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

Identifying and monitoring the development trends of emerging technologies using patent analysis and Twitter data mining: The case of perovskite solar cell technology

Xin Li[a,*], Qianqian Xie[a], Jiaojiao Jiang[b], Yuan Zhou[c], Lucheng Huang[a]

[a] *College of Economics and Management, Beijing University of Technology, Beijing, China*
[b] *School of Software and Electrical Engineering, Swinburne University of Technology, Melbourne, Australia*
[c] *School of Public Policy and Management, Tsinghua University, Beijing, China*

ARTICLE INFO

Keywords:
Emerging technologies
Technology trends
Patent analysis
Twitter data mining
Technologies emerge
Perovskite solar cell technology

ABSTRACT

Monitoring the emergence of emerging technologies helps managers and decision makers to identify development trends in emerging technologies is crucial for government research and development (R&D), strategic planning, social investment, and enterprise practices. Researchers usually use academic papers and patent data to identify and monitoring the trends of emerging technologies from a technological perspective, but they rarely make use of social media data (e.g., such as Twitter data) related to emerging technologies. Analysis of this social media data is of great significance to understand the emergence of emerging technologies and gain insight into development trends. Therefore, this paper proposes a framework that uses patent analysis and Twitter data mining to monitoring the emergence of emerging technologies and identify changing trends of these emerging technologies. The perovskite solar cell technology is selected as a case study. In this case, we used patent analysis to monitoring the evolutionary path of perovskite solar cell technology. We applied Twitter data mining to analyze Twitter users' sense of, response to, and expectations for this perovskite solar cell technology. We also identified the professional types of Twitter users and examined changes in their topics of interest over time to track the emergence of perovskite solar cell technology. We analyzed a comparison of the results of patent analysis and Twitter data mining to identify development trends of perovskite solar cell technology. This paper contributes to our understanding of how technologies emerge and develop, as well as the technology forecasting and foresight methodology, and will be of interest to solar photovoltaic technology R&D experts.

1. Introduction

In recent years, emerging technologies with disruptive characteristics have been rapidly emerging. Major technological advancements include information technology, nanotechnology, biotechnology, and new material technology. The emergence and development of these emerging technologies not only has changed existing industries but also has created new industries that have a significant impact on the socioeconomic structure (Day et al., 2000). Identifying future changing trends of these emerging technologies as early as possible is crucial for government and enterprise research and development (R&D) strategic planning to gain a first-mover advantage in the competitive business environment. Many managers and decision makers are aware of the significance of understanding of the emergence path, and identifying the future development trends of these emerging technologies for their organization's competitiveness and sustainable development (Li et al., 2015). This strategic concern issue raises one question: Can we identify the future development trends of these emerging technologies given a better understanding of how they emerge? In response to this question, this paper develops a framework to monitoring the emergence of emerging technologies and identify future changes trends of these emerging technologies by using patent analysis and Twitter data mining.

A technology trend is considered to be a continuously growing technology area with a certain pattern, and the pattern of this trend should have existed for a certain period of time (Ena et al., 2016). Technological monitoring is an important tool used to identify technological trends (Porter and Roper, 1991). Monitoring refers to the act of scanning a particular area of relevant information to understand the history of a particular technology (Momeni and Rost, 2016). Understanding the history of a technology is the basis for better identifying technological changes. Analyzing technically relevant historical

* Corresponding author.
E-mail address: lixinyz@bjut.edu.cn (X. Li).

https://doi.org/10.1016/j.techfore.2018.06.004
Received 25 January 2018; Received in revised form 16 May 2018; Accepted 4 June 2018

information can be applied to identify how changes in technological developments are influenced by current and past changes in related technologies (Kostoff, 1994; Porter and Detampel, 1995). Patents provide a source of up-to-date and reliable information for revealing technological information and development (Noh et al., 2015). Through the analysis of the technological information available in patents, we may better reveal and understand the path of technological evolution and be able to identify the development trends of technology with the help of domain experts. Thus, researchers have begun to use patent data to analyze and study these technology trends (Chen et al., 2017; Daim et al., 2006; Kajikawa et al., 2008; Kajikawa and Takeda, 2009; Li et al., 2015; Robinson et al., 2013; Yoon et al., 2013; Yoon and Kim, 2011).

The evolutionary path and development trends of a technology are the result of synergies between the actual technology and the technological environment in which it is developed (Boon and Moors, 2008; Katz, 2006; Small et al., 2014). In addition, because of the uncertainty and ambiguity inherent in these emerging technologies, studies on the development trends of these emerging technologies have focused on patent data that not only are sensitive to time but also provide limited perspectives on the multifaceted phenomenon of emerging technologies (Rotolo et al., 2015). Therefore, to better understand the evolutionary path and development trends of emerging technologies, we should pay more attention to the environmental factors that could shape or change the emergence and development process of these emerging technologies, and we need to focus on the use of timely data and up-to-date data that reflect the latest advances of emerging technologies to better identify the development trends of these emerging technologies.

With the rapid development of the Internet and information technology, a large amount of social awareness[1] data that represents the public's sense of and response to emerging technologies is available in social media platforms, such as Twitter. Because social media provides an enormous amount of timely data and up-to-date data (Injadat et al., 2016), businesses and research communities have paid considerable attention to social media (Aral et al., 2013; Kalampokis et al., 2013). Social media data have been used to make stock price predictions (Daniel et al., 2017), prevent epidemics, monitoring early events (Hughes and Palen, 2009), predict elections, manage crises, and conduct brand management by businesses and in research communities (Inauen and Schoeneborn, 2014; Williams et al., 2013). Social media is an interesting data resource for early detection and identification of these emerging technologies (Breitzman and Thomas, 2015). In the emergence and development process of emerging technology, public's sense of and response to emerging technologies contained in social media could affect other public's sense of and response to emerging technologies, such as the public's expectations for, comments about, and anticipation of emerging technologies contained in social media that may be stimulating and driving the behavior of stakeholders in these emerging technology (Borup et al., 2006). The expectations for the future development of emerging technologies will attract social investment, and an increase in social investment can accelerate the emergence and development of emerging technologies (Rotolo et al., 2015). Therefore, the analysis of social awareness data contained in social media is of great significance for understanding the emergence of emerging technologies and identifying the future development trends of these emerging technologies.

Twitter is representative of the rapid growth of social media, and it is an important data resource to study social public's sense of and response to emerging technologies. Twitter users can post opinions about topics on Twitter platform, such as products, services, and television programs (Ikeda et al., 2013). As for a particular emerging technology, researchers, technology experts, inventors, and other members of the public who use Twitter can share their expertise, opinions, comments, and expectations about emerging technology on Twitter. These tweets feature Twitter users' objective description of emerging technologies (e.g., Twitter users' sense of emerging technologies) as well as corresponding comments (e.g., Twitter users response to emerging technologies), and expectations. The attribute characteristics of different Twitter users and changes in their behavior over time also are recorded in the Twitter platform. Through the data mining of tweets and Twitter users' attribute characteristics, we can study change characteristics of Twitter users' sense of, response to, and expectations for emerging technologies. This information may help us to better understand the emergence of emerging technologies from an environmental perspective, and could be crucial to provide intelligent assistance and support for government and enterprise R&D strategic planning to seize future business opportunities.

Some previous studies exploring the evolutionary path and development trends of emerging technologies are based on patents alone from a technological perspective, whereas few scholars have combined information from both Twitter data and patents to monitoring the emergence of emerging technologies and to identify future changes trends from a technical-environmental perspective. 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 trend (Ena et al., 2016). Thus, the analysis of only patents is not sufficient to fully understand the evolutionary path and development trends of emerging technologies. Through comparative analysis of the technical information contained in patents and the awareness information of public's sense of and response to emerging technologies contained in Twitter, we may better reveal and understand the emergence of emerging technologies and gain insight into development trends by the analysis of gaps between patents and Twitter data. Therefore, this paper proposed a framework that used patent analysis and Twitter data mining to monitoring the emergence of emerging technologies and to identify future changes trends of emerging technologies from a technical-environmental perspective. We selected perovskite solar cell technology as the case study against for the rapid development of the field of photovoltaic technology.

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 and concludes the paper.

2. Literature review

2.1. Patent analysis

Patent documents are considered to be a fruitful source of data for understanding technological trends (Noh et al., 2015). Patent analysis has been applied to analyze the intelligence information of the technology, and it has been considered as a vital tool for identifying and forecasting technological trends (Yoon and Kim, 2011). Thus, various patent analysis approaches have been developed to help researchers and practitioners obtain intelligence information about the technology from patent documents to support R&D strategies (Yoon et al., 2013). Among these patent analysis approaches, text mining has been widely used because of its ease of use and effectiveness to analyze the technological intelligence. Many scholars have applied a text mining approach to track the history of technology development and to identify and monitoring new trends in technological development (Choi et al., 2011; Lee et al., 2008; Wang et al., 2015; Wu and Leu, 2014; Yoon and Kim, 2012).

Patent text mining not only gives researchers access to technical information in patent documents (Madani and Weber, 2016) but also offers a more complete picture of the technological emergence and

[1] As a sociological concept, social awareness refers to a category of spatiotemporally tagged big data that provides an observatory for the sociability and social behaviors of human beings, such as human beings' activity and movement and sense of and response to social situations (Liu et al., 2015; Xia et al., 2015). People's awareness of social situations includes two meanings: sensing the real world and response to it (Pentland, 2005).

development process. We can better understand the path of technological evolution by analyzing the progression of technical topics hidden in the patents (Callon et al., 1991; Lee et al., 2008; Wu and Leu, 2014) and identify the technological development trends (Wang et al., 2015). Topic-based clustering is an interesting text mining method that has been applied to analyze technology trends (Chen et al., 2017). 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 used the Lingo algorithm to generate technical topics from patents to monitoring the path of technological evolution and the development trends of emerging technology.

2.2. Twitter data mining

Twitter is an online social network that allows users to send and receive personal updates from other contacts (texts of up to 140 characters, known as "tweets") (Daniel et al., 2017). Because of the mobility, content sharing, simplicity, and real-time nature of the Twitter platform, the members of the public can express their opinions and comments using this platform in real time, for example, customers can follow products and services and can discuss them in Twitter (Webster, 2010). These opinions and comments can be used to raise public awareness to help the government and enterprises understand the views of the public. Therefore, tweets are an important resource to study public awareness.

Researchers and practitioners can access Twitter data using Twitter Application Programming Interface (API) (Twitter, 2013). Search and streaming APIs allow them to collect Twitter data using different types of queries, including keywords and user profiles, which has offered them opportunities to access the data needed to analyze challenging problems in diverse domains (Chae, 2015). Thus, many researchers and practitioners have begun to focus on Twitter data mining to obtain more research value and business value from this research. Jansen et al. (2009) studied tweets as a type of electronic word of mouth to share consumer opinions about brands, and they found that the potential commercial value of tweets can be extracted through this type of data. Dang et al. (2016) proposed an early detection method for emerging topics based on dynamic Bayesian networks in Twitter. The experimental results demonstrated that their method is effective and capable of detecting emerging topics on Twitter one to two hours earlier than other methods. Daniel et al. (2017) identified popular events in the financial community by sentiment analysis of the tweets published by specific financial communities. Twitter also can be used to predict event trends. Skoric et al. (2012) applied the words and word frequency analysis approach to analyze public opinion in Twitter to predict the Singapore 2011 election results. Song and Kim (2013) used keyword analysis to analyze public comments contained in Twitter to predict the Republic of Korea's general election. Oliveira et al. (2017) presented a method to assess the value of tweet data, showing that tweet data can be used to predict the stock market behavior.

Twitter users can post their opinions, comments, and expectations on emerging technologies in Twitter. Their profile data also are recorded in the Twitter platform. In this paper, we used Twitter data mining to analyze Twitter users' sense of, response to, and expectations for emerging technology, and we examined how Twitter users' profiles change over time to track the emergence of emerging technology and to gain insight into the future changes trends of these emerging technology.

2.3. Sentiment analysis

Sentiment analysis (also called opinion mining, review mining, or attitude analysis) is the task of detecting, extracting, and classifying opinions, expectations, and emotional reactions about different topics, as expressed in textual input (Ravi and Ravi, 2015). Sentiment analysis has become one of the most popular research approaches used for practical applications and academic research. Numerous examples of practical applications of sentiment analysis range from consumer buying behavior forecasting (Doubleclick, 2005), through movie reviews analysis (Pang et al., 2002) to social networks discussions analysis (Condliffe, 2010).

Behavioral finance and investor sentiment theory have firmly established that investor behavior can be shaped by whether they feel optimistic or pessimistic about future market values (Bollen and Mao, 2011). Public attitude and responses to an event that are shared on social media can affect other people's behavior, which in turn may have a significant impact on the outcome of an event. Therefore, analysis of public sentiment regarding an event can contribute to identify hot topics (Daniel et al., 2017) and predict event trends (Bollen and Mao, 2011). A lot of emotional information is contained in tweets. The sentiment analysis of tweets can perceive market value and forecast the behavior of investors (Oliveira et al., 2017). Thus, many researchers and practitioners have begun to focus on sentiment analysis of tweets to generate more research value and business value from this information. Twitter messages have many unique attributes, however, which makes it difficult to conduct sentiment analysis. One of these attributes is that the maximum length of a Twitter message is 140 characters, and another attribute is that the frequency of misspelled words and slang in tweets is much higher than in other online domains (Go et al., 2009). Thus, some researchers have developed various sentiment analysis methods based on these unique attributes of tweets. Go et al. (2009) presented machine learning algorithms to classify the sentiment of Twitter messages using distant supervision. Tang et al. (2014) proposed to build large-scale sentiment lexicon from Twitter using a representation learning approach. The sentiment analysis methods presented by Go et al. (2009) and Tang et al. (2014) have high accuracy in the analysis of tweets.

In this paper, we used sentiment 140 which is a Twitter sentiment analysis tool to mine Twitter users' emotional reaction to emerging technology contained in the tweets and to study the changing characteristics of Twitter users' expectations of emerging technology over time. This analysis of Twitter users' expectations and sentiments may help us to gain insight into the future development trends of emerging technology.

2.4. Author-topic over time model

Latent Dirichlet allocation (LDA) is a topic model that uses random mixtures of latent topics to represent a document, and each topic is characterized by a distribution of words. LDA is a generative probabilistic model for the collection of discrete data, and it can be used to uncover latent topics in a large-scale document (Blei et al., 2003). LDA has proven to be an effective unsupervised learning method for uncovering implicit topics among documents (Chen et al., 2017). These documents, however, contain a lot of external attribute information in addition to content such as time and authors. The ability to identify this external attribute information is important for revealing a tacit knowledge about a document. On the basis of this LDA model, many researchers have developed a derived topic model by introducing corresponding external attribute information.

Because LDA does not consider time parameters, it cannot effectively identify possible changes in the topic of the document over time (Wang and Mccallum, 2006). Therefore, researchers have developed various LDA-based time topic models based on LDA and time signals. Wang and Mccallum (2006) proposed a topic over time (TOT) model to uncover changes in the latent topics across a huge number of documents over time. Researchers have combined author information with LDA models to develop author-topic (AT) models. With the use of an appropriate AT model, we can assess the research topics of authors and uncover which authors are likely to have written documents similar to the observed document as well as which authors published similar articles. The AT model can effectively identify topics that are relevant to different documents as well as identify which topics are of interest to which authors (Rosen-Zvi et al., 2004). AT models, however, are devoted to discovering static latent topics and authors' research interests, and they combine only single external attribute information-author information with LDA. The authors' areas of research interests are not static, but rather they change over time. To analyze this dynamic of the evolution of authors' research interests, Xu et al. (2014) proposed an author-topic over time (ATOT) model based on the AT and TOT models. The ATOT model combines the topics, the authors, and the time information of the document to form a three-dimensional analysis model to study the evolution of authors' research interests from a dynamic perspective (Xu et al., 2014). This model not only reveals the relationships between authors and topics but also uncovers the changing patterns of authors' research interests over time.

The AT model has high accuracy in topic extraction of short text (Hong and Davison, 2010), and the ATOT model is suitable for analyzing tweets. Therefore, in this paper, we used the ATOT model to analyze Twitter users' profile and the topics of Twitter users' tweets related to the emerging technology. First, we applied the ATOT model to mine the topics of twitter users' tweets to uncover Twitter users' sense of and response to emerging technology. Then, we constructed a user-topic model to study the changing patterns of users' attribute characteristics and corresponding interest in topics over time. With the analysis of Twitter users' profiles and tweets, we mined information about who pays attention to emerging technology, the topics on which users focused, and the changing patterns in their topics of interest about the emerging technology over time.

3. Methodology

Patent text mining provides researchers access to technical information in patent documents (Madani and Weber, 2016) and offers a picture of the emergence of technology as well as the development process. So we can monitoring the emergence and development process of emerging technology by analyzing the technical themes hidden in patent documents over time from a technological perspective. However, patents have a time lag and provide limited perspectives on the multifaceted phenomenon of emerging technologies, so we need to pay more attention to the use of up-to-date data that reflect the latest advances in emerging technologies as we try to identify the development trends of emerging technologies. Twitter users post their opinions about emerging technologies in the Twitter platform. The huge volume of opinions, expressed in real time, can be considered as the up-to-date data that reflect the latest advances of emerging technologies. When automatically extracting these opinions, we wanted to discriminate based on user demographics because the ratio of positive and negative opinions differed depending, for example, on age, gender, residence area (Ikeda et al., 2013), and professional types. In the emergence and development process of emerging technology, different types of professionals who use social media (e.g., Twitter users) focused their attention of different aspects of emerging technologies. So, in this paper, we focused on an analysis of the type of professional user when we examined these opinions.

This paper proposed a framework that used patent analysis and Twitter data mining to monitoring the emergence of emerging technologies and to identify future changes trends of these emerging technologies. The framework is illustrated in Fig. 1. The steps of the framework are as follows:

3.1. Stage 1. Patent analysis to monitoring the emergence and trends of emerging technology

3.1.1. Step 1

Retrieve patent data. We used the Derwent Innovations Index (DII) databases as the data source to collect data and used the search query related to the subject of the study (i.e., perovskite solar cell technology) to download the relevant patents. We obtained a collection of patents related to perovskite solar cell technology.

3.1.2. Step 2

Preprocess the obtained patent data. We needed to preprocess the patents set obtained from Step 1. For example, we removed patents related to dye-sensitized cells which used perovskite as a dye. The preprocessed data were divided by year and the cleaned patents were converted into text format compatible for text mining.

3.1.3. Step 3

Cluster topics with the Lingo algorithm. We applied the Lingo algorithm to cluster the cleaned patents obtained from Step 2. Lingo algorithm is performed in three steps. First, we used a modified version of semantic hierarchical clustering (SHOC) algorithm to discover phrases and single terms. Second, the vector space model (VSM) and singular value decomposition (SVD) were applied to induce cluster labels. Third, we applied the classic cosine distance to calculate the cosine similarity of the clustering label and input documents. If the similarity exceeds a predefined threshold, the documents were allocated to the corresponding clusters. Two Perovskite solar cell technology domain experts who have focused on perovskite solar cell technology studies for more than 5 years were invited to filter the clustering results to better reflect the technology topics. To refine the correlation of extracted topics, these topics were rendered with the Carrot2 workbench for visualization, which is an open source software that includes several text clustering algorithms and integrated a variety of visualization tools (Saracoglu, 2015). The visualization results can show every topic's connections to other topics, which indicate their correlations. Then these extracted topics were clustered based on their correlations and named by domain experts.

3.1.4. Step 4

Construct the evolution map of emerging technology based on patents. On the basis of the results of topic clustering obtained in Step 3, we constructed an evolution map of perovskite solar cell technology based on the patents with the help of domain experts.

3.2. Stage 2. Twitter data mining to monitoring the emergence and trends of emerging technology

3.2.1. Step 1

Retrieve Twitter data. We crawled and collected data using the search query related to the study subject (i.e., perovskite solar cell technology) from the Twitter platform by using Twitter Search API, and we downloaded the related tweet data locally as a data resource.

3.2.2. Step 2

Preprocess the obtained Twitter data. We used a text content deduplication method to preprocess the tweet data, removing duplicate tweets and cleaning the tweet text of redundant information. Then we classified the cleaned tweet text data according to time.

3.2.3. Step 3

Construct words document distribution. We combined Natural Language Toolkit (NLTK) and Stanford NLP to analyze tweet texts. We installed Standford NLP under NLTK, and called the Standford Tokenizer class contained in Python NLTK to tokenize tweets obtained from Step 2. Meanwhile, we eliminated the noise data, including

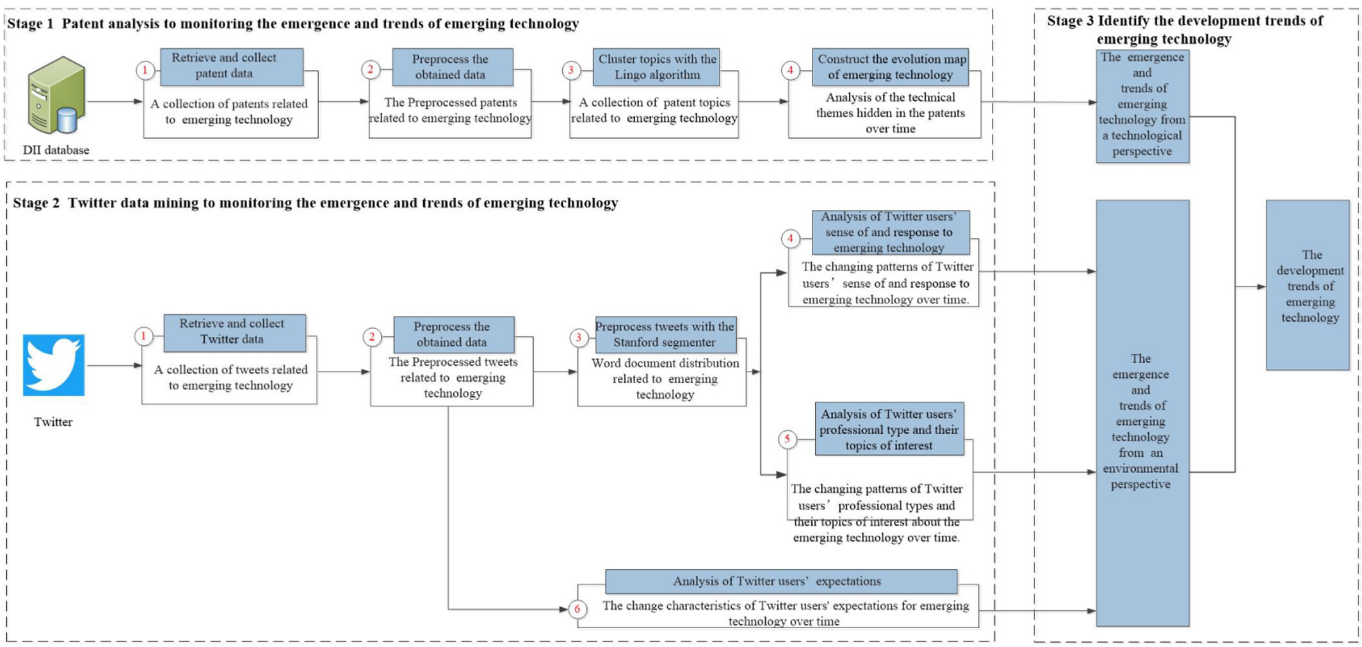

Fig. 1. Framework to identify and monitoring development trends of emerging technologies.

unifying singular and plural, and removal of stop words. The purpose of this step was to obtain words document distribution related to perovskite solar cell technology.

3.2.4. Step 4

Analysis of Twitter users' sense of and response to emerging technology. We employed the ATOT model to analyze the preprocessed tweets obtained from Step 3, and obtained the topic-feature words probability distribution and topic-users probability distribution. We used the word processing of the natural language processing module of TDA to process keywords contained in topics, and we obtained the parts-of-speech (POS) of every keywords. According to the POS of keywords, we extracted nouns, adjectives and adverbs from each topics for each year to obtain awareness keywords that reflected the Twitter users' sensing of and response to emerging technology. By analyzing the changing patterns in these awareness keywords over time, we studied the changing characteristics of Twitter users' sense of and response to the emerging technology.

3.2.5. Step 5

Analysis of Twitter users' professional type and topics of interest. According to the user's profile, we analyzed the Twitter users' professions that were identified by the ATOT model with the help of experts. We then analyzed on which topics these different types of professionals focused, as well as the changing patterns in their topics of interest about emerging technology over time.

3.2.6. Step 6

Analysis of Twitter users' expectations. In this step, we used sentiment analysis to mine Twitter users' emotional reactions to emerging technology contained in the tweets. We studied the changing characteristics in Twitter users' expectations for emerging technology over time.

3.3. Stage 3. Identify the development trends of emerging technology

After we obtained the results of the evolutionary path and trends of the emerging technology based on patent analysis; identified Twitter users' sense of, response to, and expectations for emerging technology; and attained the professional types of Twitter users and their topics of interest, we carried out the differences analysis of these results to gain insight into the future development trends of the emerging technology. First, we made a comparison analysis of the results of patent analysis and Twitter users' sense of and response to emerging technology. Then, we compared these gaps and combined the results of changing patterns in Twitter users' response to emerging technology with Twitter users' expectations for the technologies and changes in the professional types of Twitter users and their topics of interest over time to identify development trends of emerging technology.

4. Case study

In this paper, we selected perovskite solar cell technology as the case study to illustrate our proposed methodology. Perovskite solar cell technology is one type of emerging solar cell technology in the field of solar photovoltaic technology. With the rapid increase of photoelectric conversion efficiency, perovskite solar cell technology has been called the “new hope in the photovoltaic field” (Wei et al., 2014). The journal *Science* selected it as one of the ten most important scientific and technological advances in 2013. *Nature* magazine called it one of the most anticipated breakthroughs in science and technology in 2014. In 2016, the World Economic Forum (WEF) listed it among the top ten emerging technologies that will change human life. With the rapid development of perovskite solar cell technology, it is considered to have the potential to have a disruptive impact on the photovoltaic industry and would transform the existing energy industry. Therefore, monitoring the emergence of perovskite solar cell technology is essential to understand and detect changing trends early on the development of this new technology. This information is crucial for government R&D strategic planning, social investment, and enterprises practices.

4.1. Data collection

In this paper, we used the DII databases as the patent data source for data collection. We uses the terms ((perovskite*) and ((solar cell*) or (solar cells*) or (photovoltaic cell*) or (photovoltaic cells*)) as the query to search the patent data from DII and retrieved 634 issued patents from the database from 2009 to 2016. The search was done on March 20, 2017.

We crawled and collected tweets mentioning the perovskite solar

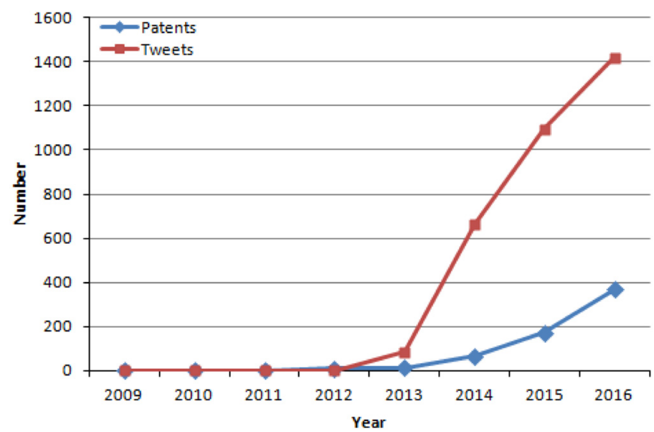

Fig. 2. Annual number of patents and tweets related to perovskite solar cell technology.

cells from May 11, 2009, until March 23, 2017, using the Twitter API to retrieve the following queries: (perovskite) AND ((solar cell) OR (solar cells) OR (photovoltaic cell) OR (photovoltaic cells)). The resulting data set consisted of 3646 tweets with the keywords about perovskite solar cells sent by 663 users. The profiles of these users on Twitter also had been crawled. We downloaded Tweet ID, user ID, user name, and tweet and release time locally as a data resource. We then used a text content deduplication method to process tweets and cleaned redundant tweet information, and we obtained 3268 tweets. Finally, we classified the preprocessed tweets according to time slices. The annual number of patents and tweets related to perovskite solar cell technology are shown in Fig. 2.

As evident in Fig. 2, patents related to perovskite solar technology first appeared in 2009, and tweets related to perovskite solar cell technology first appeared in 2010. The number of patents and tweets grew rapidly each year, in particular, the number of patents and tweets increased rapidly in 2013 as the power conversion efficiency of perovskite solar cell improved significantly. The number of patents increased exponentially in the period from 2013 to 2016 (2983% increase for patents). This high rate of patents showed that R&D for perovskite solar cell technology was active in the past three years. At the same time, the numbers of tweets showed a linear growth trend to about 1419 annual filings in 2016 (a 1569% increase for tweets). We interpreted the growing number of tweets to be significant, and it demonstrated that the development of perovskite solar cell technology generated a significant level of public attention in the past three years.

4.2. Data analysis

The data analysis consisted of two parts: (1) patent analysis to monitoring the emergence and trends of perovskite solar cell technology, and (2) Twitter data mining to monitoring the emergence and trends of perovskite solar cell technology.

4.2.1. Patent analysis to monitoring the emergence and trends of perovskite solar cell technology

We first used the Lingo clustering algorithm to extract technical topics from the annual patents related to perovskite solar cell technology. Second, we invited two perovskite solar cell technology experts to review these technical topics and we collected these extracted technical topics. Then these extracted technical topics were clustered based on their similarity, and named by experts. Finally, based on the results of topics clustering, we analyzed the technical themes hidden in the patents over time and constructed an evolution map of perovskite solar cell technology with the help of domain experts. From the evolution map, we gained a better understanding of the emergence and development of the perovskite solar cell technology from a technological perspective.

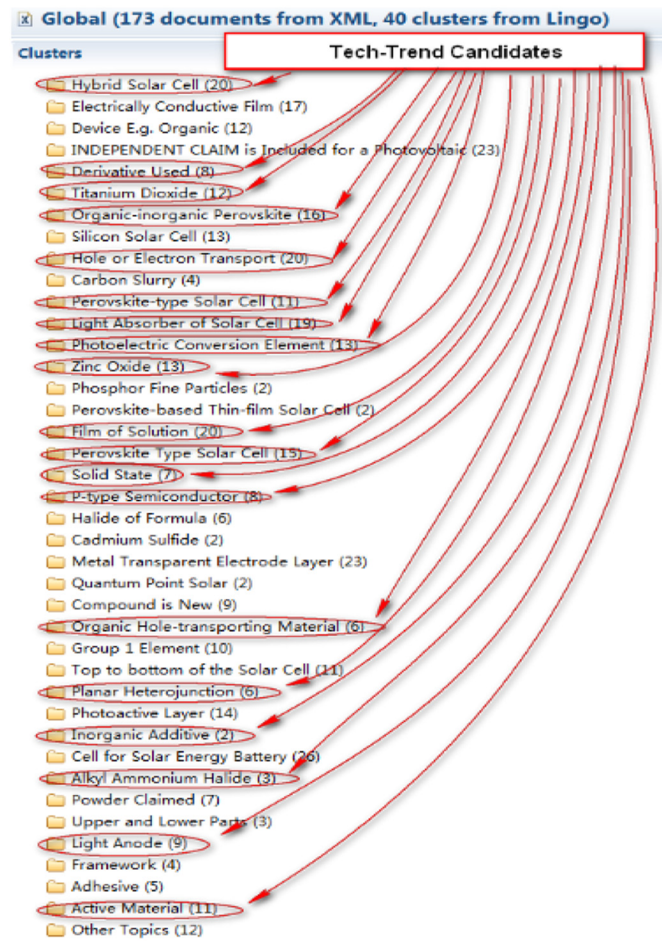

Fig. 3. Carrot clustering results of patents in 2015.

4.2.1.1. *Topics clustering.* We processed the annual patents and converted the preprocessed data into an XML format compatible with Carrot2 software. Carrot2 is an open source software project that includes several text clustering algorithms and integrates a variety of visualization tools (Saracoglu, 2015). To get better clustering results, we needed to set the control parameters of the Lingo algorithm in the Carrot2 software. We set minimum cluster size = 2, cluster merging threshold = 0.7, and size-score sorting ratio = 1.0. We conducted the selection of the control parameters for clustering algorithms using a variation and subsequent expert validation of the results in relation to the specified control parameters (Ena et al., 2016). 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, such as perovskite and charge, and two domain experts who have focused on perovskite solar cell technology studies for more than 5 years were invited to screen the clustering results, and obtained annual technical topics. Fig. 3 shows a part of the clustering results in 2015. The red circle indicates the topics selected by the perovskite solar cell technology experts. Because the related patent data from 2009 to 2011 was too small, we made extraction results only from 2012 to 2016.

Based on domain experts' knowledge, we found that there are interconnections between selected topics. To refine the interconnections of selected topics, these topics were rendered with the Carrot2 workbench for visualization. Fig. 4 shows part of the visualization results from 2015. In Fig. 4, different topics are represented by different colors, and a circle represents a document. The single number in brackets

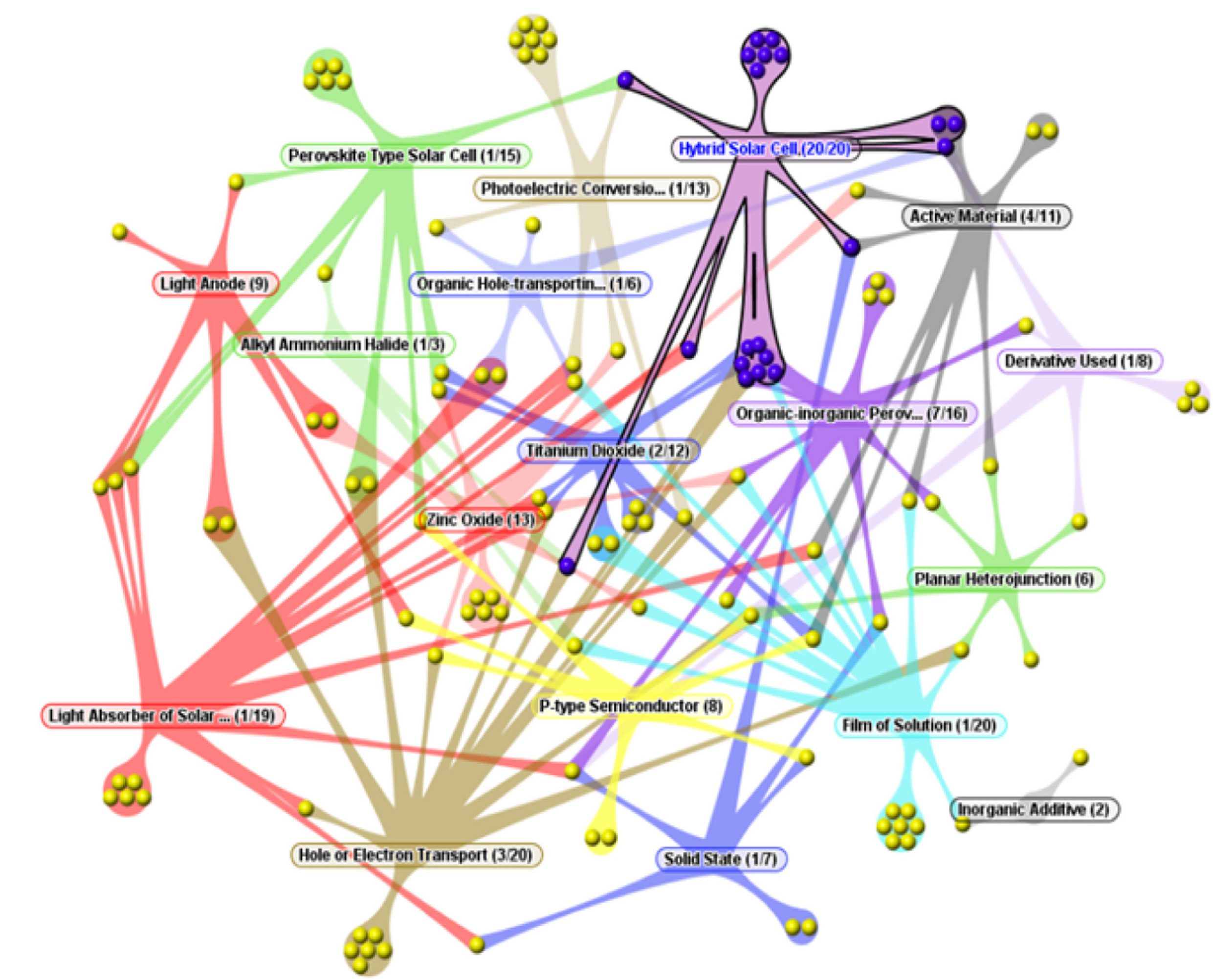

Fig. 4. Carrot2-workbench Aduna Cluster Map visualization (2015).

Table 1
The results of annual topics harmonization.

2012	2013	2014	2015	2016
Organic perovskite materials (2)	Solid organic hole transfer material (2)	Organic Inorganic hybrid materials (8)	Organic Inorganic hybrid materials (36)	Organic perovskite materials (111)
Electrolyte (2)	Mesoporous structure(2)	Organic electron transporting materials (8)	Solid organic hole transport material (6)	Organic Inorganic hybrid materials (65)
Metal Oxides (2)		Metal Oxides (12)	Inorganic hole transport material (8)	Inorganic materials(58)
		Mesoporous structure(3)	Metal Oxides (2)	Metal Oxides (104)
		The solution method (6)	Insulating material (13)	Insulating material (21)
			Organic electron transporting materials (19)	Organic electron transporting materials (17)
			Heterojunction structure(6)	Inorganic hole transport material (51)
			The solution method (6)	Heterojunction structure(21)
				Three-dimensional structure(46)
				The solution method (26)
				The deposition method(52)
				The evaporation method(2)

represents the number of documents addressing the topic, for example, the topic "Light Anode" is addressed by 9 documents. The double numbers in brackets have different meanings, the front number represents the number of documents shared with other topics, and the back number represents the total number of documents addressing the topic. For example, the topic "Organic-Inorganic perovskite" is addressed by 7 shared documents with the selected topic "Hybrid Solar Cell", and the topic "Organic-Inorganic perovskite" is addressed by a total of 16 documents. The shared circle between topics indicates their correlations to other topics (Ena et al., 2016). The more circles shared by two topics, the stronger the relationships between the two topics. These topics were clustered based on their correlations and harmonized by domain experts. We clustered the topics which have the higher correlations, and harmonized the clustered topics with the help of domain experts. The results of the topic harmonization are shown in Table 1 of Appendix A. The annual topic harmonization results are

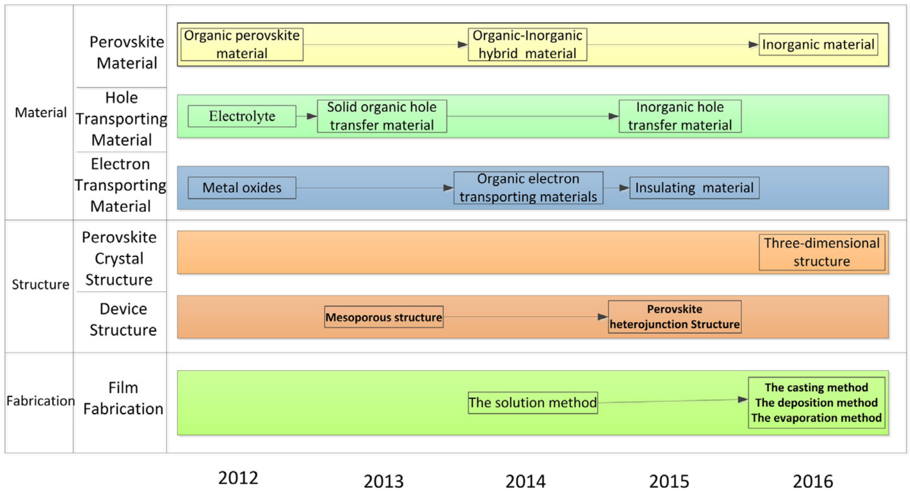

Fig. 5. Evolution map of perovskite solar cell technology based on patents.

given in Table 1. In Table 1, the values in the brackets represent the number of documents addressing the noted topic.

4.2.1.2. Analysis of the evolutionary path of the technology based on patents. To gain a better understanding of the evolutionary path of perovskite solar cell technology, we constructed a map of its evolution based on the topic clustering results obtained from patents (as given in Table 1 of Appendix A) and relying on domain experts' knowledge. The evolution map of the perovskite solar cell technology is shown 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, which were named on the basis of expert opinion. The horizontal axis represents the first time a topic appeared. The different colors represent different technical topics. From these technical topics, we could understand the emergence and development process of perovskite solar cell technology by analyzing variations in the technical topics for each layer over time.

As shown in Fig. 5, the main technical topic in the patents related to perovskite solar cell technology was the perovskite material in 2012. In 2013, two new technical topics appeared for the first time, that is, solid organic hole transfer material and mesoporous structure. In 2014, three new technical topics appeared for the first time, including the solution method, organic materials, and organic-inorganic hybrid material. In 2016, many new technical topics appeared for the first time, including inorganic material, three-dimensional structure, and the deposition method. From the emergence of the new technical topics in different years, it is evident that the applied research moves from materials to structure and then to fabrication, which also shows the technology evolution paths and development trends of perovskite solar cell technology from 2012 to 2016.

At the material level, three types of materials are related to perovskite solar cell technology, that is, perovskite material, hole transporting material, and electron transporting material. The research of perovskite material initial focused on organic materials in 2012. Organic-inorganic materials have better heat resistance than organic materials, which endows the perovskite solar cell with great performance and long life. Thus, the research of perovskite material gradually gave way to organic-inorganic materials in 2014. The inorganic perovskite materials generated a high level of research attention in 2016, because they were considered to be more economical and environmentally friendly compared with organic-inorganic materials. From the shift in these technical topics, we found that the evolution paths of perovskite material gradually shifted from the initial organic materials in 2012 to organic-inorganic perovskite hybrid materials in 2014 and then to inorganic materials in 2016.

The research on hole-transporting materials initially focused on liquid electrolyte material in 2012. Because solid organic hole-transporting materials can improve the stability of perovskite solar cells compared with liquid electrolyte material, in 2013, researchers paid greater attention to these solid organic hole-transporting materials. The cost of the solid organic hole-transporting material is relatively high, so researchers looked for a low-cost hole-transporting materials, and in 2015, they paid attention to inorganic hole-transporting materials with low cost. From the shifting of the technical topics, we found that the evolution paths of hole-transporting material gradually changed from the initial liquid electrolyte material in 2012 to solid organic hole-transporting material in 2013, and then to solid inorganic hole transporting material in 2015.

The research on electron-transporting materials initially focused on metal oxide in 2012. Because organic materials can significantly improve the power conversion efficiency of perovskite solar cells, researchers paid greater attention to organic materials in 2014. Insulation materials have a significant electron transport property that allows the photocurrent of these perovskite solar cells to reach saturation more quickly and fully, and this greatly improves efficiency. Therefore, in 2015, researchers paid attention to the insulation materials. From the shifting of the technical topics, we found that the evolution paths of electron-transporting material gradually changed from the initial metal oxide in 2012 to organic materials in 2014 and then to insulation materials in 2015. From the above analysis of the shift in the technical topics concerning perovskite material, hole-transporting material, and electron-transporting material, we found that we can better understand the evolution paths of perovskite solar cell technology materials, and materials that are low cost, stable, and environmentally friendly and that have improved conversion efficiency would represent development trends of perovskite solar cell technology.

At the structural level, two main types of structures are related to perovskite solar cell technology, that is, device structure and perovskite crystal structure. The research on the device structure initially focused on mesoporous structure in 2013. The planar heterojunction structure can increase the flexibility of these devices compared with a mesoporous structure and can simplify the structure and interface. The planar heterojunction structure contributes to the large-scale production of perovskite solar cells, and at the same time, it also offers the possibility to prepare flexible devices with high efficiency. Thus, researchers paid more attention to the planar heterojunction structure in 2015. Research on this device structure gradually changed from the initial mesoporous structure in 2013 to a planar heterojunction structure in 2015. The perovskite crystal structure received attention in 2016 as the three-dimensional structure appeared in this year, which indicated that researchers began to focus on the perovskite crystal structure. This shift in technical topics about device structure and perovskite crystal structure indicated that structures that are simple and flexible and that have improved conversion efficiency represent future development trends of perovskite solar cells technology.

At the fabrication level, research on fabrication initially focused on the solution method to prepare films of perovskite solar cell in 2014. The coating method, the evaporation method, and the deposition method generated high levels of attention in 2016. The coating method overcomes the solution method's defect, which needs a higher annealing temperature. The evaporation method mainly has been applied to the preparation of planar heterojunction solar cells, and it makes it possible to produce large area, low-cost perovskite solar cells. From the shifting of the technical topics, we found that the evolution paths of film fabrication of perovskite solar cells developed from the solution method in 2014 to the coating method, the evaporation method, and the deposition method in 2016. These indicated that fabrications that are a simple and economic preparation process represent future development trends of perovskite solar cells technology.

4.2.2. *Twitter mining to monitoring the emergence and trends of perovskite solar cell technology*

We cleaned the preprocessed tweets obtained from the section 4.1 with NLTK and Standford NLP, and we employed the ATOT model to analyze the cleaned tweets. In this paper, the ATOT model parameters were set to $\alpha = 50/K$, $\beta = 0.1$, and Gibbs sampling was run for 2000 iterations in the analysis of tweets. The K indicates number of topics, α indicates Dirichelet priors hyperparameter to the multinomial distribution of topics specific to the author, and β indicates Dirichelet priors hyperparameter to the multinomial distribution of words specific to the topic K (Xu et al., 2014). After the repeated experiments of dividing topics, we found that feature words have a higher probability and a better discrimination when the number of topics was set to 10, and β was fixed at 0.1 according to words appearing times in texts. Finally, we obtained the topic-feature words probability distributions and topic-users probability distributions. Table 2 shows partial results of the co-occurrence probability of keywords and users corresponding with topics in 2016. In Table 2, we extracted the top 10 users from the topic-users probability distributions.

As for the output of the ATOT model (partial results are shown in Table 2), we found that there are correlations between the 10 topics each year, such as the topics related to efficiency, some topics focus on materials to improve efficiency, meanwhile, other topics focus on structure to improve efficiency. Thus, we needed to merge these related topics when we wanted to obtain clear and accurate topics. In order to merge these related topics, we calculated the similarity between these related topics. We represented every topic in term of a vector of terms (Liu and Chen, 2013), and the similarity of two related topics was calculated by cosine similarity. The closer the cosine value is to 1, the more similarity exists between the two topics. We merged related topics based on the similarity threshold set by domain experts based on the similarity of topics. Finally, we divided these topics into seven classifications, and the topic name was labeled by the high probability feature word. The seven topics included perovskite solar cells (PSC) material, PSC structure, basic research, efficiency, PSC stability, PSC application, and PSC commercialization. Then, we divided the seven topics into three aspects of the linear model of technology-oriented innovation processes, science, technology, and industry (Shibata et al., 2010), by analyzing the merged topics. These three aspects are listed as follows: the technology R&D aspect, the technology application aspect, and the technology commercialization aspect.

4.2.2.1. *Analysis of Twitter users' sense of and response to perovskite solar cell technology*. Among the tweets, there are Twitter users' objective descriptions of perovskite solar cell technology (e.g., Twitter users' sense of perovskite solar cell technology) as well as corresponding

Table 2
The partial results of the co-occurrence probability of keywords and users correspond with topics in 2016.

Topic	Keywords	Probability	Users	Probability	Topic	Keywords	Probability	Users	Probability	Topic	Keywords	Probability	Users	Probability
Topic 0 th	Cells	0.164	Joelalexsmith	0.176	Topic 1 th	Future	0.047	MediosTec21_Vva	0.191	Topic 2 th	Efficiency	0.117	Osadirect	0.364
	solar	0.123	shahzada80	0.161		energy	0.035	solarcurator	0.149		efficient	0.040	coolingZONE11	0.231
	perovskite	0.014	ConingsBert	0.157		scientists	0.028	PawelRzym	0.145		new	0.038	pvbuzzmedia	0.192
	technologies	0.013	minitaxi	0.155		new	0.016	AsiaMaterials	0.145		cell	0.021	SemiEngineering	0.182
	paper	0.009	EPFL_en	0.155		cells	0.012	talius	0.145		perovskite	0.016	NanoSchaft	0.171
	stable	0.008	azom	0.154		power	0.010	markcote873	0.143		material	0.015	KavliFoundation	0.162
	hit	0.006	SolarPowerAust	0.149		closer	0.010	TinaMCasey	0.141		researchers	0.014	OTEM2000toni	0.158
	news	0.005	AZoCleantech	0.149		next	0.010	MiradorLtd	0.140		crystals	0.013	Solar_Quotes	0.153
	greener	0.005	kevinmcintyre09	0.148		nanocrystals	0.009	KatrinAtWork	0.140		develop	0.011	mendotagroup	0.153
	rival	0.005	CECHR_UoD	0.147		world	0.009	FNGhadaki	0.140		scientists	0.011	CnvCurmudgeon	0.148
	percent	0.005				record	0.007				design	0.011		
	conversion	0.005				growth	0.007				materials	0.010		
	composition	0.004				perovskite	0.007				outperform	0.009		
	rubidium	0.004				renewables	0.006				development	0.008		
	generation	0.004				air	0.006				push	0.008		
	old	0.004				big	0.006				commercial	0.008		
	highest	0.004				science	0.006				research	0.006		
	great	0.004				panels	0.006				top	0.006		
	world	0.004				achieve	0.006				make	0.006		
	make	0.004				high	0.006				graphene	0.006		

comments (e.g., Twitter users' response to perovskite solar cell technology). The nouns, adjectives, and adverbs of these tweets can be used to reflect the themes and emotional tendencies of the tweets (Zhu, 2014). The tweets' subjects are mainly embodied in the form of nouns, and emotional tendencies are mainly embodied in the form of adjectives and adverbs. On this basis, in the process of analyzing the Twitter users' sense of and response to perovskite solar cell technology, we treated nouns, adjectives, and adverbs as awareness keywords, and we treated nouns as a reflection of Twitter users' sense of the technology and treated adjectives and adverbs to reflect Twitter users' response to the technology. Then we applied the natural language processing module of TDA to process keywords contained in topics, and we obtained the parts-of-speech (POS) of every keywords. According to the POS of keywords, we extracted nouns, adjectives and adverbs from each of the seven topics for each year to obtain awareness keywords that reflected the Twitter users' sense of and response to perovskite solar cell technology. By analyzing the changing pattern in these awareness keywords over time, we studied the changing characteristics of Twitter users' sense of and response to perovskite solar cell technology over time.

Because of different knowledge backgrounds, different types of professionals that use Twitter may pay attention to different aspects of perovskite solar cell technology. Therefore, according to the Twitter user profiles, we divided the users output by the ATOT model into eight classifications of professional types, which included researchers, magazines, public organizations and agencies, news outlets, solar power enterprises, consultation platforms, suppliers and distributors, and investors. The results are given in Table 2 of Appendix A.

To uncover the dynamic evolutionary process of Twitter users' sense of and response to perovskite solar cell technology, we constructed a topic-awareness-user evolution map based on the ATOT model. Because the number of related tweets from 2010 to 2012 is too small, we analyzed tweets only from 2013 to 2016. The result is shown in Fig. 6. In Fig. 6, the vertical axis represents the technology R&D aspect, the technology application aspect, and the technology commercialization aspect. The horizontal axis represents time. The different colors of the boxes represent different topics, and every topic is composed of the public's sense of and response to the technology, which corresponded with the awareness keywords. PSC is the abbreviation of perovskite solar cell technology.

As shown in Fig. 6, eight different professional types of Twitter users paid attention to seven different topics related to perovskite solar cell technology. Twitter users' sense of and response to perovskite solar cell technology have changed in the technology R&D aspect, the technology application aspect, and the technology commercialization aspect over time, which revealed the difference between the awareness content and characteristics of the awareness stage. In 2013, researchers, magazines, public organizations and agencies, news outlets, and solar power enterprises focused on the technology R&D aspect and focused on the material, structure, basic research, and efficiency topics related to perovskite solar cell technology. In addition to the technology R&D aspect, in 2014, Twitter users began to pay attention to the technology application aspect with significant developments in this technology. The consultation platform paid attention to the technology of the PSC material, basic research, efficiency, and PSC application; and the suppliers and distributors paid attention to the efficiency of the technology. Twitter users started to focus on the technology commercialization aspect in 2016, and investors began to pay attention to the development of this technology. From these shifts in the topics that the eight different professional types of Twitter users focused on, we found that the topics about perovskite solar cell technology that concerned Twitter users gradually developed from the technology R&D aspect in 2013 to the technology application aspect in 2014 and then to the technology commercialization aspect in 2016. This shift revealed that the perovskite solar cell technology has been developed rapidly in recent years and has the potential for commercialization. At the same time, Twitter users changed from researchers, magazines, public organizations and agencies, news outlets, and solar power enterprises in 2013 to include consultation platforms and suppliers and distributors in 2014, and then to include investors in 2015, which also indicated that the development of perovskite solar cell technology had generated a high level of public attention in recent years. On the basis of topics that Twitter users were focused on, we found that they continued to pay attention to the technology R&D aspect and the technology application aspect. In particular, Twitter users started to focus on PSC commercialization in 2016, which showed that the potential for technology commercialization was increasing.

To better understand the detailed changing characteristics of Twitter users' sense of and response to perovskite solar cell technology, we analyzed the changing characteristics of the awareness keywords contained in each topic Twitter users were concerned with over time. As can be seen from Fig. 6, for PSC material (topic I), the awareness keywords perovskite, lead, silicon, and tin appeared in 2013 and 2014; graphene appeared in 2015; and hybrid materials appeared in 2016. The changes in these awareness keywords showed that Twitter users' sense of PSC material shifted from conventional natural chemical materials in 2013 and 2014, gradually evolved to focus on new materials in 2015, and then shifted to synthetic new hybrid materials in 2016. Awareness keywords such as efficient, high, and excited appeared in 2013; power and stable appeared in 2014; and then potential, boost, develop, renewable, and improved appeared in 2015 and 2016. Changes in these awareness keywords showed that Twitter users' response to PSC material changed from the efficient development of materials in 2013, gradually evolved to focus on the stable development of materials in 2014, and then shifted to the development of recycled materials in 2015 and 2016. This shift in the awareness keywords concerning perovskite solar cell technology indicated that Twitter users' attention to PSC material initially focused on efficient materials, changed to stable materials, and then shifted to environmentally friendly materials.

As for PSC structure (topic II), the awareness keywords such as atom, metal appeared in 2013; atom, hybrids appeared in 2015; and nanocrystals, big, panels appeared in 2016. The change in these awareness keywords showed that Twitter users' sense of PSC structure shifted from the initial mesoporous structure in 2013 and 2015, and gradually evolved to focus on the planar heterojunction structure in 2016. The awareness keywords such as high, cheap, and conventional appeared in 2013; printed, flexible, and magic appeared in 2015; and growth, high, and power appeared in 2016. This change in awareness keywords showed that Twitter users' response to PSC structure initially focused on highly efficient and cheap structures in 2013, gradually evolved to focus on the flexible structure and the printable structure in 2015, and then changed to the new structure in 2016. This shift in awareness keywords concerning perovskite solar cell technology indicated that Twitter users' attention to PSC structure shifted from the efficient, cheap traditional mesoporous structure to the flexible, printable, highly efficient new structure.

As for basic research (topic III), the awareness keywords such as breakthrough, oxides, and scientist appeared in 2013; lead, tin, cost, and researchers appeared in 2014; and conversion, paper, and rubidium appeared in 2016. This change in awareness keywords showed that Twitter users' sense of basic research initially focused on the research of materials such as oxides in 2013, gradually evolved to focus on cutting the cost of materials in 2014, and then changed to research on improving efficiency in 2016. The awareness keywords such as cheaper and new appeared in 2013; cheap, recycling, promise, durable, and renewable appeared in 2014; and great, stable, and highest appeared in 2016. This change in awareness keywords showed that Twitter users' response to basic research changed from cheap and new materials in 2013; gradually evolved to focus on recycling materials, renewable materials, and cheaper cost in 2014; and then shifted to high efficiency in 2016. This shift in awareness keywords concerning perovskite solar

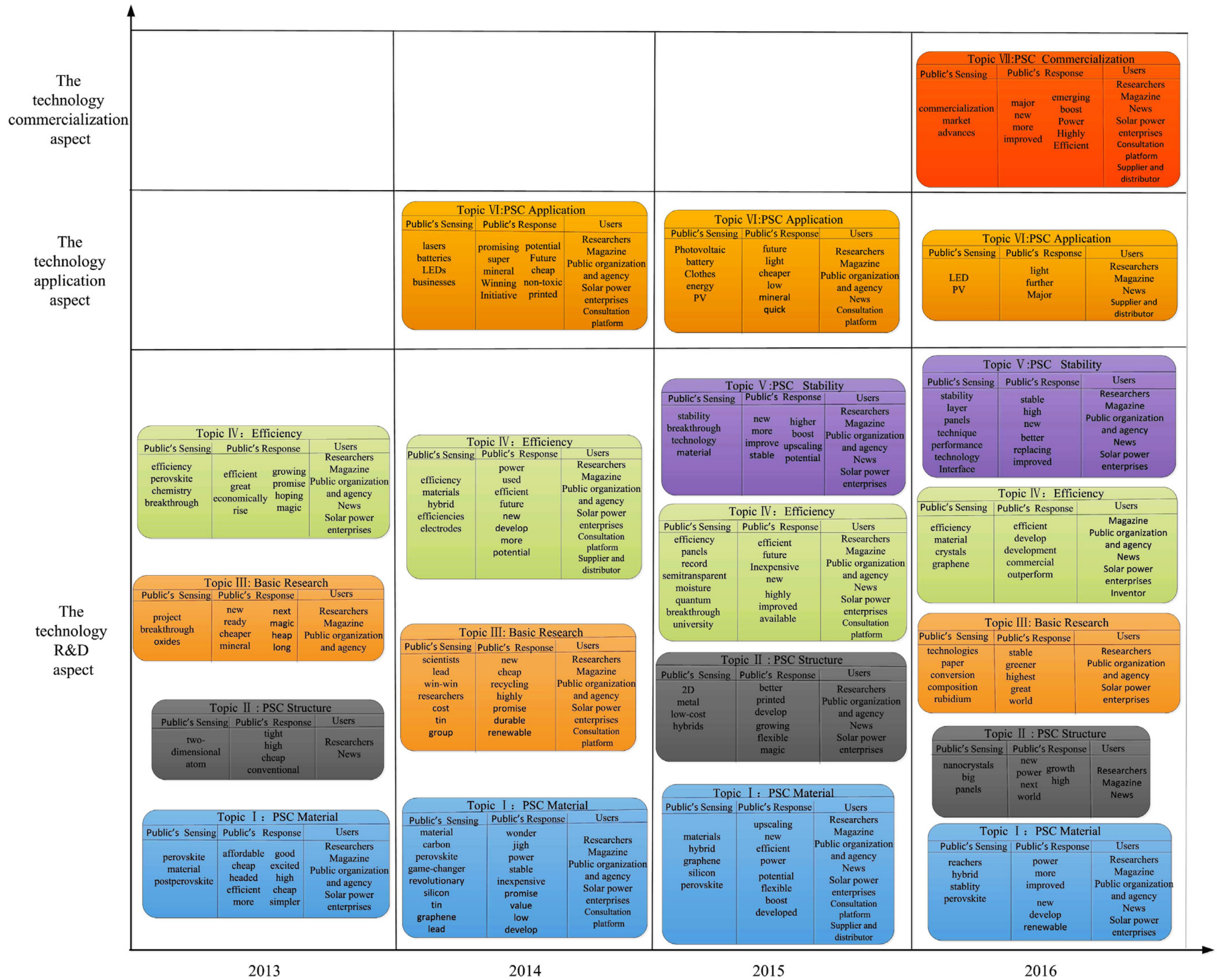

Fig. 6. The evolution map of Twitter users' sense of and response to perovskite solar cell technology.

cell technology indicated that Twitter users' attention to basic research shifted from the basic research of materials to the research of lower costs materials and then to research on the stability and efficiency of the technology.

As for efficiency (topic IV), the awareness keywords such as efficiency and perovskite appeared in 2013; materials and hybrid appeared in 2014; panels appeared in 2015; and crystals appeared in 2016. This change in awareness keywords showed that Twitter users' sense of efficiency initially focused on changing materials to improving efficiency in 2013 and 2014, and gradually evolved to focus on changing structure to improving efficiency in 2015 and in 2016. The awareness keywords such as efficient, great, growing, and promise appeared in 2013; future and potential appeared in 2014; inexpensive, highly, and improved appeared in 2015; and commercial and outperform appeared in 2016. This change in awareness keywords showed that Twitter users' response to efficiency shifted from the greatness and promise of efficiency in 2013 and 2014, and gradually evolved to focus on the commercial potential of efficiency in 2015 and in 2016. This shift in awareness keywords concerning perovskite solar cell technology indicated that Twitter users' attention to efficiency shifted from changing materials to improve efficiency in 2013 and 2014, and gradually evolved to focus on changing structure to improve efficiency in 2015 and in 2016. Twitter users believed that the efficiency of perovskite solar cells improved quickly and thus it has the potential for commercialization in the future.

As for stability (topic V), the awareness keywords such as stability, technology, breakthrough, and material appeared in 2015; and layer, panels, interface, and performance appeared in 2016. This change in awareness keywords showed that Twitter users' sense of stability initially focused on changing materials to improve the stability of the technology and gradually evolved to focus on changing layer, panels, and interface to improve the stability of the technology. The awareness keywords such as stable, improve, and upscaling appeared in 2015; and new, better, replacing, and improved appeared in 2016. This change in awareness keywords showed that Twitter users' response to stability shifted from the upscaling of technology stability in 2015 to the replacement and improvement of technology stability in 2016. This shift in awareness keywords concerning perovskite solar cell technology indicated that Twitter users' attention to stability shifted from changing materials to improve technology stability to changing the structure to improve technology stability. Twitter users believed that technology stability improved significantly.

As for PSC application (topic VI), the awareness keywords such as lasers, batteries, and LEDS appeared in 2014; photovoltaic and clothes appeared in 2015; and LED and PV appeared in 2016. This change in awareness keywords showed that Twitter users' sense of PSC application shifted from lasers, batteries, and LEDS in 2014 and gradually evolved to photovoltaic, clothes and photovoltaic in 2015 and 2016. The awareness keywords such as potential, promising, nontoxic, and printed appeared in 2014; cheaper and low appeared in 2015; and future and light appeared in 2016. This change in awareness keywords showed that Twitter users' response to PSC application changed from initially focusing on promising and nontoxic technology and gradually evolved to focus on low-cost technology. This shift in awareness keywords concerning perovskite solar cell technology indicated that Twitter users' attention to PSC application shifted from the initial focus on applying perovskite solar cells to lasers and LEDS and gradually developed to apply perovskite solar cells to wearable devices and the photovoltaic field. With this shift in the application field, Twitter users' awareness of and expectations for the technology's application changed from nontoxic to low cost.

As for PSC commercialization (topic VII), the emergence of awareness keywords such as commercialization, market, and advances indicated that Twitter users paid attention to the potential commercialization of perovskite solar cell and its future influence on the global photovoltaic market. The emergence of awareness keywords such as major, new, improved, boost, and emerging indicated that Twitter users had high expectations for the future developments of perovskite solar cell technology.

4.2.2.2. Analysis of Twitter users' professional type and their topics of interest. In this section, we analyzed the professional types of Twitter users identified by the ATOT model based on the crawling user's profile. We analyzed these different users' professional types and their focus on topics related to perovskite solar cell technology. We examined changing patterns in their interest in topics about the technology over time. The purpose of our analysis is to gain insight into the future changes trends of the perovskite solar cell technology. We mainly analyzed five types of Twitter users: researchers, news outlets, solar power enterprises, suppliers and distributors, and investors.

As can be seen in Fig. 6, in the period from 2013 to 2016, researchers continuously focused on the technology R&D aspect, such as PSC material (topic I), PSC structure (topic II), basic research (topic III), and efficiency (topic IV). Researchers paid attention to the PSC application since 2014 and paid attention to PSC commercialization starting in 2016. Researchers mainly engaged in the research and development of perovskite solar cell technology, and researchers possessed the latest developments and cutting-edge knowledge in this technology. The topics on which researchers focused changed from the technology R&D aspect to the technology application aspect, and then to the commercialization aspect, which indicated that the R&D of perovskite solar cell technology was active and that the technology has the potential for commercialization and great market prospects.

News outlets continuously focused on PSC structure (topic II) and basic research (topic III) in the period from 2013 to 2016. News outlets paid attention to the PSC application (topic VI) since 2014 and paid attention to PSC commercialization (topic VII) starting in 2016. News outlets can accurately reflect the research results and new developments in perovskite solar cell technology. The topics on which news outlets focused changed from the technology R&D aspect to the technology application aspect, and then to the commercialization aspect, which indicated the rapid development of and future trends in perovskite solar cell technology. Because of the objectivity, authenticity, and real-time characteristics of the news, it spreads quickly and has significant influence. Some of the descriptions of, comments about, and expectations for perovskite solar cell technology contained in news could have a significant impact on the behavior of perovskite solar cell technology stakeholders, and in turn, stakeholder behavior will affect the emergence and development of perovskite solar cell technology.

Solar power enterprises have focused on PSC material (topic I) and efficiency (topic IV) since 2013. Solar power enterprises paid attention to PSC stability (topic V) from 2015 to 2016, and solar power enterprises paid attention to PSC commercialization (topic VII) in 2016. Solar power enterprises mainly focus on solar power areas of technological innovation that can bring economic benefits, and solar power enterprises play an important role in bringing perovskite solar cell technology to market for commercialization. The topics on which solar power enterprises focused changed from the technology R&D aspect to the technology application aspect, and then to the commercialization aspect, which indicated that the application of perovskite solar cell technology was developing rapidly and has nice commercial prospects.

Supplier and distributors paid attention to efficiency (topic IV) starting in 2014, and paid attention to PSC material (topic I) in 2015. They focused on PSC application (topic VI) and PSC commercialization (topic VII) in 2016. Suppliers and distributors provide solar products and services to consumers, and as such, their focus shifted from the technology R&D aspect to the technology application aspect, and then to the commercialization aspect, which indicated that perovskite solar cell technology has great development potential and has the potential to achieve mass production.

Investors paid attention to PSC structure (topic II) from 2015 to 2016. The structure of perovskite solar cell will affect the efficiency and performance of this technology, and this efficiency and performance will affect the development and commercialization prospects of the technology. Investors generally invest in promising projects to maximize benefit from investments. The attention investors gave to perovskite solar cell technology indicated the relatively fast growth and prominent impact of this technology.

From this analysis, we found that Twitter users' awareness of the rapid development of perovskite solar cell technology increased, and the breadth and depth of this awareness has been enhanced in recent years. Twitter users have begun to pay attention to the commercialization of perovskite solar cell technology, which demonstrated the rapid development trends and commercial potential of perovskite solar cell.

4.2.2.3. Tweets sentiment analysis. The mapping of expectations of emerging technologies using qualitative analysis of documents can provide important insights on the uncertainty and ambiguity and the prominent impact attributes of emergence (Rotolo et al., 2015). In this section, we analyzed Twitter users' emotional reaction to perovskite solar cell technology contained in tweets and studied the changing characteristics of Twitter users' expectation for this technology over time. By analyzing Twitter users' expectations, we may better understand the emergence of technology and gain insight into the future development trends of perovskite solar cell technology.

Sentiment140 is a Twitter sentiment analysis tool, which allows you to automatically classify the sentiment of tweets. Sentiment140 uses a maximum entropy classifier method and can achieve high accuracy for classifying sentiment of tweets (Go et al., 2009). Therefore, in this paper, we applied Sentiment140 to study changes in Twitter users' expectations for perovskite solar cell technology. The analysis steps were as follows: First, we processed the obtained tweets related to perovskite solar cell technology according to the method proposed by Go et al. (2009). The preprocessed datasets were processed using author-developed sentiment analysis codes, and we used API provided by Sentiment140 to obtain the results of sentiment analysis. Then by examining the output results of Sentiment 140, we found that Sentiment 140 is primarily a good classification of tweets with emotional words; Tweets, however, contain sentences that state facts, and the sentiment analysis method does not make a clear classification these statements (Go et al., 2009). For example, solar cells made of perovskite could hit 20–30% efficiency within the next few years. This tweet is to state a fact, and there is no emotional vocabulary in it. So the sentiment analysis method identified this tweet as neutral, but in fact it is a positive expression. Other tweet “New efficiency record set for perovskite solar cells” was also identified as neutral. To resolve these problems, based on the specific context related to perovskite solar cell technology, we marked

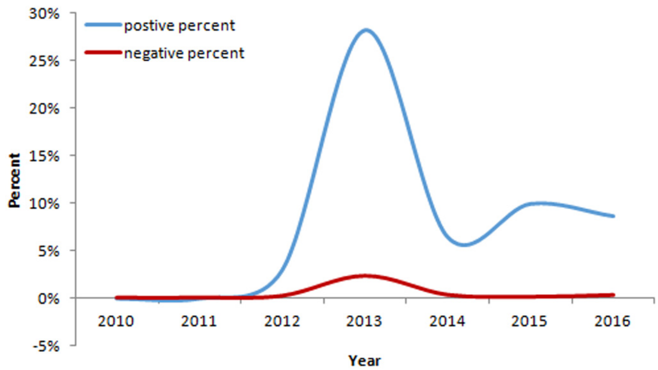

Fig. 7. The proportion of positive and negative tweets related to perovskite solar cell technology.

the tweet as positive if it contained any word related to the attributes of perovskite solar cell technology, including "breakthrough," "cheap," or "efficiency." This was in accordance with the public's awareness of the reality of the technology. We are interested in the positive view and the negative view of perovskite solar cell technology, so we neglected to include neutral tweets in our analysis. Fig. 7 shows the proportion of positive and negative tweets related to perovskite solar cell technology.

As can be seen from Fig. 7, Twitter users have higher positive emotion to the perovskite solar cell technology than negative emotion, which indicated that Twitter users had a positive emotion toward and attitude about the emergence and development of this technology. As also can be seen from Fig. 7, both positive and negative emotion experienced significant fluctuations in 2013. To analyze the cause of these emotional fluctuations, we reviewed the related literature and the corresponding historical events of this technology. In the period from 2012 to 2013, when *Science* named it one of the ten most important technological advancements, the efficiency of perovskite solar cell technology increased from 9.7% in 2012 (Kim et al., 2012) to 16.2% in 2013 (Jeon et al., 2014). The rapid improvement of the efficiency of perovskite solar cell technology generated a high level of Twitter users' attention, and Twitter users had numerous discussions about the development and prospects of this technology. Early on, Twitter users paid attention to the breakthrough of the technology, in some cases, exaggerating its advantages and benefits, and many Twitter users were optimistic about the future development of this technology. Some Twitter users, however, felt pessimistic about the future development of the technology. Therefore, in 2013, the positive and negative emotion to perovskite solar cell technology reached a peak of attention. In the period from 2013 to 2014, the efficiency of the technology increased slowly, and Twitter users found that their expectations for this technology did not conform to the actual development of the technology. As a result, the proportion of positive emotions about perovskite solar cell technology dropped more drastically. When the perovskite solar cell technology was assessed in 2014 by *Nature* to be one of the most anticipated breakthroughs in science and technology, it had an impact on the Twitter users' expectations for and emotion about the technology. At the same time, because of the development of the technology, Twitter users' positive emotion about this technology gradually began to rise and then to stabilize.

Fig. 7 illustrates an interesting phenomena, which is that the proportion curve of Twitter users' positive emotions about perovskite solar cell technology conforms to the hype cycle. After 2014, Twitter users' positive emotions about the perovskite solar cell technology showed a rising trend and then gradual stability, and the proportion of negative emotion also tended to be stable and far below the proportion of positive emotions, which indicated that the hype period of Twitter users on the perovskite solar cell technology might have passed, and Twitter users' expectations for and perception of the technology may have reached a steady state and tended to be more rational. The proportion of positive emotions is greater than the proportion of negative emotions to the perovskite solar cell technology, which shows that Twitter users are full of hope for future development and are optimistic about the market value of the technology. Twitter users' expectations for future development of the perovskite solar cell technology may attract more entrepreneurs, investors, and researchers to join the R&D and application of this technology, which could accelerate the emergence and development of perovskite solar cell technology.

4.3. Comparison analysis of the results of patent analysis and twitter data mining related to the perovskite solar cell technology

To monitoring the emergence of perovskite solar cell technology and identify future changing trends of this technology, we compared the results of patent analysis and Twitter users' sense of and response to perovskite solar cell technology. The analysis results are shown in Table 3. In Table 3, technical topics are derived from patent analysis. Awareness keywords are derived from tweet text mining. The arrows represent the evolutionary relationship between the elements.

From the evolutionary relationship between the topics analysis of patents, as shown in Table 3, we found that the research on perovskite solar cell technology focused only on material in 2012; researchers started to focus on the structure of perovskite solar cell technology in 2013; and they paid attention to the material, structure, and the fabrication of perovskite solar cell technology from 2014 to 2016. From the perspective of technology, the research on perovskite solar cell technology gradually developed from material to structure and then to fabrication from 2012 to 2016, which also showed the path of technological evolution and developing trends of perovskite solar cell technology. Meanwhile, from the topics analysis of tweets mining, we found that Twitter users focused on topics about perovskite solar cell technology such as PSC material and PSC structure in 2013. Twitter user started to pay attention to PSC application in 2014, and they focused on topics such as PSC material, PSC structure, PSC application, and PSC commercialization in 2016. The topics with which Twitter users were concerned related to perovskite solar cell technology changed in the technology R&D aspect, the technology application aspect, and the technology commercialization aspect over time, which was in line with trends in the evolution of topics addressed in patents.

In addition, based on the domain experts' knowledge, we know that the technology topics addressed in patents evolve to improve efficiency, improve stability, and reduce the cost of perovskite solar cell technology. From our analysis of Twitter users' sense of and response to perovskite solar cell technology, as shown in Table 3, we found that Twitter users' attention to efficiency (topic IV) shifted from changing materials to improve efficiency in 2013 and 2014, and gradually evolved to focus on the changing structure to improve efficiency in 2015 and in 2016. Twitter users' attention to stability (topic V) shifted from changing materials to improve technology stability in 2015 and evolved to change structure for improving technology stability in 2016. The trend in Twitter users' sense of efficiency and stability was consistent with the trend in the technology from material research to structure research. From this comparison analysis, we found that trends in Twitter users' sense of topics related to perovskite solar cell technology were in line with trends in the technology topics addressed in patents.

To compare the evolution of technical topics in patents and the evolution of topics perceived by Twitter users, we compared the results of patent analysis and Twitter users' sense of perovskite solar cell technology in detail.

(1) As shown in Table 3, from the material level of patent analysis, we found that the evolution paths of perovskite material gradually shifted from the initial organic materials in 2012 to organic-inorganic perovskite hybrid materials in 2014 and then to inorganic materials in 2016; the evolution paths of hole-transporting material gradually changed from the initial liquid electrolyte material in 2012 to solid

Table 3
Comparison analysis of the results of patent analysis and twitter data mining.

		Patents					Tweets								
		Topics						Public's sensing				Public's response			
		2012	2013	2014	2015	2016	Topics	2013	2014	2015	2016	2013	2014	2015	2016
Material	Perovskite Material	Organic material		Organic-Inorganic hybrid material	Organic-Inorganic hybrid material	Organic material Inorganic material Organic-Inorganic hybrid material	Topic I: PSC material	Perovskite Material	Lead Silicon Tin	Graphene	Hybrid material	Efficient High Excited	Power Stable	Promise Boost Potential	Renewable Develop
	Hole Transporting Material	Electrolyte	Solid organic hole transfer material	Solid organic hole transfer material	Solid organic hole transfer material Inorganic hole transfer material	Solid organic hole transfer material Inorganic hole transfer material									
	Electron Transporting Material	Metal oxides	Metal oxides	Metal oxides Organic materials	Metal oxides Organic materials Insulating material	Metal oxides Organic materials Insulating material									
Structure			Mesoporous structure	Mesoporous structure	Mesoporous structure perovskite heterojunction structure	Mesoporous structure perovskite crystal structure perovskite heterojunction structure	Topic II: PSC Structure	Atom Metal		Atom Hybrids	Nanocrystals Big Panels	High Cheap Conventional		Printed Flexible Magic	Growth High Power
Fabrication				The solution method	The solution method	The casting method The deposition method The evaporation method	Topic VI: PSC application		Lasers Batteries LEDS	Photovoltaic Clothes	Energy PV		Non-toxic Promising Potential	Cheaper Low	Future Light
The evolution of topics		Material	Material Structure	Material Structure Fabrication	Material Structure Fabrication	Material Structure Fabrication	Topic IV: Efficiency	Efficiency Materials	Perovskite Hybrids	Panels	Crystals	Efficient Great Promise	Promise Potential	Inexpensive High Improved	Commercial Outperform
							Topic V: Stability			Stability Technology Breakthrough Material	Layer Panels Interface			Stable Improver Upscaling	Replacing High Improved

organic hole-transporting material in 2013, and then to solid inorganic hole transporting material in 2015; the evolution paths of electron-transporting material gradually changed from the initial metal oxide in 2012 to organic materials in 2014 and then to insulation materials in 2015. We also found that the research on perovskite solar cell technology material gradually changed from high-efficiency materials to performance materials and then shifted to environmental friendly and low-cost materials in the period from 2012 to 2016.

From the PSC material topic analysis of tweets mining, we found that Twitter users' sense of PSC material gradually shifted from the perovskite, lead, silicon, and tin in 2013 and 2014 to graphene appeared in 2015, and then to hybrid materials appeared in 2016. The changes in these awareness keywords showed that Twitter users' sense of PSC material shifted from conventional chemical materials in 2013 and 2014, gradually evolved to focus on new materials in 2015, and then shifted to synthetic new hybrid materials in 2016. We learned that Twitter users' sense of PSC material changed from an initial focus on efficient materials to stable materials and then to environmental friendly materials in the period from 2012 to 2016. The comparison analysis showed that the evolution of Twitter users' sense of PSC material (topic I) was in line with the trend in research topics on technology materials addressed in patents.

(2) As shown in Table 3, from the structure level of patent analysis, we found that the evolution paths of perovskite solar cell technology structure gradually shifted from the initial mesoporous structure in 2013 to a planar heterojunction structure in 2015, and then to perovskite crystal structure in 2016.

From the PSC structure topic analysis of tweets mining, we found that Twitter users' sense of PSC structure gradually shifted from the atom and metal in 2013 to atom and hybrids in 2015, and then to nanocrystals, big, and panels in 2016. The change in these awareness keywords showed that Twitter users' sense of PSC structure shifted from the initial mesoporous structure in 2013, and gradually evolved to focus on the planar heterojunction structure and perovskite crystal structure in 2016. The comparison analysis showed that the trend in Twitter users' sense of PSC structure (topic II) was basically in line with the trend in the research of structural topics addressed in patents.

(3) As shown in Table 3, we found that, in 2014, the solution method was addressed in patents, which revealed that the fabrication of the perovskite solar cell technology had appeared. At the same time, from the topics analysis of tweets mining, we found that Twitter users began to focus on this aspect of technology application in 2014. In 2015, research on planar heterojunction structures appeared in patents, and Twitter users focused on the application of perovskite solar cell technology in clothes in 2015. In 2016, the evaporation method appeared in patents, which mainly has been applied to the preparation of planar heterojunction solar cells, and it makes it possible to produce large area, and low-cost perovskite solar cells. Meanwhile, Twitter users paid attention to the photovoltaic field and the commercialization of perovskite solar cell technology in 2016. Perovskite solar cell technology with a planar heterojunction structure is flexible, prepared easily, and suitable for large-scale application, and this technology could be applied to wearable devices and large-scale photovoltaic power generation devices. The comparison analysis showed that the trend in Twitter users' sense of PSC application (topic VI) was basically in line with the trend in the research of fabrication topics addressed in patents.

From the above comparison analysis, we found that the evolution trend in Twitter users' sense of topics related to perovskite solar cell technology followed basically the same evolution as technical topics addressed in patents. This phenomenon indicated that Twitter users' sense of and response to perovskite solar cell technology could be used to track the emergence and development of this technology.

4.4. Analysis of the development trends of the perovskite solar cell technology based on Twitter data mining

Based on our comparison of the results of patent analysis and twitter data mining related to perovskite solar cell technology, we analyzed the future development trends of the technology by analyzing changing patterns in Twitter users' response to and expectation for this technology, as well as changes in the professional types of Twitter users and their interest in various topics.

(1) From the results of Twitter users' response to and expectations for perovskite solar cell technology, we found that Twitter users tended to pay more attention to environmentally friendly materials and to the flexible, printable, high-efficiency new structure. Twitter users paid attention to the potential for commercialization of perovskite solar cell technology and its future influence on the global photovoltaic market. From the results of Twitter users' expectations for perovskite solar cell technology, we found that Twitter users felt optimistic about the application of this technology, and they expected the technology to be applied to wearable devices and in the photovoltaic power generation fields. They had high expectations for future development and were optimistic about market prospects for this technology. Through tweets sentiment analysis, we found that it is possible to monitor the changes of Twitter users' emotional reaction to the technology and gain insights into the future trends of the technology.

Twitter users' response to and expectations for this technology might steer the development of this technology. Because Twitter users' expectations may reflect potential market demand, which could stimulate and drive the behavior of technology stakeholders. Therefore, Twitter users' response to and expectations for perovskite solar cell technology could be treated as a weak signal to identify the future development trends of this technology.

(2) From the results analysis of the professional types of Twitter users, we found that Twitter users change from researchers, magazines, public organizations and agencies, news outlets, and solar power enterprises to consultation platforms, suppliers and distributors, and then to investors, which also indicated that many different groups of social media users paid attention to the development of perovskite solar cell technology in recent years. In particular, investor attention to the perovskite solar cell technology may have indicated relatively fast development of this technology and also could cast light on the expected impact of the technology.

From the topics of interest to Twitter users, we found that they continued to pay attention to the technology R&D and the technology application, and they focused on the technology commercialization in 2016. Notably, the topics with which researchers were concerned changed from the technology R&D aspect to the technology application aspect, and then to the commercialization aspect, which showed that the perovskite solar cell technology developed rapidly in recent years and has the potential for commercialization. The results of changes in the different professional types of Twitter users and their topics of interest could be used as an indicator of forecasting development trends of this technology.

From the above results analysis of patent analysis and twitter data mining related to perovskite solar cell technology, we found that we can better monitoring and gain insight into the future development trends of the technology.

5. Discussion and conclusions

This study attempted to move the technological emergence and technology forecasting research field forward by using patent analysis and a Twitter data mining method. To do so, we proposed a framework that used patent analysis and Twitter data mining to monitoring the emergence of emerging technologies and to identify future changes trends of these emerging technologies from a technical-environmental perspective. We used patent analysis to monitoring the evolutionary path of these emerging technologies from a technical perspective, and we applied Twitter data mining to analyze the Twitter users' sense of, response to, and expectation for emerging technologies and examined the professional types of Twitter users and their changing topics of interest over time to track the emergence of emerging technologies from an environmental perspective. We made a comparison analysis of the results of patent analysis and Twitter users' sense of and response to emerging technologies, and combined awareness information with Twitter users' expectations for the technologies and changes in the professional types of Twitter users and their topics of interest over time to identify development trends of emerging technologies. We selected perovskite solar cell technology as a case study, and found the proposed framework to be valid and flexible.

Some key findings and contributions are listed as follows:

1. The framework provides a tool to monitoring the emergence of emerging technologies and to identify development trends of emerging technologies based on patent analysis and Twitter data mining. Many researchers have used academic papers and patent data to detect and identify trends in emerging technologies from a technology perspective, although they rarely have made use of social media data focused on emerging technologies. Social media data are important data resources to study the emergence and development trends of emerging technologies. Because the publics' sense of, response to, and expectations for emerging technologies as communicated in social media may reflect potential demand in the market and may indicate the development of intelligence information about the technology. This, in turn, could stimulate and drive the behavior of technology stakeholders and could shape or change the emergence and development of emerging technologies. By analyzing social media data of emerging technologies, we can better understand the public's awareness of emerging technologies. Therefore, combining an analysis of social awareness of emerging technologies with a patent analysis of emerging technologies holds great significance for understanding the emergence of these emerging technologies and for identifying development trends of these emerging technologies. This framework also provides a new method and research perspective to identify technology trends.
2. We constructed an evolution map of the perovskite solar cell technology by using patent analysis and expert judgment, and developed an evolution map of Twitter users' sense of and response to perovskite solar cell technology using Twitter data mining and expert judgment. The maps offer a better understanding of the emergence and development of this technology.
3. We found that trends in Twitter users' sense of topics related to perovskite solar cell technology basically followed the same trends as technical topics addressed in patents when we compared the evolution of technical topics in patents and the evolution of topics perceived by Twitter users. This indicated that we could use Twitter users' sense of and response to perovskite solar cell technology to track the emergence and development of this technology. In the past, researchers applied patent text mining to analyze changes in technical themes over time and to identify potential changes trends in these emerging technologies with the help of domain experts. The reasons behind the changes relied solely on expert knowledge. Through a comparison analysis of the results of Twitter users' sense of and response to perovskite solar cell technology and an analysis of topics addressed in patents, we found that Twitter users' response to the technology could reflect the reasons of the changes in technology topics and potential trends. Meanwhile, combining awareness information with Twitter users' expectations for the technology and changes in the professional types of Twitter users and their topics of interest, we tracked the emergence of this technology and gained insight into the future development trends of the technology. These findings help us acquire greater insights into the use and potential role of Twitter for identifying the future development trends of emerging technology.

This method has some limitations and issues to be considered. First, for the data resources related to emerging technologies, in this paper, we used only Twitter data to analyze social awareness. In the future, the study of social awareness of emerging technologies also should be analyzed using other social media data, including Facebook, Wechat, and blogs. These data could be integrated with Twitter data, which would enable us to better monitoring the emergence of emerging technologies and identify trends. Second, we should pay attention to the behavior of Twitter users. By mining the data traces left by Twitter users in the Twitter platform, we could study the relationships between changing patterns in Twitter users' awareness of emerging technologies and changing characteristics in their behaviors. This would enable us to capture weak signals regarding future development trends of emerging technologies.

Acknowledgments

A previous version of this paper was presented in the 2017 Portland International Conference on Management of Engineering and Technology (PICMET). We are grateful to Alan Porter for his helpful comments and suggestions, and to the two anonymous reviewers for their valuable comments. Many thanks to Dr. Binxin Zhao and Mr. Zhaofu Li for their valuable work on the data mining, and to Prof. Hongyi Li, Prof. Qingbo Meng, and Ms. Jingjing Wang for their helpful consultations. 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 Social Science Foundation of Beijing University of Technology (BJUT201512).

Appendix A

Table 1
Obtained results of topics extraction and the results of the topics harmonization.

The name of the layer	The harmonized name of the topic		Element
Material	Perovskite material	Organic perovskite materials	PbI_3, $CH_3NH_3PbI_3$, Halide Perovskites
		Organic-inorganic hybrid materials	Organic inorganic, Comprising Organic-inorganic perovskite, Hybrid Solar Cell, Hybrid perovskite
		Inorganic materials	Inorganic Solar Cell
	Electronic transport material	Metal oxides	Tin Oxide Perovskite, Titanium Oxide, Carbon Nanotube
		Insulating material	Zinc Oxide
		Organic electron transporting materials	Derivative Used, Zinc Derivative
	Hole transport material	Electrolyte	Electrolyte solution
		Solid organic hole transport material	spiroOMeTAD, Organic Hole-transporting Material,
		Inorganic hole transport material	P-type Semiconductor, Comprises hole transport material
Structure	Device structure	Mesoporous structure	ZnO Nanorod, Mesoporous Scaffold, Scaffold
		heterojunction structure	Planar Heterojunction, Heterojunction Solar Cell
	Perovskite crystals structure	Three-dimensional structure	Perovskite Crystal
Fabrication	The film Fabrication	The solution method	Precursor Solution, Film of Solution
		The casting method	Coating Perovskite,
		The deposition method	Depositing perovskite
		The evaporation method	Sputtering Gas

Table 2
The classification of Twitter users.

	2013	2014		2015		2016	
Researchers	JuliaPercival	RobyFumagalli;	lonepair	DawnSantoianni;	cdtpv	Lonepair;	he_solartube
	parkm2000	JuliaPercival;	bioluisinho	KamatlabND;	PsiFi41	oleh_vybornyi;	JulienBarrier
	rgproductdotcom	SevNHabis;	aashish_999	CarbonDuke;	calestous	TinaMCasey;	joelalexsmith
	d_loris;	DrScottWatkins;	hakala_mo	TinaMCasey;	lonepair	shahzada80;	ConingsBert
	SolarPowerEng	SolarPowerEng;	dlbessette	hakala_mo;	shahzada80	KamatlabND;	PV_Physicist
	Rgproductdotcom	KamatlabND		ShStrumann;	JerzyBuzek	rzumf;	derznovich
	sim0nRedfern			DesmondKLoke		TinaMCasey;	4SSolar
	rgproductdotcom						
Magazines	WR_Systems	ChemPhysChem;	acsnano	ACSCatalysis;	acsnano	EnergyTechnol;	MiradorLtd
	acsnano	ACSPhotonics		ChemMater		ChemPlusChem	pvbuzzmedia
	sciencemagazine	JphysChem		EcoGenMag		SolarNovusAnne;	ChemAsianJ
	CarolynEd1	hyggemediauk		CambridgeCore		ACSEnergyLett;	ACSPhotonics
		SolarNovus		ambUP_engineer		CnvCurmudgeon;	AsiaMaterials
		ChemSusChem		ChemistryWorld		sunwindenergy1;	ChemSusChem
		AlcuinBramerton		EnergyTechnol		ChemPubSoc_Euro;	acsnano
		SolarNovusAnne		ACSPhotonics		PHOTON_Magazine;	
Public organizations and agencies	CECHR_UoD	SurreyChemistry;	SC_Gray		EPFL_en	KAUST_News;	StanfordEnergy
	SurreyOutreach	dailyfusion;	OxfordPV		freeOPV	Materials_MRS;	CECHR_UoD
	EnergyClimateDE	ERAscience;	innovateuk		SLAClab	KavliFoundation;	EPFL_en;
	SurreyChemistry	David_J_L3wis			snsf_ch	musacchios;	QUTSESS

(continued on next page)

Table 2 (*continued*)

	2013	2014		2015		2016	
				StanfordEnergy; CECHR_UoD; theowenlab; Materials_MRS; LosAlamosNatLab BrownUniversity			
News outlets	dailyfusion Investor_Intel	EandTmagazine; TF_GreenLiving; robertmueller74; Thermal_Plant; EnergySouth; dailyfusion	HZBzlog solar_org Innorama GeekGawk TGTechno	physorg_com; bot_innovation; andreasmenn; SolarNovus; solarcurator; perovskiteinfo Investor_Intel	bobaceti SfEP osadirect solar_org TGTechno	coolingZONE11; AusSolarQuotes; energyenviro; Perovskiteinfo; realnewsvideos; WeArePlugIn; physorg_chem; MediosTec21_Vva;	osadirect Adalidda NanoSchaft solarcurator Lujiferondo astar_research scienceanglr Nanowerk
Solar power enterprises	BonnieHoffman BenMarketLine	aiace96 is_ober maxthabiso richardbacon SRPElectricptbo		pvbuzzmedia_ cheynman Surflightroy AdvEnergy SurfaceMSystems CleanEnergyPwys FreshDialogues		AlphaDealClean; Krishna_Mishra; SolarPowerAust; OTEM2000toni; mendotagroup; CleanTechCncpts; ufoenergy	CPFuk jaanvanv minitaxi Solar_Quotes wonderfest grampo
Consultation platforms		3DPYellow; sunnybloke EES_journal; jksaraswat AbsoluteGreenC robinbrittain		solarsolnsteam Business_Save JoeCrawfordBSBA		SolarPowerAust	
Suppliers and distributors		DuncanRenewable NRGCleanPower		EternalSunNL		Rogers_Corp InovateusSolar	
Investors				nateqminar		FNGhadaki	

References

Aral, S., Dellarocas, C., Godes, D., 2013. Social media and business transformation: a framework for research. Inf. Syst. Res. 24, 3–13.

Blei, D.M., Ng, A.Y., Jorden, M.I., 2003. Latent dirichlet allocation. In: Machine Learning Research Archive. vol. 3. pp. 993–1022.

Bollen, J., Mao, H., 2011. Twitter mood as a stock predictor. Computer 44 (10), 91–94.

Boon, W., Moors, E., 2008. Exploring emerging technologies using metaphors: a study of orphan drugs and pharmacogenomics. Soc. Sci. Med. 66 (9), 1915–1927.

Borup, M., Brown, N., Konrad, K., Van Lente, H., 2006. The sociology of expectations in science and technology. Technol. Anal. Strat. Manag. 18 (3), 285–298.

Breitzman, A., Thomas, P., 2015. The emerging clusters model: a tool for identifying emerging technologies across multiple patent systems. Res. Policy 44 (1), 195–205.

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.

Chae, B., 2015. Insights from hashtag #supply chain and Twitter analytics: considering Twitter and Twitter data for supply chain practice and research. Int. J. Prod. Econ. 165, 247–259.

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, S., Yoon, J., Kim, K., Lee, J.Y., Kim, C.H., 2011. SAO network analysis of patents for technology trends identification: a case study of polymer electrolyte membrane technology in proton exchange membrane fuel cells. Scientometrics 88 (3), 863–883.

Condliffe, J., 2010. Flaming drives online social networks. New Sci. 12, 11–12.

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.

Dang, Q., Gao, F., Zhou, Y.D., 2016. Early detection method for emerging topics based on dynamic Bayesian networks in micro-blogging networks. Expert Syst. Appl. 57, 285–295.

Daniel, M., Neves, R.F., Horta, N., 2017. Company event popularity for financial markets using Twitter and sentiment analysis. Expert Syst. Appl. 71, 111–124.

Day, G.S., Shoemaker, P.J.H., Gunther, R.E., 2000. Wharton on Managing Emerging Technologies. Wiley & Sons, Inc, New York.

Doubleclick, 2005. Search before the purchase: understanding buyer search activity as it builds to online purchase. http://www.doubleclick.com.

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.

Go, A., Bhayani, R., Huang, L., 2009. Twitter Sentiment Classification Using Distant Supervision. Cs224n Project Report.

Hong, L., Davison, B.D., 2010. Empirical study of topic modeling in Twitter. In: Proceedings of the First Workshop on Social Media Analytics, pp. 80–88.

Hughes, A.L., Palen, L., 2009. Twitter adoption and use in mass convergence and emergency events. Int. J. Emerg. Manag. 6, 248–260.

Ikeda, K., Hattori, G., Ono, C., Asoh, H., Higashino, T., 2013. Twitter user profiling based on text and community mining for market analysis. Knowl.-Based Syst. 51, 35–47.

Inauen, S., Schoeneborn, D., 2014. Twitter and Its Usage for Dialogic Stakeholder Communication by MNCs and NGOs. 6. Emerald Group Publishing Limited, pp. 283–310.

Injadat, M.N., Salo, F., Nassif, A.B., 2016. Data mining techniques in social media: a survey. Neurocomputing 214, 654–670.

Jansen, B., Zhang, M.M., Sobel, K., 2009. Twitter power: tweets as electronic word of mouth. J. Am. Soc. Inf. Sci. Technol. 60 (11), 2169–2188.

Jeon, N.J., Noh, J.H., Kim, Y.C., et al., 2014. Solvent engineering for high-performance inorganic-organic hybrid perovskite solar cells. Nat. Mater. 13, 897–903.

Kajikawa, Y., Takeda, Y., 2009. Citation network analysis of organic LEDs. Technol. Forecast. Soc. Chang. 76 (8), 1115–1123.

Kajikawa, 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.

Kalampokis, E., Tambouris, E., Tarabanis, K., 2013. Understanding the predictive power of social media. Internet Res. 23, 544–559.

Katz, J.S., 2006. Indicators for complex innovation systems. Res. Policy 35 (7), 893–909.

Kim, H.S., Lee, C.R., Im, J.H., et al., 2012. Lead iodide perovskite sensitized all-solid-state submicron thin film mesoscopic solar cell with efficiency exceeding 9%. Sci. Rep. 2, 591.

Kostoff, R.N., 1994. Research impact quantification. R&D Manag. 24, 207–218.

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.

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, S., Chen, C.M., 2013. The differences between latent topics in abstracts and citation contexts of citing papers. J. Assoc. Inf. Sci. Technol. 64 (3), 627–639.

Liu, Y., Liu, X., Gao, S., 2015. Social sensing: a new approach to understanding our socioeconomic environments. Annals of the association of American geographers. Ann. Assoc. Am. Geogr. 3 (105), 512–530.

Madani, F., Weber, C., 2016. The evolution of patent mining: applying bibliometrics analysis and keyword network analysis. World Patent Inf. 46, 32–48.

Momeni, A., Rost, K., 2016. Identification and monitoring of possible disruptive technologies by patent-development paths and topic modeling. Technol. Forecast. Soc. Chang. 104, 16–29.
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.
Oliveira, N., Cortez, P., Areal, N., 2017. The impact of microblogging data for stock market prediction: using Twitter to predict returns, volatility, trading volume and survey sentiment indices. Expert Syst. Appl. 73, 125–144.
Osinski, S., Weiss, D., 2005. A concept-driven algorithm for clustering search results. IEEE Intell. Syst. 20 (3), 48–54.
Pang, B., Lee, L., Vaithyanathan, S., 2002. Thumbs up sentiment classification using machine learning techniques. In: Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP), pp. 79–86.
Pentland, A., 2005. Socially aware computation and communication. IEEE Comput. 3 (38), 33–40.
Porter, A.L., Detampel, M.J., 1995. Technology opportunities analysis. Technol. Forecast. Soc. Chang. 49, 237–255.
Porter, A.L., Roper, A.T., 1991. Forecasting and Management of Technology. John Wiley & Sons.
Ravi, K., Ravi, V., 2015. A survey on opinion mining and sentiment analysis: tasks, approaches and applications. Knowl.-Based Syst. 89, 14–46.
Robinson, D.K.R., Huang, L., Guo, Y., Porter, A.L., 2013. Forecasting innovation pathways (FIP) for new and emerging science and technologies. Technol. Forecast. Soc. Chang. 80 (2), 267–285.
Rosen-Zvi, M., Griffiths, T., Steyvers, M., Smyth, P., 2004. The author-topic model for authors and documents. In: Proceedings of the 20th Conference on Uncertainty in Artificial Intelligence, pp. 487–494.
Rotolo, D., Hicks, D., Martin, B., 2015. What is an emerging technology? Res. Policy 44 (10), 1827–1843.
Saracoglu, B.O., 2015. An experimental research of small hydropower plant investments selection in Turkey by Carrot2, DEXi, DEXiTree. J. Invest. Manag. 4 (1), 47–60.
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.
Skoric, M., Poor, N., Achananuparp, P., 2012. Tweets and votes: a study of the 2011 Singapore general election. In: Proceedings of 45th Hawaii International Conference on Social Intelligence and Technology (SOCIETY). Hawaii, pp. 2583–2591.
Small, H., Boyack, K.W., Klavans, R., 2014. Identifying emerging topics in science and technology. Res. Policy 43 (8), 1450–1467.
Song, M., Kim, M.C., 2013. RTˆ 2 M: real-time Twitter trend mining system. In: Proceedings of 2013 International Conference on Social Intelligence and Technology, Pennsylvania, pp. 64–71.
Tang, D., Wei, F., Qin, B., Zhou, M., Liu, T., 2014. Building large-scale twitter-specific sentiment lexicon: a representation learning approach. In: Proceedings of COLING, the 25th International Conference on Computational Linguistics: Technical Papers, pp. 172–182.
Twitter, 2013. Twitter developer documentation. https://dev.twitter.com/docs.
Wang, X., Mccallum, A., 2006. Topics over time: a non-Markov continuous-time model of topical trends. In: Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 424–433.
Wang, X.F., Qiu, P.G., Zhu, D.H., Mitkova, L., Lei, M., Porter, A.L., 2015. 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.
Webster, T., 2010. Twitter Usage in America: 2010: The Edison Research/Arbitron Internet and Multimedia Study. Edison Research.
Wei, J., Zhao, Q., Li, H., 2014. Perovskite solar cells: promise of photovoltaics. Sci. Sinica 44 (8), 801–821.
Williams, S., Terras, M., Warwick, C., 2013. What Do People Study When They Study Twitter?. Classifying Twitter Related Academic Papers. vol. 69. pp. 384–410.
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.
Xia, F., Liu, L., Li, J., 2015. Socially aware networking: a survey. IEEE Syst. J. 9 (3), 904–921.
Xu, S., Shi, Q.W., Qiao, X.D., Choi, S.P., 2014. Author-topic over time (AToT): a dynamic users' interest model. In: Lecture Notes in Electrical Engineering. vol. 274. pp. 58–61.
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, J., Park, H., Kim, K., 2013. Identifying technological competition trends for R&D planning using dynamic patent maps: SAO-based content analysis. Scientometrics 94 (6), 313–331.
Zhu, J., 2014. Research on Related Problems of Emotional Analysis Based on Twitter. Wuhan University of Technology, Wuhan.

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, 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.

Jiaojiao Jiang is a postdoctoral researcher in the School of Software and Electrical Engineering, Swinburne University of Technology, Australia. Her recent research interests include data mining, and social media analysis.

Yuan Zhou is an associate professor in the School of Public Policy and Management, Tsinghua University, China. His recent research interests include public policy, innovation policy, and technology roadmapping.

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