic-inorganic (28)
			Photoactive layer (14)	Aode and cathode (26)
				Precursor solution (26)
				Heterojunction solar cell (21)
				Zinc (21)
				Derivative used (17)
				Removing solvent (14)
				P-type Semiconductor (10)
				Light-absorbing agent (5)
				Sputtering gas (2)
				Cation of formula (28)
				Charge mobility (7)
				Carrier unit (5)

the evolution of perovskite solar cell technology. Based on the hierarchical structure of the perovskite solar cell technology and the annual technical topics shown in Table 3, we classified topics into appropriate phases of TRM and located them on the map, and we finished the map of the evolution of perovskite solar cell technology based on scientific papers in Fig. 5. In Fig. 5, the vertical axis represents the material layer, the structural layer, and the fabrication layer of the perovskite solar cell technology obtained from the hierarchical structure's first level. The horizontal axis represents time. The different colors represent different technical topics. The hexagon represents the technical topics' names. The key elements in each layer, each year, and the different key components are located on the map with different colors. From these elements we are able to understand the development process of perovskite solar cell technology by analyzing the variation of elements in each layer over time.

As shown in Fig. 5, in the material level, there are four types of materials related to perovskite solar cell technology: perovskite material, ETM, HTM, and perovskite layer material. Technical topics such as

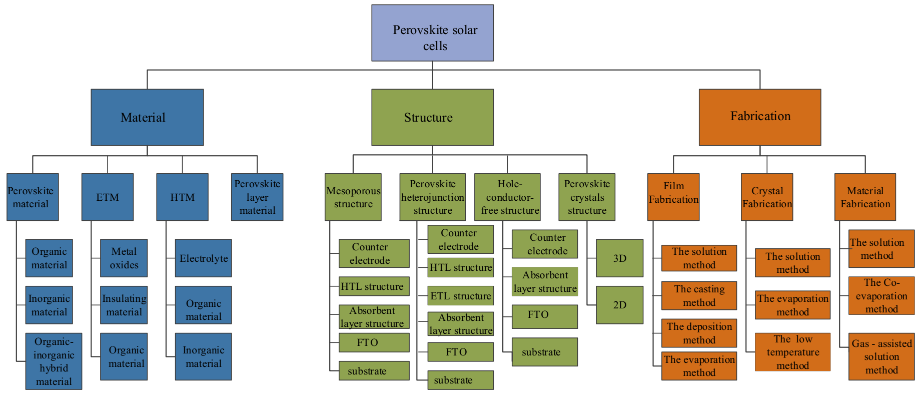


Fig. 4. Hierarchical structure of perovskite solar cell technology based on text mining and expert knowledge.

CH3NH3PbI3, organic-inorganic perovskites, halide perovskites, inorganic organic, hybrid organic-inorganic, halide perovskite materials, and inorganic hybrid perovskites have appeared with changes over time, which shows that perovskite material gradually changes from the initial organic materials in 2012 to organic-inorganic hybrid perovskite materials and inorganic perovskite materials after 2014. Organic-inorganic hybrid materials have better heat resistance than the corresponding organic semiconductor materials, which endows the perovskite solar cell with better device stability and longer life. Technical topics such as fullerene, ZnO, nanoparticles, and mesoporous TiO2 have appeared with changes over time, which shows that ETM application has changed from the fullerene application in 2013 to the insulating material applications such as Al2O3 and ZnO in 2014. Compared to fullerene, the photocurrent of the perovskite battery with ZnO as the electron transporting material can reach saturation more quickly and fully. Changing the ETM can improve the conductivity of ETM, thereby improving the power conversion efficiency of the perovskite solar cells. Technical topics such as spirobifluorene spiroOMeTAD, hole transport material, and transport material in perovskite solar cells appeared with changes over time, showing that the HTM of the perovskite solar cell changes from liquid electrolyte to solid organic hole transport material. The shifting of technical topics shows the details of the development process of perovskite solar cell technology material. We can see that researchers focused on perovskite materials in 2012 and began to study ETM, HTM, and perovskite layer materials after 2013. The high power conversion efficiency is not only determined by the properties of the perovskite material but also by ETM and HTM. Researchers try to improve the perovskite solar cells' performance and efficiency by developing different materials. We can see that ETM, HTM, and perovskite layer materials will receive more attention from researchers in the future, and the application of inorganic materials may be the future direction of materials research related to perovskite solar cell technology.

In the structure level, there are seven types of structures related to perovskite solar cell technology. The mesoporous structure appeared in 2012, and perovskite heterojunction and hole-conductor-free structure appeared in 2014. These results show that the structure of the perovskite solar cell has changed from the mesoporous structure in 2012 to the hole-conductor-free structure and the perovskite heterojunction structure in 2014. Currently, there are a variety of perovskite solar cell structures, and the coexistence of multiple structures will continue to be a trend of the technology in the short term. Compared to the mesoporous structure, the perovskite heterojunction structure and hole-conductor-free structure receive more researchers' attention, and the development of the new structure of perovskite solar cells would be a future research direction. The perovskite crystal appeared in 2014; crystal structure appeared in 2015; and crystal perovskite films and single crystals appeared in 2016, which show that researchers have committed to research of the perovskite crystal structure. The research trend gradually transitions from the three-dimensional perovskite structure to the two-dimensional perovskite structure. We predict that the two-dimensional perovskite structure may be the future research trend concerning perovskite crystals, and the perovskite solar cells' structure will be simplified and made more economical. The perovskite heterojunction structure, hole-conductor-free structure, and two-dimensional structure of the perovskite crystal will receive more research attention in the future.

In the fabrication level, there are three types of fabrication related to perovskite solar cell technology. The atomic layer deposition, two-step deposition technique appeared in 2013, which show that the film fabrication gradually develops from one-step solution preparation to two-step solution preparation to obtain perovskite films with uniform surface morphology, high film coverage, and strong crystallinity. Technical topics such as room temperature, crystallization process, and processing temperature have appeared with changes over time, which show that the crystal fabrication tends towards simplicity and ease of control.

4.2.3.2. Analysis of the path of the technology evolution based on patents. With the help of the two domain experts and the one senior researcher of our expert panel, we extracted the technical topics from the patents based on the classification results (as shown in Table 2 in Appendix I), and obtained the technical topics in the period 2012–2016, as shown in Table 4. In Table 4, the blue five-pointed star (★) represents the first time the technical topics appeared. The number in the first brackets represents the number of documents containing the topic, and the number in the second brackets represents the percentage increase of the topic with respect to the previous year. The numbers in the second brackets are obtained by calculating the documents increment compared with the previous year in the first brackets.

As shown in Table 4, the main technical topic in the patents related to perovskite solar cell technology is perovskite material in 2012. In 2013, two new technical topic appeared for the first time, namely, ETM and HTL structure. In 2014, many new technical topics appeared for the first time, including perovskite layer material, mesoporous structure, ETL structure, material fabrication, and film fabrication. In 2015, three new technical topics appeared for the first time, namely, HTM, absorbent layer structure, and perovskite heterojunction structure. In 2016, only one new technical topic appeared for the first time: perovskite crystal structure. From the emergence of the new technical topics in different years, we can see that the applied research moves from material to structure and then to fabrication, which shows that the research on the application of perovskite solar cell technology is accelerating. In addition to the research on all aspects of the technology, perovskite solar cell technology has the potential for commercialization in the future.

As shown in Table 4, the applied research related to perovskite solar cell technology at the material, structural, and fabrication levels is increasing from 2012 to 2016. Among them, perovskite materials have

Table 3
Technical topics of scientific papers in the period 2012–2016.

2012	2013	2014	2015	2016
★Perovskite material (2)	★HTM (11)	ETM (37)(1750)	ETM (343)(827)	ETM (498) (45)
★Mesoporous structure (2)	Perovskite material (6)(200)	HTM (50)(355)	Perovskite material (506)(92)	HTM (231)(362)
	★Perovskite layer material (4)	Perovskite material (264)(4300)	Perovskite crystal structure (206)(207)	Perovskite material (1026)(103)
	Mesoporous structure (9) (350)	Perovskite layer material (108)(2600)	Perovskite layer material (141)(31)	Perovskite layer material (823)(484)
	★Absorbent layer structure (8)	Mesoporous structure (52)(478)	HTL structure (199)(99)	Perovskite crystal structure (659)(220)
	★HTL structure (7)	★Hole-conductor-free structure (112)	Hole-conductor-free structure (141)(26)	Hole-conductor-free structure (823)(484)
	★ETL structure (8)	Absorbent layer structure (81)(913)	Perovskite heterojunction structure (125)(2)	HTL structure (296)(49)
	★Film fabrication(4)	HTL structure (100)(1329)	ETL structure (528)(205)	ETL structure (1849)(250)
	★ETM (2)	★Perovskite crystal structure (67)	Film fabrication (306)(115)	Film fabrication (210)(−31)
		★Perovskite heterojunction structure (123)	Crystal fabrication (155)(546)	Crystal fabrication (351)(126)
		ETL structure (173)(2063)	Material fabrication (182)(160)	Material fabrication (405)(123)
		Film fabrication (142)(3450)		
		★Crystal fabrication (24)		
		★Material fabrication (70)		

shown a high percentage of growth since 2014; ETM has a high growth percentage in both 2014 and 2016; HTL structure has a high growth percentage in 2015; and film fabrication has a high growth percentage in both 2015 and 2016. The high growth percentage indicates that these technical topics have great technological growth potential, which also means that these technical topics may be the applied research focuses related to perovskite solar cell technology in the future.

To better understand the detailed evolution path of perovskite solar cell technology, we constructed a map of its evolution based on the topic clustering results obtained from patents and domain experts' knowledge. Based on the hierarchical structure of the perovskite solar cell technology and the annual technical topics shown in Table 4, and with the help of our TRM panel, we classified topics into appropriate phases of TRM and located them on the map, and we finished the map of the evolution of perovskite solar cell technology based on patents in Fig. 6. In Fig. 6, the vertical axis represents the material, structural, and fabrication layers of the perovskite solar cell technology. The horizontal axis represents time. Different colors represent different technical topics. The hexagon represents the technical topics' name. The key elements in each layer, each year, and the key components are located in the map with different colors. From this we are able to understand the development process of perovskite solar cell technology by analyzing the variation of elements in each layer over time.

As shown in Fig. 6, in the material level, there are four types of materials related to perovskite solar cell technology. Technical topics such as organic-inorganic perovskite, inorganic solar cell, and comprising organic-inorganic have appeared with changes over time, which shows that the research of perovskite materials related to perovskite solar cell technology gradually developed from organic materials in 2012 to organic-inorganic hybrid materials in 2014 and then to inorganic materials in 2016. The performance of organic-inorganic hybrid materials is more stable than that of organic materials, and inorganic material is more economical than organic materials. From the shifting of the technical topics, we can see that the design and development of cheap, stable and environmentally friendly perovskite materials is the future research trend, and the inorganic materials will get more researchers' attention. Technical topics such as organic hole-transporting material, P-type semiconductor and P-type semiconductor comprises hole transport material have appeared with changes over time, which shows that HTM changes from organic material to comprised material. The technical topic titanium oxide appeared in 2014, and zinc oxide, derivative used appeared in 2015, which shows that ETM application has changed from the commonly used Tio2 material to ZnO and fullerene derivative application. This can improve the power conversion efficiency of the perovskite solar cells. The shifting of the technical topics concerning perovskite materials, ETM, and HTM indicates that materials that are cheap, stable, environmentally friendly, easy to prepare, and of improved performance represent future development trends of perovskite solar cells technology, and inorganic materials are a direction for future development.

In the structure level, there are six types of structures related to perovskite solar cell technology. The scaffold appeared in 2014, and perovskite heterojunction appeared in 2015. These results show that the structure of perovskite solar cells has transitioned from the mesoporous structure to the planar heterojunction structure. The planar heterojunction structure simplifies the structure and interface of the perovskite solar cell, which is to improve the power conversion efficiency. Perovskite crystal appeared in 2016, which indicates that researchers are beginning to focus on the perovskite crystal structure. A variety of perovskite solar cell structures indicates that the coexistence of multiple structures will continue to be the trend of the technology in the short term.

In the fabrication level, there are two types of fabrication related to perovskite solar cell technology. Precursor solution appeared in 2014, and coating perovskite, depositing perovskite, and sputtering gas appeared in 2016, which shows that the film fabrication develops from

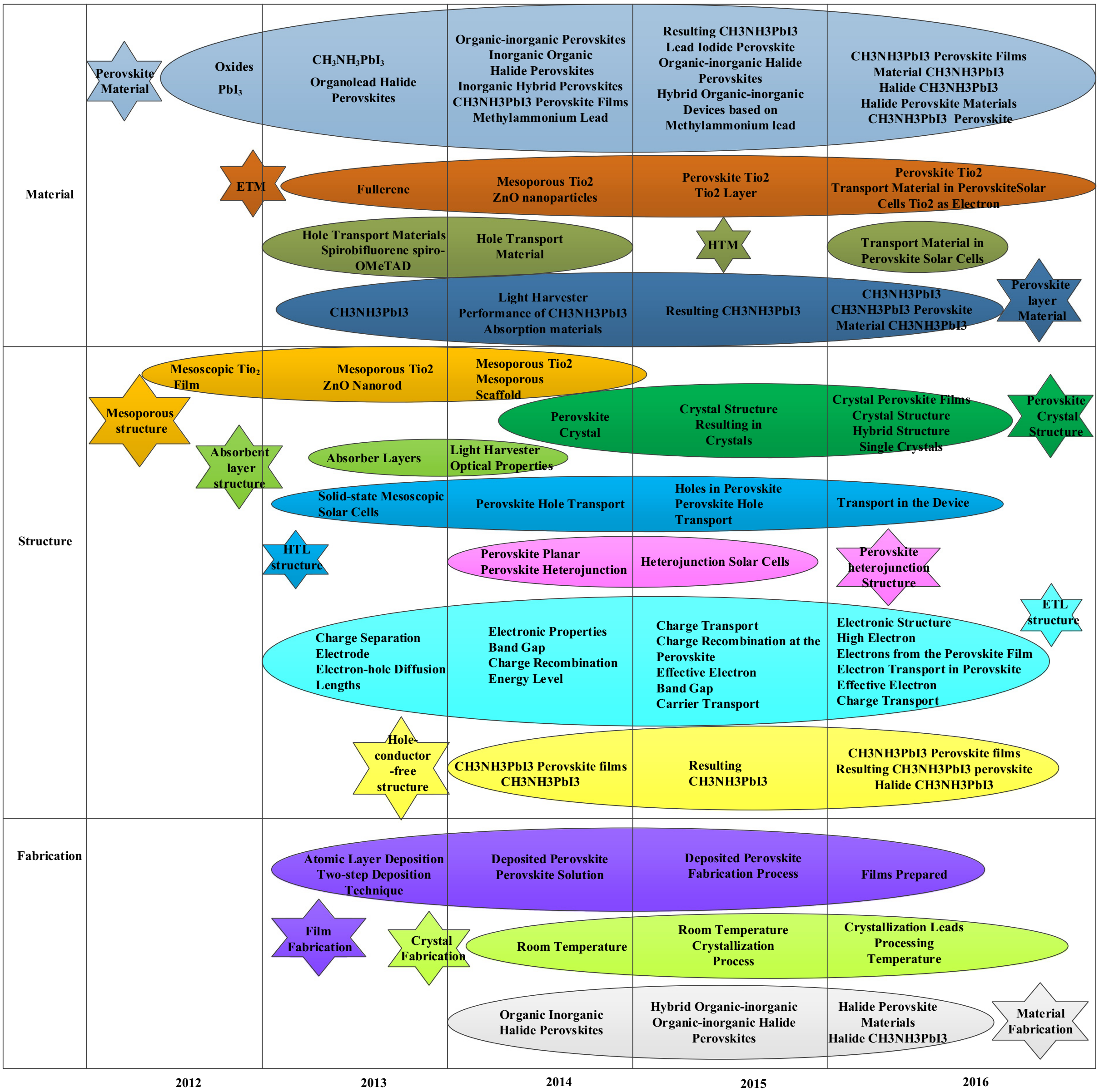


Fig. 5. Evolution map of perovskite solar cell technology based on scientific papers text mining.

the solution method to the coating method and the vapor deposition method. The coating method and vapor deposition method are designed to obtain perovskite films with high coverage, uniform surface, and strong crystallinity. Compared to the solution method, vapor deposition method and coating method have the advantages of being economic and simple, and are more conducive to the commercialization of the perovskite solar cell. Therefore, the fabrication development trend will involve a simple and economic preparation process for perovskite solar cell technology.

Table 4
Technical topics of patents in the period 2012–2016.

2012	2013	2014	2015	2016
★Perovskite material (2)	★ETM (2) ★HTL structure (2)	ETM (12)(500) ★Perovskite layer material (7) Perovskite material (8)(300) Mesoporous structure (3)★ ★ETL structure (7) ★Film fabrication (6) ★Material fabrication (8) HTL structure (7)(250)	ETM (33)(175) ★HTM (14) Perovskite layer material (30)(329) Perovskite material (36)(350) ETL structure (20)(186) HTL structure (27)(286) ★Absorbent layer structure (14) Film fabrication (20)(233) Material fabrication (16)(100) ★Perovskite heterojunction structure (6)	ETM (111)(236) HTM (10)(−29) Perovskite layer material (66)(120) Perovskite material (176)(388) ★Perovskite crystal structure (46) ETL structure (38)(90) Absorbent layer structure (58)(314) HTL structure (27)(0) Perovskite heterojunction structure (21)(250) Film fabrication (169)(745) Material fabrication (159)(894)

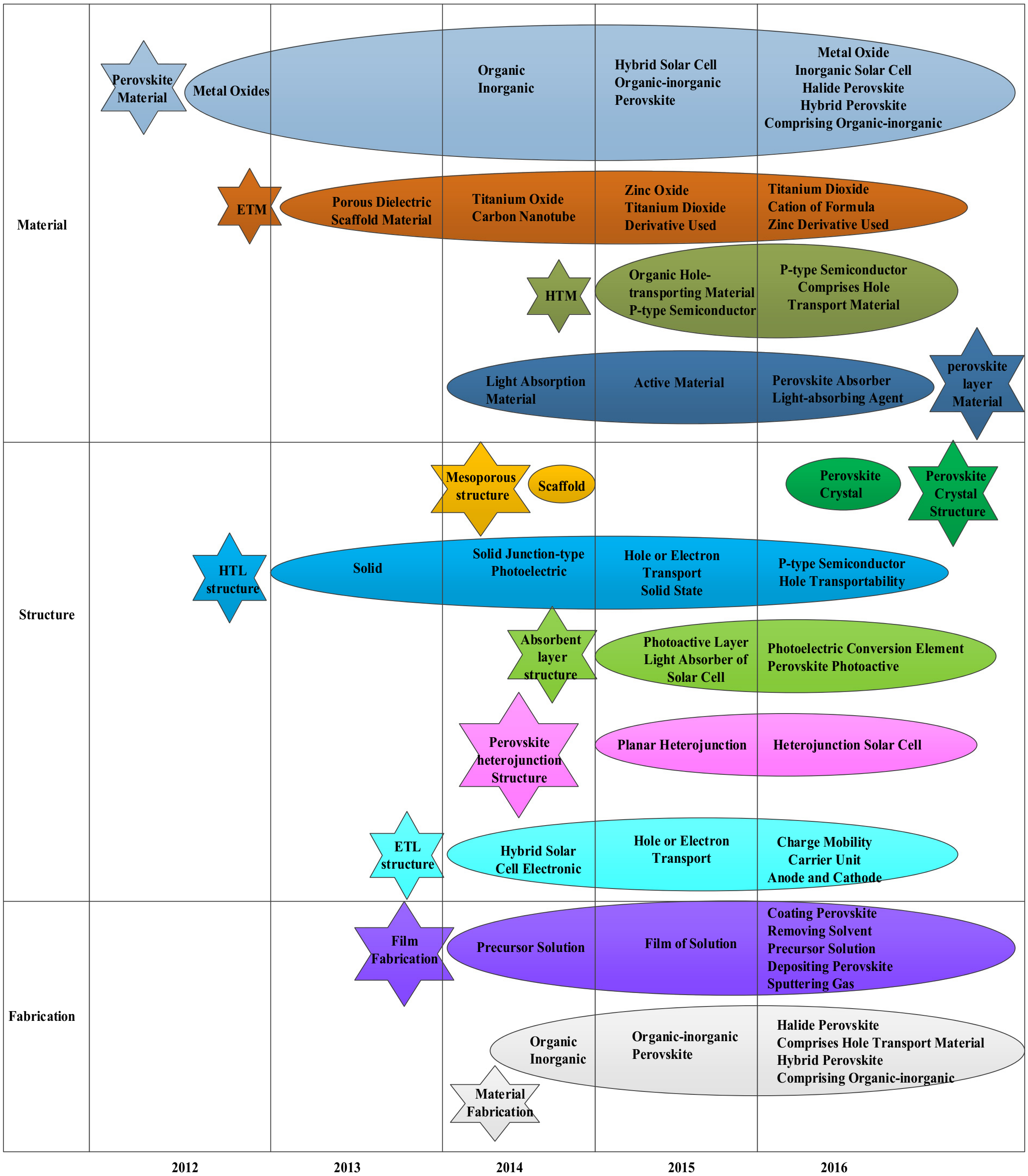


Fig. 6. Map of the evolution of perovskite solar cell technology based on patent text mining.

4.2.4. Gaps between science and technology analysis for forecasting technology trends

As shown in Fig. 5 and Fig. 6, the different evolution paths of perovskite solar cell technology based on scientific papers and patents text mining are evident. This shows that the description and research of the perovskite solar cell technology in scientific papers and patents on technology differ. Scientific papers have tended to focus on basic research, while patents have focused on the application of technology research (Shibata et al., 2010). Understanding the relationship between science and technology has become a key task for R&D managers and policy makers focusing on future technology (Shibata et al., 2010). Therefore, Shibata et al. (2010) extracted the commercialization gap between science and technology, and proposed that in the active technical research field, topics that exist in papers but not in patents are considered as technological opportunities. Technological opportunities can determine technological development (Olsson, 2005); thus, it is critical to identify technological opportunities for forecasting technology development trends. Based on the definition of technological opportunity proposed by Shibata et al. (2010), this paper makes a comparative analysis of the first appearance of the technical topics in scientific papers and in patents, as shown in Table 5. By comparing the time difference between the first appearance of the technical topics in scientific papers and in patents, we can effectively identify the technological opportunities in the field of perovskite solar cell technology, which is significant for better understanding the technology's development trends.

Table 5
Comparison of technology trends analysis based on scientific papers and patents.

Topics	Time of topics' appearance in the papers	Time of topics' appearance in the patents	Time lag
Perovskite material	2012	2012	0
Mesoporous structure	2012	2014	2
ETM	2013	2013	0
HTL structure	2013	2013	0
Film fabrication	2013	2014	1
Perovskite layer material	2013	2014	1
ETL structure	2013	2014	1
Absorbent layer structure	2013	2015	2
HTM	2013	2015	2
Material fabrication	2014	2014	0
Perovskite heterojunction structure	2014	2015	1
Perovskite crystal structure	2014	2016	2
Hole-conductor-free structure	2014	–	
Crystal fabrication	2014	–	

In addition to technological opportunities, the high growth rate of technical topics is also valuable when it comes to forecasting potential development trends of technology. The high growth rate of certain technical topics indicates that they have a higher growth potential and they will be able to catch the researchers' attention after a few years (Bildosola et al., 2017). Therefore, in this paper, we compared the growth rates of technical topics obtained from scientific papers and patents and tried to find the potential growth trends of these technical topics. We treated the top five topics as the technical topics with high growth rate based on the percentage increase of the topics in Tables 3 and 4. Due to there are too many new topics first appeared in 2013, therefore, we only considered the topics with high growth rate during the period from 2014 to 2016. The annual five technical topics with the highest growth rate obtained in scientific papers and patents are shown in Table 6. In Table 6, the number in brackets represents the growth rate of the topic with respect to the previous year.

To better understand the gaps between science and technology, and forecasting technology trends based on the gaps analysis, we incorporated the results of the first appearance of the technical topics in scientific papers and in patents and rapidly growth rates of technical topics obtained from scientific papers and patents, and the results are shown in Fig. 7. In Fig. 7, five-pointed star in different colors represents different topics, the correspondence relationships of the technical topics are connected by dotted lines, such as, red dotted line connects corresponding topics which the time lag between them is 2 years, and green dotted line connects corresponding topics which the time lag between them is 1 years. The topics in red circle mean technological opportunity which exists in scientific papers but not in patents. The annual numbers of topics with high growth rates are shown in combination chart.

As shown in Fig. 7, in 2012 , new technical topics such as perovskite material, and mesoporous structure appeared in the scientific papers, while only perovskite material appeared in that year's patents. Therefore, the mesoporous structure can be considered as technological opportunity after 2012. In the patents, the mesoporous structure appeared in 2014. The result validated the predicted technological opportunity.

As shown in Table 5 and Fig. 7, in 2013, new technical topics such as ETM, HTL structure, film fabrication, perovskite layer material, ETL structure, absorbent layer structure, and HTM appeared in the scientific papers, but only ETM and HTL structure appeared in that year's patents. Therefore, the film fabrication, perovskite layer material, ETL structure, absorbent layer structure, and HTM can be considered as technological opportunities after 2013. In the patents, the film fabrication, perovskite layer materials, and ETL structure appeared in 2014, and absorbent layer structure and HTM appeared in 2015. These results also validated the predicted technological opportunities. Therefore, it is possible to identify technological opportunities according to the time lag between the technical topics' first appearance in scientific papers and in patents. In 2014, new technical topics appeared in scientific papers, such as material fabrication, perovskite crystals structure, perovskite heterojunction structure, crystal fabrication, and hole-conductor-free structure, while there is only the material fabrication in patents. Therefore, other new technical topics can also be considered as technological opportunities after 2014. In 2015, the perovskite heterojunction structure appeared in the patents, and the perovskite crystal structure appeared in the patents in 2016. As there are no patents corresponding to the hole-conductor-free structure and the crystal fabrication before 2016, these two technical topics could be regarded as technological opportunities after 2016.

As shown in Table 6 and Fig.7, the perovskite layer material in the scientific papers has a high growth rate in 2014 and in 2016, while it has a high growth rate in patents in 2015; the ETM in the scientific papers has a high growth rate both in 2014 and in 2015, while in the patents it has a high growth rate both in 2014 and in 2016. The ETL structure in the scientific papers has a higher growth rate in the period of 2014 to 2016, while in the patents it has a higher growth rate in 2015. The film fabrication in the scientific papers has a high growth rate in 2014, while in the patents it has a high growth rate in 2015 and in 2016. Based on these results, we can see that the corresponding topics in the scientific papers and patents follow the same trends. The technical topics with high growth rates first appeared in scientific papers, and then they followed the same growth trends in the patents. Therefore, we may predict the technical topics' growth trends in patents according to their development trends in scientific papers.

Based on above analysis, we can predict the technology development trends by combining the analysis of corresponding relationships between the technical topics in the scientific papers and patents, and the analysis of rapidly growing technical topics in the scientific papers and patents. Therefore, as shown in Tables 5, 6 and Fig. 7, we can obtain the technology development trends by analyzing the gaps

Table 6
Rapidly growing topics in scientific papers and patents from 2014 to 2016.

	2014	2015	2016
Scientific papers	Perovskite material (4300)	ETM (827)	Hole-conductor-free structure (484)
	Film fabrication (3450)	Crystal fabrication (546)	Perovskite layer material (484)
	Perovskite layer material (2600)	Perovskite crystal structure (207)	HTM (362)
	ETL structure (2063)	ETL structure (205)	ETL structure (250)
	ETM (1750)	Material fabrication(160)	Perovskite crystal structure (220)
Patents	ETM (500)	Perovskite material (350)	Material fabrication (894)
	Perovskite material (300)	Perovskite layer material (329)	Film fabrication (745)
		HTL structure (286)	Perovskite material (388)
		Film fabrication (233)	Perovskite heterojunction structure ((250)
		ETL structure (186)	ETM (236)

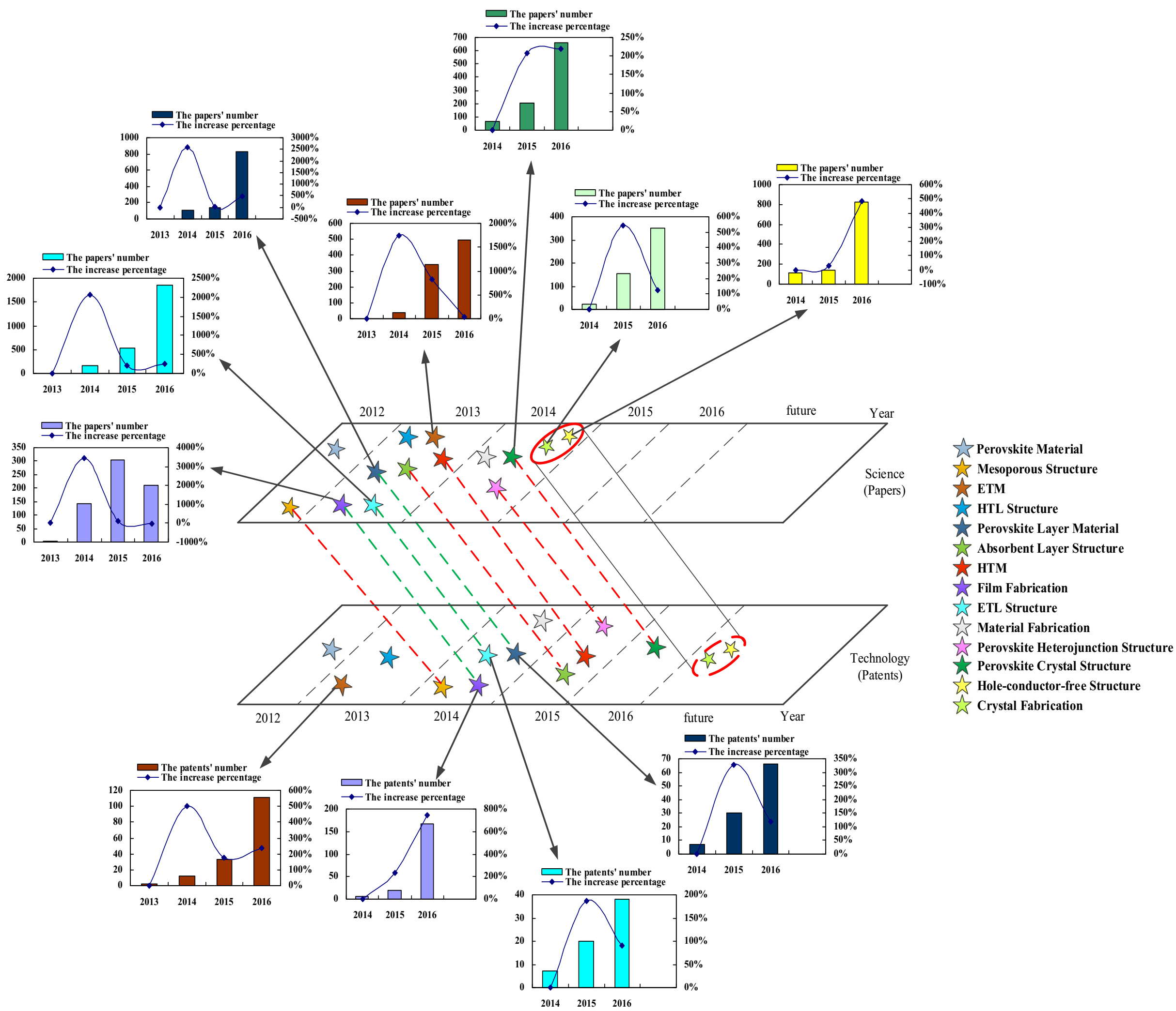


Fig. 7. Gaps between science and technology analysis for forecasting technology trends.

between science and technology, and the results are as follows:

(1) The crystal fabrication first appeared in scientific papers in 2014, and showed a higher growth rate in 2015. Meanwhile, the number of scientific papers containing this topic has increased in 2015 and in 2016. While it didn't appear in patents before 2016, we can predict that it will appeared in patents after 2016 with high growth potential.
(2) The hole-conductor-free structure first appeared in scientific papers in 2014, and showed a higher growth rate in 2016. Meanwhile, the number of scientific papers containing this topic has increased in 2015 and in 2016. While it didn't appear in patents before 2016, we can predict that it will appeared in patents after 2016 with high growth potential.
(3) The perovskite crystal structure first appeared in scientific papers in 2014, and showed a higher growth rate in 2015 and in 2016. While it appeared in patents in 2016, we can predict that it will have high growth potential in patents after 2016.
(4) The ETL structure first appeared in scientific papers in 2013, and showed a higher growth rate in the period of 2014 to 2016. While it appeared in patents in 2014, and showed a higher growth rate in 2015. We can predict that it will have high growth potential in patents after 2016.
(5) The perovskite layer material first appeared in scientific papers in 2013, and showed a high growth rate in 2014 and in 2016. Meanwhile, the number of scientific papers containing this topic has increased in 2015 and in 2016. While it appeared in patents in 2014, and showed a higher growth rate in 2015. We can predict that it will have high growth potential in patents after 2016.
(6) The ETM first appeared in scientific papers in 2013, and showed a higher growth rate in 2014 and in 2015. Meanwhile, the number of scientific papers containing this topic has increased in the period of 2014 to 2016. While it also appeared in patents in 2013, and showed a higher growth rate in 2014 and in 2016. The number of patents containing this topic also has increased in the period of 2014 to 2016. We can predict that it will continue have high growth potential in patents after 2016.

From the results of analyzing the gaps between science and technology, we can predict that the crystal fabrication, hole-conductor-free structure, perovskite crystal structure, ETL structure, perovskite layer material, and ETM will be future development trends of perovskite solar cell technology.

5. Discussion

Science provides a fundamental basis for technology development, and the science is considered as seeds of technology and innovation in the linear innovation model (Shibata et al., 2010). Much information of technological development can be obtained from analysis of the development of science. Many empirical studies show that the interdependencies and interactions between science and technology have been increasing. Despite this interdependence, gaps also exist between them (Wang et al., 2015a). Thus, how to discover the gaps between science and technology through the mining of technical information contained in scientific papers and patents, and to study technology development trends is an important issue that scholars need to pay attention. Therefore, this paper proposes a framework that integrates scientific papers and patents as data resources to study technology trends. In the framework, topics-based text mining and experts' knowledge approaches are applied to mining the technical knowledge and information contained in scientific papers and patents respectively, and to identify the technology evolution path, and a gaps analysis between science and technology is used to forecast technology trends within the short term. The perovskite solar cell technology is selected as a case study. Through the analysis of the case, the evolution path and development trends of perovskite solar cell technology have been developed, and the proposed framework has also been proven to be valid and robust.

The framework provides a tool for understanding the evolution path of technology and forecasting technology development trends by analyzing the gaps between science and technology. For the gaps analysis, we mainly analyzed two different gaps: One gap shows the difference between technical topics' first appearance in scientific papers and in patents, and the other is the differences analysis between the technical topics with high growth rates extracted from scientific papers and patents. For the gaps analysis of technical topics' first appearance in scientific papers and in patents, we found that the technical topics that appeared in patents generally can be found the corresponding knowledge base and the same technical topics in scientific papers, but sometimes only the first appearance time is different. Therefore, it is possible to predict technology development trends according to the time lag between the technical topics' first appearance in scientific papers and in patents. For the gaps analysis of the technical topics with high growth rates extracted from scientific papers and patents, we found that the technical topics with high growth rates first appeared in scientific papers, and then they followed the same growth trends in the patents. Based on these results, we can see that the corresponding topics in the scientific papers and patents follow the same trends. Therefore, we may predict the technical topics' growth trends in patents according to their development trends in scientific papers.

In the analysis of gaps between science and technology, we found that the technical topics extracted from patents verify the development trends raised in the scientific papers, which indicates that the framework is valid and flexible to predict the development trend of technology. Therefore, when we study technology development trends, we can apply topics-based text mining and experts' knowledge approaches to mine the technical knowledge and information contained in scientific papers and patents respectively, and construct technological evolution path maps based on scientific papers and patents with the help of domain experts. Based on understanding the different characteristics of the two technological evolution paths over time, we can predict technology development trends by analyzing the differences between the two technological evolution paths and comparing the gaps (e.g. One gap is the difference between technical topics' first appearance in scientific papers and in patents, and the other is the differences analysis between the technical topics with high growth rates extracted from scientific papers and patents) between science and technology. This is the key contribution of this paper and the difference between this paper and previous studies, and the framework also provides a new perspective for forecasting technology trends.

We constructed the evolution path of the perovskite solar cell technology by the use of text mining method and expert judgment approach, which offers a better understanding of the emergence and future trends of this technology. We found three significant technology trends of perovskite solar cell technology by combining the gaps analysis results with the detailed evolution paths of perovskite solar cell technology based on scientific papers and patents, which may be interest to solar photovoltaic technology R&D experts, supporting R&D strategy, and decision-making. The technology trends are as follows: (1) For the material level, ETM and perovskite layer material are the future development trend of materials related to perovskite solar cell technology. The design and development of materials related to perovskite solar cell technology tend to be cheap, stable, environmentally friendly and of higher performance. (2) For the structural level, the hole-conductor-free structure, ETL structure, and perovskite crystal structure are future trends of structural development related to perovskite solar cell technology, and the perovskite solar cells' structure tends towards becoming simple and economical. (3) For the fabrication level, the crystal fabrication that is economic, easy to control, and simple represent the future trend of fabrication development related to perovskite solar cell technology.

6. Conclusions and future study

This paper proposes a framework that uses scientific papers and patents as data resources, and integrates text mining and expert judgment approach to forecast technology trends by identifying the gaps between science and technology. The text mining and expert judgment approaches were applied to analyze the technical topics appearing in scientific papers and patents, and gaps analysis between science and technology was used to forecast technology development trends. Perovskite solar cell technology was selected as a case study, through which the proposed framework was proven to be valid and flexible. This paper contributes to the technology forecasting methodology and sheds light on the emergence and future trends of technology studies and R&D policy analysis.

There are some limitations and issues regarding our method that need to be considered. First, for the clustering method of text mining, in this paper, we used only the Lingo algorithm to cluster technical topics. In the future, clustering algorithm based on semantic analysis may be applied to cluster technical topics, which may improve the effectiveness and accuracy of clustering results. Second, in the analysis of gaps between science and technology, we found that the first appearance of the technical topics extracted from patents has a time lag compared with their first appearance in the scientific papers. In the future, more case studies might reveal and summarize the time lag of the technical topics' appearances in scientific papers and patents, which may be useful for forecasting technology trends. Third, the methodology of this paper is based on the existing research on the relationship between science and technology; that is, the development of technology depends on the science and technological information that can be captured from the scientific literature. This relationship may be applicable for forecasting trends in knowledge-intensive science technology, while it may not be suitable for less knowledge-intensive technology. Therefore, the applicable scope of the proposed methodology may be limited to knowledge-intensive science technology. In addition, we note that the use of scientific papers and patents to predict technology trends also suffers from several limitations, which may be listed as follows: (1) not all science and technology developments are published, and not all records are valuable (Porter and Detampel, 1995), and (2) timely and important information may also be missing because of publishing lag times (Huang et al., 2014). Therefore, expert judgment and scenario-planning method combined with text mining may be useful in further research if we wish to predict technology trends and possible events that shape the future development of technology.

Acknowledgements

This paper is supported by the National Natural Science Foundation of China (Grant 71673018), Ministry of Education Social Science Youth Foundation of China (Grant 14YJC630071), Ministry of Education

Social Science Foundation of Beijing (Grant SM201610005001), and International Cooperation Seed Fund of Beijing University of Technology (2018B23). We are grateful to the four anonymous reviewers for their helpful comments and suggestions. Many thanks to Dr. Binxin Zhao, Prof. Hongyi Li, Prof. Qingbo Meng, Dr. Yuan Zhou and Ms. Jingjing Wang for their valuable consultations.

Appendix I

Table 1
Short name of the topic clustering.

Clustering theme	Short name
Electronic transport material	ETM
Hole transport material	HTM
Zinc oxide	ZnO
Titanium dioxide	TiO2
Lead iodide	PbI3
Perovskite solar cell	PSC
Hole transport layer	HTL
Electronic transport layer	ETL
Fluorine-doped tin oxide layer	FTO
3 Dimension	3D
2 Dimension	2D

Table 2
Classification of the topic clustering.

The name of the layer		Elements
Material	Perovskite material	Oxides, PbI_3, $CH_3NH_3PbI_3$, Organolead halide perovskites, Organic-inorganic perovskites, Inorganic organic, Halide perovskites, Inorganic hybrid perovskites, Perovskite films, Methylammonium lead, Organic-inorganic halide perovskites, Hybrid organic-inorganic, Halide perovskite materials, Metal oxides, Hybrid solar cell, Inorganic solar cell, Comprising organic-inorganic, $CH_3NH_3PbI_3$ perovskite films, Resulting $CH_3NH_3PbI_3$, Devices based on methylummonium lead, Material $CH_3NH_3PbI_3$, Halide $CH_3NH_3PbI_3$, $CH_3NH_3PbI_3$ perovskite
	ETM	Tio_2, Fullerene, Tio2 $CH_3NH_3PbI_3$, Mesoporous Tio2, ZnO nanoparticles, Perovskite Tio2, Tio2 layer, Perovskite Tio2, Transport material in perovskite solar cells, Tio2 as electron, Tin oxide perovskite, Porous dielectric scaffold material, Titanium oxide, Zinc oxide, Titanium dioxide, Derivative used, Titanium dioxide, Cation of formula, Zinc derivative used, Carbon nanotube
	HTM	Spirobifluorene spiro-OMeTAD, Hole transport material, Transport material in perovskite solar cells, Organic hole-transporting material, P-type semiconductor, Comprises hole transport material
	Perovskite layer material	Absorber layers, Light harvester, Perovskite absorber, Light absorption material, Active material, Light-absorbing agent, Light-absorbing layer, $CH_3NH_3PbI_3$, Absorption material, Resulting $CH_3NH_3PbI_3$, Performance of $CH_3NH_3PbI_3$, $CH_3NH_3PbI_3$ perovskite, Material $CH_3NH_3PbI_3$
Structure	Perovskite crystals structure	Perovskite crystal, Crystal structure, Resulting in crystals, Crystal perovskite films, Hybrid structure, Single crystals
	Mesoporous structure	Mesoscopic Tio_2 film, Mesoporous Tio2, ZnO nanorod, Mesoporous scaffold, Scaffold
	Perovskite heterojunction structure	Perovskite planar, Perovskite heterojunction, Heterojunction solar cells, Planar heterojunction, Heterojunction solar cell
	Hole-transport-free structure	Performance of $CH_3NH_3PbI_3$, $CH_3NH_3PbI_3$ perovskite films, Resulting $CH_3NH_3PbI_3$, $CH_3NH_3PbI_3$, Halide $CH_3NH_3PbI_3$
	HTL structure	Holes in perovskite, Perovskite hole transport, Transport in the device, Hole-transporting, Solid, Solid junction-type photoelectric, Hole or electron transport, Solid state, Hole transportability, Solid-state mesoscopic solar cells, P-type semiconductor
	ETL structure	Electron transport in perovskite, Effective electron, Charge transport, Charge separation, Electrode, Electron-hole diffusion lengths, Electronic properties, Band gap, Charge recombination, Energy level, Charge recombination at the perovskite, Carrier transport, Electronic structure, High electron, Electron from the perovskite film, Hybrid solar cell electronic, Hole or electron transport, Charge mobility, Carrier Unit, Anode and cathode
	Absorbent layer structure	Absorber layers, Light harvester, Photoactive layer, Light absorber of solar cell, Photoelectric conversion element, Perovskite photoactive, Optical properties
Fabrication	Film fabrication	Atomic layer deposition, Two-step deposition technique, Deposited perovskite, Perovskite solution, Fabrication process, Films prepared, Precursor solution, Film of solution, Coating perovskite, Removing solvent, Sputtering gas
	Crystal fabrication	Room temperature, Crystallization process, Crystallization leads, Processing temperature
	Material fabrication	Organic inorganic, Halide perovskites, Hybrid organic-inorganic, Halide perovskite materials, Halide CH3NH3PbI3, Organic-inorganic perovskite, Comprises hole transport material, Comprising organic-inorganic, Halide perovskite

References

Behkami, N., Daim, T., 2012. Research forecasting for health information technology (HIT), using technology intelligence. Technol. Forecast. Soc. Chang. 79 (3), 498–508.

Bildosola, I., Río-Bélver, R.M., Garechana, G., Cilleruelo, E., 2017. TeknoRoadmap: an approach for depicting emerging technologies. Technol. Forecast. Soc. Chang. 117, 25–37.

Breschi, S., Catalini, C., 2010. Tracing the links between science and technology: an exploratory analysis of scientists' and inventors' networks. Res. Policy 39 (1), 14–26.

Callon, M., Courtial, J.P., Laville, F., 1991. Co-word analysis as a tool for describing the network of interactions between basic and technological research: the case of polymer chemistry. Scientometrics 22 (1), 155–205.

Carvalho, M.M., Fleury, A., Lopes, A.P., 2013. An overview of the literature on technology roadmapping (TRM): contributions and trends. Technol. Forecast. Soc. Chang. 80 (7), 1418–1437.

Chang, P.L., Wu, C.C., Leu, H.J., 2010. Using patent analyses to monitor the technological trends in an emerging field of technology: a case of carbon nanotube field emission display. Scientometrics 82 (1), 5–19.

Chen, S.H., Huang, M.H., Chen, D.Z., 2012. Identifying and visualizing technology evolution: a case study of smart grid technology. Technol. Forecast. Soc. Chang. 79 (6), 1099–1110.

Chen, H.S., Zhang, G.Q., Zhu, D.H., Lu, J., 2017. Topic-based technological forecasting based on patent data: a case study of Australian patents from 2000 to 2014. Technol.

Forecast. Soc. Chang. 119 (7), 39–52.
Choi, J., Hwang, Y.S., 2014. Patent keyword network analysis for improving technology development efficiency. Technol. Forecast. Soc. Chang. 83 (1), 170–182.
Choi, C., Park, Y., 2009. Monitoring the organic structure of technology based on the patent development paths. Technol. Forecast. Soc. Chang. 76 (6), 754–768.
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.
Daim, T.U., Rueda, G., Martin, H., Gerdsri, P., 2006. Forecasting emerging technologies: use of bibliometrics and patent analysis. Technol. Forecast. Soc. Chang. 73 (8), 981–1012.
Daim, T.U., Iskin, I., Li, X., Zielsdoff, C., Bayraktaroglu, A.E., Dereli, T., Durmusoglu, A., 2012. Patent analysis of wind energy technology using the patent alert system. World Patent Inf. 34 (1), 37–47.
Dereli, T., Durmusoglu, A., 2009. Classifying technology patents to identify trends: applying a fuzzy-based clustering approach in the Turkish textile industry. Technol. Soc. 31 (3), 263–272.
Ding, W., Chen, C., 2014. Dynamic topic detection and tracking: a comparison of HDP, C-word, and cocitation methods. J. Assoc. Inf. Sci. Technol. 65 (10), 2084–2097.
Dotsika, F., Watkins, A., 2017. Identifying potentially disruptive trends by means of keyword network analysis. Technol. Forecast. Soc. Chang. 119, 114–127.
Ena, O., Mikova, N., Saritas, O., Sokolova, A., 2016. A methodology for technology trend monitoring: the case of semantic technologies. Scientometrics 108 (3), 1013–1041.
Garcíalillo, F., ÚbedaGarcía, M., Marcolajara, B., 2016. The intellectual structure of research in hospitality management: a literature review using bibliometric methods of the journal international journal of hospitality management. Int. J. Hosp. Manag. 52, 121–130.
Gibson, E., Blommestein, K., Kim, J., Daim, T., Garces, E., 2017. Forecasting the electric transformation in transportation. Tech. Anal. Strat. Manag. 29 (10), 1103–1120.
Golembiewski, B., Stein, N.V., Sick, N., Wiemhofer, H.D., 2015. Identifying trends in battery technologies with regard to electric mobility: evidence from patenting activities along and across the battery value chain. J. Clean. Prod. 87 (1), 800–810.
Hao, J., Yan, Y., Gong, L., Wang, G., Lin, J., 2014. Knowledge map-based method for domain knowledge browsing. Decis. Support. Syst. 61, 106–114.
Huang, L., Zhang, Y., Guo, Y., Zhu, D.H., Porter, A.L., 2014. Four dimensional science and technology planning: a new approach based on bibliometrics and technology roadmapping. Technol. Forecast. Soc. Chang. 81 (1), 39–48.
Hussain, M., Tapinos, E., Knight, L., 2017. Scenario-driven roadmapping for technology foresight. Technol. Forecast. Soc. Chang. 124.
Jaewoo, C., Woonsun, K., 2014. Themes and trends in Korean educational technology research: a social network analysis of keywords. Procedia Soc. Behav. Sci. 131, 171–176.
Kajikawa, Y., Takeda, Y., 2008. Structure of research on biomass and bio-fuels: a citation-based approach. Technol. Forecast. Soc. Chang. 75 (9), 1349–1359.
Kajikawa, Y., Takeda, Y., 2009. Citation network analysis of organic LEDs. Technol. Forecast. Soc. Chang. 76 (8), 1115–1123.
Kajikawa, Y.Y., Yoshikawa, J., Takeda, Y., Matsushima, K., 2008. Tracking emerging technologies in energy research: toward a roadmap for sustainable energy. Technol. Forecast. Soc. Chang. 75 (6), 771–782.
Kostoff, R.N., 1998. The use and misuse of citation analysis in research evaluation. Scientometrics 43 (1), 27–43.
Kostoff, R.N., Schaller, R.R., 2001. Science and technology roadmaps. IEEE Trans. Eng. Manag. 48 (2), 132–143.
Kostoff, R.N., Solka, J.L., Rushenberg, R.L., Wyatt, J.A., 2008. Literature-related discovery (LRD): water purification. Technol. Forecast. Soc. Chang. 75 (2), 256–275.
Lee, S., Kang, S., Park, E., Park, Y., 2008. Applying technology road-maps in project selection and planning. Int. J. Qual. Reliab. Manag. 25 (1), 9–51.
Lee, S., Yoon, B., Lee, C., Park, J., 2009. Business planning based on technological capabilities: patent analysis for technology-driven roadmapping. Technol. Forecast. Soc. Chang. 76 (6), 769–786.
Lee, P.C., Su, H.N., Wu, F.S., 2010. Quantitative mapping of patented technology–the case of electrical conducting polymer nanocomposite. Technol. Forecast. Soc. Chang. 77 (3), 466–478.
Li, X., Zhou, Y., Xue, L., Huang, L.C., 2015. Integrating bibliometrics and roadmapping methods: a case of dye-sensitized solar cell technology-based industry in China. Technol. Forecast. Soc. Chang. 97, 205–222.
Liu, X., Jiang, T.T., Ma, F.C., 2013. Collective dynamics in knowledge networks: emerging trends analysis. J. Informet. 7 (2), 425–438.
Madani, F., Weber, C., 2016. The evolution of patent mining: applying bibliometrics analysis and keyword network analysis. World Patent Inf. 46, 32–48.
Marzi, G., Dabic, M., Daim, T., Garces, E., 2017. Product and process innovation in manufacturing firms—a thirty-year bibliometric analysis. Scientometrics 113 (2), 673–704.
McMillan, G.S., Narin, F., Deeds, D.L., 2000. An analysis of the critical role of public science in innovation: the case of biotechnology. Res. Policy 29 (1), 1–8.
Narin, F., Noma, E., 1985. Is technology becoming science? Scientometrics 7 (3–6), 369–381.
Noh, H., Jo, Y., Lee, S., 2015. Keyword selection and processing strategy for applying text mining to patent analysis. Expert Syst. Appl. 42 (9), 4348–4360.
Noh, H., Song, Y.K., Lee, S., 2016. Identifying emerging core technologies for the future: case study of patents published by leading telecommunication organizations. Telecommun. Policy 40 (10 − 11), 956–970.
Olsson, O., 2005. Technological opportunity and growth. J. Econ. Growth 10 (1), 31–53.
Osinski, S., Weiss, D., 2005. A concept-driven algorithm for clustering search results. IEEE Intell. Syst. 20 (3), 48–54.
Phaal, R., Farrukh, C.J., Probert, D.R., 2004. Technology roadmapping—a planning framework for evolution and revolution. Technol. Forecast. Soc. Chang. 71 (1), 5–26.
Phaal, R., O'Sullivan, E., Routley, M., Ford, S., Probert, D., 2011. A framework for mapping industrial emergence. Technol. Forecast. Soc. Chang. 78 (2), 217–230.
Porter, A.L., Detampel, M.J., 1995. Technology opportunities analysis. Technol. Forecast. Soc. Chang. 49 (3), 237–255.
Ravikumar, S., Agrahari, A., Singh, S.N., 2015. Mapping the intellectual structure of scientometrics: a co-word analysis of the journal scientometrics (2005–2010). Scientometrics 102 (1), 929–955.
Rezaeian, M., Montazeri, H., Loonen, R.C.G.M., 2017. Science foresight using life-cycle analysis, text mining and clustering: a case study on natural ventilation. Technol. Forecast. Soc. Chang. 118, 270–280.
Rifkin, J., 2011. The Third Industrial Revolution: How Lateral Power is Transforming Energy, the Economy, and the World. Palgrave Macmillan, New York.
Shibata, N., Kajikawa, Y., Takeda, Y., Matsushima, K., 2008. Detecting emerging research fronts based on topological measures in citation networks of scientific publications. Technovation 28 (11), 758–775.
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 (7), 1147–1155.
Teh, Y.W., Jordan, M.I., Beal, M.J., Blei, D.M., 2006. Hierarchical Dirichlet processes. J. Am. Stat. Assoc. 101 (476), 1566–1581.
Teufel, S., Siddharthan, A., Dan, T., 2006. An annotation scheme for citation function. Proceedings of Sigdial workshop on discourse and. Dialogue 80–87.
Tijssen, R.J.W., 2001. Global and domestic utilization of industrial relevant science: patent citation analysis of science-technology interactions and knowledge flows. Res. Policy 30 (1), 35–54.
Tsai, H.H., 2012. Global data mining: an empirical study of current trends, future forecasts and technology diffusions. Expert Syst. Appl. 39 (9), 8172–8181.
Tseng, Y.H., Lin, C.J., Lin, Y.I., 2007. Text mining techniques for patent analysis. Inf. Process. Manag. 43 (5), 1216–1247.
Wang, B., Liu, S., Ding, K., Liu, Z.Y., Xu, J., 2014. Identifying technological topics and institution-topic distribution probability for patent competitive intelligence analysis: a case study in LTE technology. Scientometrics 101 (1), 685–704.
Wang, M.Y., Fang, S.C., Chang, Y.H., 2015a. Exploring technological opportunities by mining the gaps between science and technology: microalgal biofuels. Technol. Forecast. Soc. Chang. 92, 182–195.
Wang, X.F., Qiu, P.G., Zhu, D.H., Mitkova, L., Lei, M., Porter, A.L., 2015b. Identification of technology development trends based on subject–action–object analysis: the case of dye-sensitized solar cells. Technol. Forecast. Soc. Chang. 98, 24–46.
Watts, R.J., Porter, A.L., 1997. Innovation forecasting. Technol. Forecast. Soc. Chang. 56 (1), 25–47.
Wu, C.C., Leu, H.J., 2013. Exploring the technological trends for a novel technology through patent network analysis: the case of carbon nanotubes. Inf. Jpn. 16 (7), 5291–5301.
Wu, C.C., Leu, H.J., 2014. Examining the trends of technological development in hydrogen energy using patent co-word map analysis. Int. J. Hydrog. Energy 39 (33), 19262–19269.
Yoon, J., Kim, K., 2011. Identifying rapidly evolving technological trends for R&D planning using SAO-based semantic patent networks. Scientometrics 88 (1), 213–228.
Yoon, J., Kim, K., 2012. TrendPerceptor: a property–function based technology intelligence system for identifying technology trends from patents. Expert Syst. Appl. 39 (3), 2927–2938.
Yoon, B., Park, Y., 2005. A systematic approach for identifying technology opportunities: keyword-based morphology analysis. Technol. Forecast. Soc. Chang. 72 (2), 145–160.
Zhang, Y., Guo, Y., Wang, X., Zhu, D., Porter, A.L., 2013. A hybrid visualisation model for technology roadmapping: bibliometrics, qualitative methodology and empirical study. Tech. Anal. Strat. Manag. 25 (6), 707–724.
Zhang, Y., Robinson, D.K.R., Porter, A.L., 2015. Technology roadmapping for competitive technical intelligence. Technol. Forecast. Soc. Chang. 110, 175–186.
Zhang, Y., Zhang, G.Q., Chen, H., Porter, A.L., Zhu, D.H., Lu, J., 2016. Topic analysis and forecasting for science, technology and innovation: methodology with a case study focusing on big data research. Technol. Forecast. Soc. Chang. 105, 179–191.
Zhang, Y., Zhang, G., Zhu, D., Lu, J., 2017. Science evolutionary pathways: identifying and visualizing relationships for scientific topics. J. Assoc. Inf. Sci. Technol. 68 (8), 1925–1939.

Xin Li is an associate researcher in the College of Economics and Management, Beijing University of Technology, China. His recent research interests include technology forecasting, data mining, emerging technologies, and technology roadmapping.

Qianqian Xie is a master candidate in the College of Economics and Management, Beijing University of Technology, China. Her recent research interests include technology forecasting, and data mining.

Tugrul Daim is a professor in the Department of Engineering and Technology Management, Portland State University Portland, OR, USA. His recent research interests include technology forecasting, and technology roadmapping. He is also a leading research fellow at the National Research University Higher School of Economics, Moscow, Russia and an Honorary Chair Professorat the Chaoyang University of Technology, Taichung, Taiwan

Lucheng Huang is a professor in the College of Economics and Management, Beijing University of Technology, China. His recent research interests include emerging technologies, and technology innovation management.