

Contents lists available at ScienceDirect

Technological Forecasting & Social Change

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



# R&D trend analysis based on patent mining: An integrated use of patent applications and invalidation data



Xiaotong Han<sup>a</sup>, Donghua Zhu<sup>a,\*</sup>, Ming Lei<sup>a</sup>, Tugrul Daim<sup>b</sup>

<sup>a</sup> School of Management and Economics, Beijing Institute of Technology, No.5 South Zhongguancun Street, Haidian District, Beijing 100081, China <sup>b</sup> Department of Engineering and Technology Management, Portland State University, 1900 4th Avenue, Portland, OR 97207-0751, USA

#### ARTICLE INFO

Keywords: Technology opportunity analysis Patent invalidation Patent mining Topic model R&D trend

# ABSTRACT

Formulating good R&D strategy requires sound knowledge of the past and present R&D trends in various industry sectors. Therefore, this paper outlines a framework for mining industry level R&D trends from patents that were designed for enterprises. Unlike the current alternatives, the approach presented here considers both patent applications and invalidated patents, i.e., those patents that have expired, lapsed, or been revoked. The result is a richer and more comprehensive analysis that covers the full lifespan of a targeted technology from emergence to decline. The framework comprises of a LDA topic model that identifies the technologies and sub-technologies, and of each individual patent and invalidated patent. Then, two specifically designed measures chart the stages of the technologies' life. An application metric reflects annual levels of interest in an area, while an invalidation metric traces waning interest. The output is a series of trend maps that show the levels of interest and disinterest in different avenues of inquiry over time. Charted on different axes, these two metrics create two distinct trend lines that reflect the different changes over a technology's lifecycle. A case study that focused on China's 3-D printing technology illustrates the approach. The analysis results are highly consistent with the present technology trends across industries, which indicates that the method could serve as a useful reference tool for analyzing R&D trends and creating new R&D strategies.

# 1. Introduction

Innovation is recognized as the most important source of a firm's growth and economic development (Penrose, 2009; Schumpeter, 1911), and one that is especially significant for technology-based companies. However, blindly investing in an ill-thought-out R&D direction may cause problems. For example, investing in a technology only to find a market filled with competitors would be an unwelcome surprise. Pursuing technologies that are beyond the competencies of those involved would probably result in a significant waste of time and resources. A better approach is to identify R&D directions that suit the capabilities of the enterprise, and that are likely to result in the company achieving its strategic goals. In other words, developing a sound R&D strategy requires that a company accurately identifies and then assesses its technology opportunities. Part of this process is identifying the R&D strategies competitors are pursuing as well as current industry trends. These are the problems we intend to address in this paper.

Technology opportunities analysis (TOA) was first proposed by

Porter and Detampela (1995) to discover opportunities for technological innovation. Previous studies on TOA have mostly focused on specific technologies, discovering technology gaps through methods like morphology analysis (Yoon and Park, 2005) and semantic analysis (Choi et al., 2011). A few studies have paid attention to the first steps of technology innovation, which are the decisions made surrounding what to innovate. Arguably, this step is the most crucial to formulating an R&D strategy, particularly for industry newcomers.

Additionally, almost all previous TOA studies ignore a patent's legal status, i.e., whether a patent is still active or whether it is now inactive/ invalidated. Yet, a key factor in identifying technological trends are invalidated patents and, more importantly, the reason for the invalidation. For example, when the number of patent applications in a technology sharply drops, it is not possible to say whether the technology is in decline or has merely encountered a bottleneck without knowing the number of invalidated patents. A rapid increase in the number of invalidations likely means the technology is in decline. Otherwise, bottleneck is the more probable cause. At the firm-level, taking patent

https://doi.org/10.1016/j.techfore.2021.120691

Received 8 April 2020; Received in revised form 17 February 2021; Accepted 19 February 2021 Available online 1 March 2021 0040-1625/© 2021 Elsevier Inc. All rights reserved.

This article belong to the special section on Advanced Tech Mining: Technical Emergence Indicators and Measurements.

<sup>\*</sup> Corresponding author.

E-mail address: zhudh\_bit@163.com (D. Zhu).

invalidations into account can help to identify which technologies a company has stopped pursuing. For example, in China more than half of all invalidations are at the behest of the applicants, while less than 5 percent simply expire. Hence, in the same way that analyzing the patent applications for a given industry can accurately reveal hot R&D topics, considering patent invalidations can provide insights into which technologies firms have lost interest in. The result is a much more comprehensive picture of an industry's R&D landscape.

Most of the literature on invalidated patents comes from the field of legal studies, e.g., explaining the anticompetitive effects of unenforced invalidated patents (Leslie, 2006); examining why most Korean patents are invalidated (Lee and SooMee, 2012); or analyzing how to preserve market dominance after patents expire (Ii, 2006). Only a few studies use invalidated patents in technology forecasting. Scholars in China have pointed out that invalidated patents can be good indicators of technological dynamics, R&D directions, and even market trends (Fu, 2005; Gu and Yang, 2010; Lai et al., 2007). However, none have proposed a specific method.

As a remedy, we propose an approach to analyzing R&D trends directed at the industry level. The method is straightforward. First, a database consisting of both patent applications and invalidations is fed into an adapted latent Dirichlet allocation (LDA) topic model to identify the technology(s) covered in each individual patent. Two specifically constructed measures are then used to evaluate the applications and invalidation levels in each technology topic. The output is a series of two-dimensional trend maps of the applications and invalidation levels of each technology topic in different years. The shape of the trend line for each period indicates which of four possible stages of life the technology was in at that time. The four stages match the standard innovation lifecycle: emerging (low-application, low-invalidation), growing (high-application, low-invalidation), upgrading (high-application, highinvalidation), and declining (low-application, high-invalidation). The results of an analysis on any given technological topic should provide insights and advice for companies engaged in strategic R&D planning. We tested our method with a case study on China's 3D printing innovators. The results align with the sector's present technology trends, providing validity to our method.

The rest of this paper is organized as follows. In Section 2, we introduce relevant studies on the four stages of technological innovation and review the approaches and tools used in TOA. Section 3 outlines our proposed approach, followed by a case study on Chinese patented 3D printing technologies in Section 4. The discussion and conclusions are presented in the last section.

#### 2. Literature review

# 2.1. TOA in the different stages of technological innovation

Technological innovation is the process of discovering and realizing technology opportunities. A technological innovation system can be defined as "a dynamic network of agents interacting in a specific economic or industrial area under a particular institutional infrastructure and involved in the generation, diffusion, and utilization of technology" (Carlsson and Stankiewicz, 1991). Different scholars divide technological innovation into several stages; however, these divisions are relatively meaningless because what separates one stage from another is not clear. Further, the stages are not necessarily sequential, but rather are more likely to be cyclical or overlapping. In this paper, we have divided the process of technological innovation into four stages: decisions, R&D, production, and the market.

Decisions about whether to even engage in technological innovation are made, obviously, in the decision stage. Once decided, the next decision concerns which field to conduct the innovation activity in. Relevant TOA studies on this stage include: technology cluster analysis for enterprise R&D strategic planning (Hsu, 2006); developing new technology ideas based on the F-term (Song et al., 2017); and identifying potential opportunities for innovation arising from converging technological areas (Kim et al., 2017), among others.

In the R&D stage, the key benchmark points in the R&D process are mapped out and executed and appropriate investments are made according to the specific technology. A range of methods have been proposed for analyzing technology opportunities in this stage. For example, Yoon and Park (2005) improved morphology analysis by adapting text mining techniques to patent documents, while Ma et al. (2014) used text mining to predict the direction a technology would take during its development. Wang et al. (2017) proposed a similar method for detecting priority combinations of technologies. Lee et al. (2009) constructed keyword-based patent maps for use in new technology creation activities, and Yoon et al. (2013) used dynamic patent maps to show the trends in technological competition. Ma and Porter (2015) identified high potential opportunities by analyzing technology development pathways. Other researches have helped to identify technology opportunities for specific business structures, such as SMEs (Lee et al., 2014).

Almost no TOA-related research analyzes the production stage, which involves innovations in tools, machines, equipment, process flow, etc.

The last stage, bringing a technology to market, involves recognizing the characteristics and trends of innovative products to discover technology opportunities based on market demand. Studies on this stage generally link a technology with products and analyze the opportunities these links present (Yoon et al., 2015; Yoon et al., 2014). Some identify profitable markets and promising product concepts based on technology information (Jin et al., 2015) or opportunities for technology-based services (Kim et al., 2015).

Most of the studies on TOA to date have concentrated on the R&D and market stages. This includes analyzing a specific technology to reveal gaps in the market that might be filled through an R&D process – or the reverse, analyzing a market to reveal which technologies are in demand. This leaves a need in the literature for studies that explore the decision and production stages. In this paper, we focus on the decision stage, where enterprises decide whether to engage in innovation and which technology or field to choose. This is a crucial stage in any innovation endeavor, particularly for newcomers to an industry, because it is here that well-thought-out R&D strategies find their genesis.

# 2.2. Methods of technology opportunities analysis

Our research draws on TOA, which was first proposed by Porter and Detampela (1995). These authors put forward a modern version of monitoring based on bibliometric analyses to identify technology opportunities. TOA has subsequently been enriched by many scholars using a variety of methods, most of which are not independent but, rather, come together in various combinations each playing a different role in the steps of TOA.

Zhu and Porter (2002; 1999) present a specific process for TOA that they describe through five main stages: (1) data retrieval and acquisition; (2) data description; (3) extracting potential relationships; (4) visualizing those relationships; and (5) interpreting the results to identify potential opportunities.

Text mining has been widely used in the relationship extraction step, and is usually based on finding links between keywords or subject terms in the abstracts of documents (Kim et al., 2014; Lee et al., 2009; Song et al., 2017; Yoon and Park, 2005). Principal component analysis (Lee et al., 2009; Ma and Porter, 2015), cluster analysis (Kim et al., 2014), and topic models (Kim et al., 2015; Lee and Sohn, 2017; Momeni and Rost, 2016) are all aimed at standardizing and simplifying data to extract potential relationships in a more concise way. Knowledge units are selected to build up a picture of the existing relationships before identifying the potentials for new relationships. Morphology analysis (Ma et al., 2014; Yoon and Park, 2005), semantic analysis such as SAO (subject-action-object) (Wang et al., 2017; Yoon et al., 2013) and AO (action-object) (Lee et al., 2014) have also been used. More recently, other methods, like collaborative filtering (Lee and Lee, 2017; Park and Yoon, 2017) and novelty detection (Geum et al., 2013; Kim et al., 2017; Lee et al., 2015) have gained popularity as a way to discover potential opportunities through relationships.

To visualize these relationships, patent maps (Lee et al., 2015, 2009; Yoon et al., 2013) and two-dimensional portfolio maps (Kim et al., 2015) are common, and network analysis is often used to interpret the results (Kim et al., 2015).

## 3. Methodology

Patent mining is an established methodology used for analyzing R&D trends (Daim et al., 2006). It has been used on its own and has also been combined with other methods of analysis to explore many technology trends including in energy (Daim, 2012), smart buildings (Madani et al., 2017), biotechnology (Gonçalves Pereira et al., 2019), and electric cars (Gibson et al., 2017). However, as more advanced data analytics methods have emerged, more sophisticated models have been developed to evaluate technology opportunities. Example studies include Li et al.'s (2019) analysis of autonomous vehicles and Lin et al.'s (2019) analysis of solar technologies.

In this research, we propose a method based on patent applications and invalidation data to analyze R&D trends. The framework is shown in Fig. 1. It is a four-step procedure that includes: (1) data collection and pre-processing; (2) the discovery of technology topics; (3) the construction of measures to assess the number of active/approved applications and the number of invalidations; and (4) the analysis of R&D trends.

## 3.1. Step I: data collection and pre-processing

The method is ostensibly based on patent mining. In this paper, we used patent applications and invalidation data from the China Patent Database, which contains information about the legal status of each patent. Only invention patents were selected because other types of patents are not good representations of innovative technology. In China, patent applications are secret until they are made public the gap between lodging a patent application and its public release is usually a little less than 18 months. However, it is advisable to use the most recent data available, manually trimming years where the data is too sparse and may skew the results. The case study illustrates this point in more detail.

Invalidated patents means the rights over the technology in the document were not granted or are no longer valid (Xing, 2013). The four main reasons a patent may be invalidated are: non-payment of the annual fee; a deemed withdrawal due to non-payment of various expenses or failure to reply to an examination notice; rejection of the application; or the patent has reached its expiry date. In 2016, the first two reasons accounted for 87% of all invalidated patents (59% and 28% respectively). These are subjective factors caused by the applicant. For instance, a company might choose to waive its patent rights if the expected profits will not cover the annual fee, or they may simply decide not to pursue the application to final approval. Both factors indicate that the applicant has lost interest in the invention or the technological field. The last two reasons do not reflect actions by the applicant. Hence, in this study, we have only considered patents invalidated for the first two (applicant-driven) reasons.

Once the two datasets are assembled, salient terms are extracted from the patents' titles, abstracts, and main claims. General words and meaningless words are removed manually, and a term dictionary is formed. Each abstract is then segmented into words according to the term dictionary, and a patent-word matrix is generated for the topic model.

## 3.2. Step II: discovering technology topics with an LDA model

Many studies have used International Patent Classification (IPC) as a classification criteria. In China, IPC is the only standard for classifying patents. However, IPC classifications tend to be functional, which is not particularly suited to characterizing technologies. Hence, our framework prescribes classifying the technologies with a topic model instead



Fig. 1. Research framework.

of simply taking the IPC codes from the patent.

LDA topic models (Blei et al., 2003) have become a popular approach to identifying hidden topics in a corpus of text. Among its applications, Lee and Sohn (2017) used an LDA topic model for patent mining to identify emerging areas and trends in patents on financial business methods, while Wang et al. (2014) used an LDA model to find the implicit relationships and inherent links between technologies. Kim et al. (2015) applied an LDA topic model to forecast gaps in technological areas.

Topic modeling is based on an LDA using Bayesian statistics to infer what each word might mean from its neighboring words or words that frequently co-occur (Blei et al., 2003). Based on the assumption that a latent set of topics exists within every document, each word appearing in a document can be assigned to one of the topics in the set with some probability (Momeni and Rost, 2016). As Fig. 2 shows, the input for the topic model is the patent-word matrix constructed in Step I. The number of topics **K** and the hyperparameters  $\alpha$ ,  $\beta$  are established during this stage; these determine the probabilities of the Bayesian priors. The prior probabilities are continually updated through thousands of iterations of Gibbs sampling to derive the posterior probabilities, i.e., the patent-topic matrix and the topic-word matrix.

The number of topics K affects the clustering quality. A common criterion to assess the quality of the model is perplexity, which is used as a reference for determining the number of topics (Blei et al., 2003; Heinrich, 2008). The quality of the model is considered better when the perplexity is relatively low. It is calculated by

$$perplexity = \exp\left\{-\frac{\sum_{d=1}^{M} \log p(\mathbf{w}_d)}{\sum_{d=1}^{M} N_d}\right\}$$
(1.1)

where **M** is the total number of documents,  $N_d$  is the total number of words in a document d and  $p(\mathbf{w}_d)$  is the probability vector of the words in document d, which is calculated by

$$p(\mathbf{w}_d) = \sum_{t=1}^{K} p(\mathbf{t}_d) p(\mathbf{w}_t)$$
(1.2)

where **K** is the total number of topics,  $p(\mathbf{t}_d)$  is the probability vector of the topics in document **d**, and  $p(\mathbf{w}_t)$  is the probability vector of the words in topic **t**.

Settings for the hyperparameters  $\alpha$  and  $\beta$  have been reported as resulting in good model quality at  $\alpha = 50/K$  and  $\beta = 0.01$  (Griffiths and Steyvers, 2004; Heinrich, 2008).

The resulting patent-topic matrix lists the technology topics

identified in individual documents. For most of the topics in a document, the probability will be very close to 0. These topics are ancillary and are excluded according to a threshold of 1/K. Only several topics should remain, which represent the main topics discussed in the document. The topic-word matrix is used to infer the content of individual topics by checking the few words with the highest probability in each topic. Sets of technology topics that are hidden within patent documents can be explored with this method.

3.3. Step III: constructing the measures for applications and invalidation levels

The patent-topic matrix can be divided into two further matrices according to their legal status: application-topics (Fig. 3) and invalidation-topics (Fig. 4) In these two matrices, any probability smaller than 1/K is replaced with 0. We use two indicators to measure the extent to which a topic still holds an applicant's interest (*NTA*) or the applicant's loss of interest in a topic (*NTI*), these are also referred to as the application level and the invalidation level.

Both **NTA** and **NTI** are normalized to the weighted number of applications or invalidations within a technology topic and are calculated for each technology topic. Minimum-maximum normalization brings all the values into the range [0,1] with the following equations.

$$NTA_i = \frac{TA_i - TA_{min}}{TA_{max} - TA_{min}} \quad i \in [1, k]$$
(2.1)

where

$$TA_{i} = \sum_{m=1}^{M} A_{m,i} \quad i \in [1,k]$$
(2.2)

$$TA_{max} = max\{TA_i, i \in [1, k]\}$$
(2.3)

$$TA_{min} = min\{TA_i, i \in [1, k]\}$$
(2.4)

$$NTI_j = \frac{TI_j - TI_{min}}{TI_{max} - TI_{min}} \quad j \in [1, k]$$
(3.1)

where

$$TI_i = \sum_{n=1}^{N} I_{n,j} \quad j \in [1,k]$$
 (3.2)

$$TI_{max} = max\{TI_i, j \in [1, k]\}$$
 (3.3)



		PA <sub>1</sub>	PA <sub>2</sub>	PA <sub>3</sub>	 PAm	 PAM
	T <sub>1</sub>	A <sub>1,1</sub>	A <sub>2,1</sub>	A <sub>3,1</sub>	 A <sub>m,1</sub>	 A <sub>M,1</sub>
	T <sub>2</sub>	A <sub>1,2</sub>	A <sub>2,2</sub>	A <sub>3,2</sub>	 A <sub>m,2</sub>	 A <sub>M,2</sub>
Topics	T <sub>3</sub>	A <sub>1,3</sub>	A <sub>2,3</sub>	A <sub>3,3</sub>	 A <sub>m,3</sub>	 A <sub>M,3</sub>
	Ti	A <sub>1,i</sub>	A <sub>2,i</sub>	A <sub>3,i</sub>	 A <sub>m,i</sub>	 A <sub>M,i</sub>
	T <sub>k</sub>	A <sub>1,k</sub>	A <sub>2,k</sub>	A <sub>3,k</sub>	 A <sub>m,k</sub>	 A <sub>M,k</sub>

Patent applications

# Fig. 3. Application-topic matrix.

# **Invalidated** patents

		$PI_1$	$PI_2$	PI₃	 Pln	 $PI_N$
	T <sub>1</sub>	I <sub>1,1</sub>	I <sub>2,1</sub>	I <sub>3,1</sub>	 I <sub>n,1</sub>	 $I_{N,1}$
	T <sub>2</sub>	I <sub>1,2</sub>	I <sub>2,2</sub>	I <sub>3,2</sub>	 I <sub>n,2</sub>	 I <sub>N,2</sub>
Topics	T <sub>3</sub>	I <sub>1,3</sub>	I <sub>2,3</sub>	I <sub>3,3</sub>	 I <sub>n,3</sub>	 I <sub>N,3</sub>
	Tj	I <sub>1,j</sub>	I <sub>2,j</sub>	I <sub>3,j</sub>	 I <sub>n,j</sub>	 I <sub>N,j</sub>
	T <sub>k</sub>	$I_{1,k}$	I <sub>2,k</sub>	I <sub>3,k</sub>	 I <sub>n,k</sub>	 I <sub>N,k</sub>

Fig. 4. Invalidation-topic matrix.

level

$$TI_{min} = min\{TI_i, j \in [1,k]\}$$
 (3.4)

In these equations,  $TA_i(TI_i)$  is the total probability of topic i(j) in the application-topic (invalidation-topic) matrix, which is the sum of row i(j).  $TA_{min}$  ( $TI_{min}$ ) and  $TA_{max}$  ( $TI_{max}$ ) are the minimum and the maximum over all  $TA_i$  ( $TI_i$ ) in the application-topic (invalidation-topic) matrix.

The application (invalidation) levels are bounded between 0 and 1, where 1 represents the topic a company has lodged the most applications for and 0 is the least. A higher application score (more applications) means greater interest; a higher invalidation score (more invalidations) means greater disinterest.

## 3.4. Step IV: analyzing R&D trends using a two-dimensional scatter plot

Foster (1986) divides a technology's lifecycle into four stages: emerging, growth, maturity, and saturation. Subsequently, Ernst (1997) proposed a method for identifying these stages using a combination of the number of patent applicants and their applications. In addition to these four stages, Meyers (2004) added a further stage called the "innovation renewal period" where a firm's R&D focus is on rejuvenating or replacing a technology.

Drawing on these stages and Ernst's (1997) method of identifying these stages, we divided the technology topics into stages according to the two dimensions under study in this research: applications and invalidations, as shown in Fig. 5.

The emerging stage contains technology topics with both low application and invalidation levels during the study period. The technologies in this area have only just emerged so patent activity is generally low. However, low activity may also be attributed to atypical technologies, unpopular technologies, or technologies waiting for some kind of breakthrough to further progress. Therefore, a technology topic must have a rise in both application and invalidation levels to be considered an emerging technology. Note, however, that this upward trend does not need to be absolute. In many cases, there could be one or two years where the trends fluctuate before stabilizing.

The growing stage contains patents with much sharper increases in



Fig. 5. Dividing technology topics into stages.

application levels than in invalidation levels, and therefore invalidated patents should be low. These signs indicate that a technology is in ascendance and dominating the R&D direction of the industry, and therefore the natural assumption is that there should be good prospects for this technology.

The upgrading stage sees sharp increases in both application and invalidation levels. Here, both criteria have a large and fast-growing number of patents, which means that the technology is undergoing change, either through upgrades or because the components of the technology are being replaced.

The remaining stage covers declining technologies with reduced application levels, but continuing increases in invalidation levels. More patent invalidations than active applications is a signal that the technology may be outdated or has been superseded by other technologies. As an R&D prospect, this technology no longer has value. However, again, this area is complex as the technology may have encountered a bottleneck, such as lack of a breakthrough or difficulties with upgrades due financial problems, lack of R&D capability, and so on.

Categorizing technologies in this way offers companies a better understanding of the technological trends in their industry. These insights have the potential to provide a solid reference for companies to build upon when formulating their own (well thought out) R&D strategy.

# 4. Case study

# 4.1. Data description and LDA model results

3D printing, also known as additive manufacturing, emerged in the mid-1990s as a rapid prototyping technology. As a digital model, it uses a powdery material to construct the object layer by layer. 3D printing technology has applications in many fields such as jewelry design, construction, automotive engineering, aerospace, dentistry, medical technology. We chose this growing and innovative technology area as our case study to verify the feasibility of the approach.

Following the LDA method, we downloaded all patent applications associated with 3D printing in China from a Chinese commercial database "DIInspiro" (http://zldsj.com/). The total number of applications retrieved was 24,510 from 2005 to 2020. Among them, 2926 patents were invalidated for the two applicant-driven reasons. Table 1 shows the number of applications and invalidated patents for each year. We fed all the data into the LDA model, but only used the data from 2015 to 2019 for the trend analysis because there were too few patents prior to 2015 and the data for 2020 was incomplete.

We used ITGinsight software (http://en.itginsight.com/) for the text mining. Only terms that appeared more than once were included the dictionary, and generic and meaningless words were excluded. In total, 24,119 terms were selected from the titles, abstracts, and claims of 24,510 patents, resulting in a patent-word matrix of size 24,510 × 24,119. To assess perplexity (see Eq (1.1)), we evaluated the model with different numbers of topics. The results are shown in Fig. 6. As the number of topics increased, the perplexity decreased. To avoid redundancy in the topics, we set  $\mathbf{k}$  to =140 where the perplexity was relatively low, and the curve began to flatten.

# 4.2. R&D trend analysis

The next step was to calculate the application (*NTA*) and invalidation (*NTT*) levels for all 140 topics. Figs. 7-10 show some example technology topics for each stage as two-dimensional trend plots. In these figures, a plot represents a technology topic. The broken line represents the change of *NTA* and *NTI* from 2015 to 2019. We chose the medians of *NTA* and *NTI* in 2019 as the threshold for dividing the stages. If the *NTA* and *NTI* of a topic (in 2019) were both smaller than the thresholds, the topic was denoted as emerging (Fig. 7). An *NTA* greater than the threshold, but with an *NTI* smaller than the threshold, was labeled as growing (Fig. 8). If both *NTA* and *NTI* levels were greater than the thresholds, then the topic was classified as being in the upgrading stage (Fig. 9). An *NTA* level smaller than the threshold but with an *NTI* level

Table 1	1
---------	---

N	umb	er c	f	patents	of	each	ı j	year.
---	-----	------	---	---------	----	------	-----	-------

Year	Applications	Invalidations
2005	4	0
2007	1	0
2008	4	0
2009	2	0
2010	5	0
2011	9	3
2012	39	3
2013	258	4
2014	1004	7
2015	1741	23
2016	3265	134
2017	4843	350
2018	5570	530
2019	5785	1293
2020	1980	579
Total	24.510	2926

greater than the threshold implied that the topic was in the declining stage (Fig. 10). Tables 2-5 list some sample topics from each of the stages along with the most frequently mentioned terms in each topic. We inferred that the content of each topic from these terms.

# 4.2.1. Technology topics of emerging stage

As shown in Fig. 7, topics 031, 061, 091 and 097 are examples of emerging technologies. Their application and invalidation levels were both low and volatile during 2015–2019. Topics in this stage may have only recently emerged. Topic 031 involves neural networks, which is likely associated with deep learning for 3D modeling. Topic 061 is 3D printed artificial blood vessels, including veins and arteries. Topic 091 is bioactive glass as printing material, which can be used to 3D-print bone trabeculae and implants, and to repair bone tissue and defects in the body. Topic 097 involves shape memory alloy as a 3D printing material, which is a substance that can restore its original shape through thermoelasticity. These technologies are just beginning to establish a market foothold, and therefore there is less competition. Large enterprises with a strong capacity for R&D might opt to enter the marketplace with some success. On the other hand, these technologies may not be the best choice for companies with weak R&D capabilities or those just entering the industry because fostering a competitive edge in a nascent technology tends to require strong and established innovative capabilities, and a level of resources that may not be available to small businesses.

#### 4.2.2. Technology topics of growing stage

Fig. 8 and Table 3 show topics 005, 016, 041, 057, 086, 092, 095, 117, 121. These technologies are in the growth stage with a sharp increase in applications, and low invalidation levels. This indicates that these technologies were on the rise and popular during the window of analysis (in this case, 2015–2019). Topic 005 involves teeth and jaw treatments, including 3D-printed dental guides, braces, implants, dental molds, jaw surgery guides, and so on. Topic 016 is 3D-printed bone and joint prostheses. Other growing technologies are topic 041, alloy material, and topic 092, metal powder, which includes titanium alloy powder, aluminum alloy powder, and magnesium alloy powder. These types of alloys can be used for 3D printing in areas such as dentistry, orthopedics, and in light-weight part manufacturing for the automotive and aerospace industries. Topic 057 is slicing software for image layered processing.

Topic 086 is arc additive manufacturing. Metal additive manufacturing technology can be divided into three types according to heat source: laser, electron beam, and arc. Research over the past two decades has focused on the first two approaches, which may not be able to form certain structures or specific parts, or may be too costly in terms of raw materials and time. So, to meet the needs of large scale and integrated parts, arc additive manufacturing technology has attracted more and more attention for its lower costs and higher efficiency.

Topic 095 involves 3D printing as applied to building and architecture, such as developing concrete materials that can be used for 3D printing and printing actual walls. Topic 117 involves unmanned manufacturing. Automated additive manufacturing using robots is an important trend. Topic 121 is gel material, which can be used for bio-3D printing. Technologies in the growth stage are recommended for most companies because they are generally good industry prospects. Assuming a firm chooses to direct its efforts toward something aligned with its skill sets, investing human and financial resources into growing technologies is a conservative approach to R&D.

# 4.2.3. Technology topics of upgrading stage

Fig. 9 and Table 4 show Topics 008, 030, 064, 067, 130. These are technologies in the upgrading stage. Technologies undergoing change display both a sharp increase in applications and invalidation levels. Topic 008 and Topic 030, respectively, represent feeding and air exhaust devices, both of which are components of the 3D printer. Topic 064 is ceramic 3D printing materials and ceramic 3D printing technology.



Fig. 6. Perplexity results with different numbers of topics.



# Emerging stage

Fig. 7. Trends of example technology topics in emerging stage.

Topic 067 involves biological materials, such as cells. Topic 130 is personalization and customization, mainly customizing 3D printed models for patients in the medical field. For these technologies, enterprises wishing to pursue innovation to ultimately occupy a place in the market must find a novel upgrade direction to pursue. For example, they might develop more efficient, less wasteful feeding or air support devices, more flexible ceramic, or biological materials, etc.

Technologies undergoing a period of upgrade can present good opportunities for smaller, nimbler companies to capitalize on R&D if they can beat a giant to market. Moreover, many invalidated patents have not lost all their technology value, and they not protected under any laws. Newcomers might reduce costs by leveraging the knowledge of the past to avoid reinventing the wheel or by applying those inventions directly to new products.

# 4.2.4. Technology topics of declining stage

Fig. 10 and Table 5 show Topics 039, 055, 112, 123, 136. These technologies have reduced application levels and increased invalidation



Fig. 8. Trends of example technology topics in growing stage.



Fig. 9. Trends of example technology topics in updating stage.

levels and are therefore in their declining stage. They may have reached maturity, been replaced by other technologies, or have reached a bottleneck that is difficult to break through. Technologies in the declining stage warrant further scrutiny as to the cause of the decline. If the technology has no market prospects, any R&D plans in this direction should be abandoned. Alternatively, if a technological bottleneck is the cause of decline, companies might look for new ways to break through. Among the declining technologies found in our analysis, Topic 039 is plastic or thermoplastic material, and Topic 136 is photocuring resin. As plastic and resin 3D printing matures, new innovations are dwindling, and higher-tech 3D materials, such as metals and biomaterials, are taking their place. Topics 055, 112,123 represent mechanical arms, sensing devices, and material stock device, which are components of a 3D printer. These technologies may be past maturity and/or require



Fig. 10. Trends of example technology topics in declining stage.

#### Table 2

Terms of example technology topics in emerging stage.

Techi	nology topic	High probability terms
031	Neural network	Rubber, network, nerve, hardness, Silicone rubber, sulfide, array, degree, indicators, particles, neural network, network structure, tire, architecture, nerve conduits, Rubber material, network model, jack, High hardness convolution probability, rubber
061	Artificial blood vessels	Artificial, blood vessels, building materials, artery, artificial bone, branch, preparation, aorta, stents, drug, heart, inhibitors, human body, artificial teeth
091	Bioglass	Glass, active, biological activity, trabecular, glass plate, flexibility, mesoporous, vitrification, glass powder, anchor point, trabecular structure, bioglass, toughened, quartz glass, bioactive glass, mesoporous biological, metal bone, pioneer
097	Shape memory	Deformation, square, memory, ink, shape memory, rectangular, crystallization, tetragonal, slot, memory alloy, cuboid, heat, ladder, rectangle, field, square, drilling

disruptive innovation.

# 5. Discussion and conclusions

Previous studies on TOA have often neglected the legal status of patents. Yet both active patents and those that have lapsed can serve as good indicators of current interest in a technology. If a company is not willing to pursue a patent application to final approval or is unwilling to continue paying the annual maintenance fee, the reasons behind these decisions are worthy of scrutiny. Lapsing interest in a technology can be a significant indicator of industry trends, struggling technologies, or technologies that may face fierce competition once developed. The method presented in this paper is designed to identify R&D trends as a tool to help companies avoid the pitfalls of formulating an R&D strategy that ultimately fails due to market forces. We used a case study that focused on the 3D printing technology in China to illustrate the method and the insights companies might derive. With a sample of both active and invalidated patents as the dataset, we formed a dictionary of salient terms. An LDA topic model discovered 140 technology topics, and two normalized measures were used to calculate the application and invalidation levels for each topic. Based on these measures, we divided the patents into four stages of the technology lifecycle: emerging, growing,

Table	3					
-	c		1	1		

Tech	nology topic	High probability terms
005	Teeth and jaw	guide, teeth, surgery, patients, oral, bone cutting,
	treatment	jaw, braces, implant, retainer, gum, dental, dental
		implants, alveolar bone, rectification device, tooth
		model
016	Bone and joint	joints, prosthesis, chamber, femoral, shin,
	prosthesis	acetabulum, knee joint, bone, the conduction,
		patients, bone cutting, block, artificial, joint
0.41	Allow motorial	Allow Titonium allow Allow neurodan Aluminium
041	Alloy material	Alloy, Alloy material Magnesium alloy, Argon gas
		Heat treatment Ball mill Allow material increasing
		Allow composite mixed powder ingot casting
		impurities. Zinc allov
057	Slicing software	Slice, software, thickness, layer, single chip
	0	microcomputer, slicing software, sand blasting, D/a,
		hierarchical processing, slice data, height, slice
		figure, modeling
086	Arc additive	Arc, welding torch, arc additive manufacturing,
	manufacturing	particle, current, welding wire, tubular, molten pool,
		tungsten electrode, waveform, silk machine, heat
000	NF ( 1 1	source, voltage
092	Metal powder	Metal, metal powder, screening, metal parts, metal
		printing, screen mesn, sieving, particle size, alloy
		granulation allow belt
095	Building	Coagulation concrete side namel wall reinforced
0,0	Dununig	masonry, exterior wall, concave-convex, floor.
		masonry shell, building, shear wall, steel wire,
		concrete shell, shafting
117	Unmanned	Machine, robot, rectangular, man-machine, making
	manufacturing	machine, unmanned, additive manufacturing
		machine, metal plate, control cabinet, conveyor belt,
		software
121	Gel material	Solution, gel, aqueous solution, protein, gelatine,
		crosslinking agent, freeze-dried, gel materials,
		colloid, water bath, gel stent, hydrophilic, soluble,
		sait solution, colloidal

upgrading, and declining. We examined these stages in detail, offering suggestions for companies of the industry as to how these insights might influence the formulation of their own R&D strategies.

Comparing our results with the current R&D trends in 3D printing provides a reasonable barometer for the accuracy of the method. In

#### Table 4

Terms of example technology topics in updating stage.

Techr	ology topic	High probability terms
008	Feeding device	Feeding, shaft, feed port, feed tube, cone gear, bevel gear, feeding mechanism, belt, feeding device, feeding channel, feed opening, feed end, spray material, feed pipe
030	Air exhausting device	Trachea, fan, air, tuyere, exhaust, purification, diversion, hot air, duct, ventilation, outlet, dust removal, suction, air pump, cooling device, air duct, ventilation, blast
064	Ceramic materials	Ceramic, paste, powder, porcelain powder, ceramic powder, ceramic slurry, green body, ceramic materials, ball mill, skim, powder material, metal ceramic, dispersant
067	Biological materials	Biological, cells, substrate, soluble, factor, biological materials, cartilage, protein, water soluble, engineering scaffolds, bone tissue, support material, biological print, stem cells, bone cells, biological scaffolds, bone tissue engineering, gelatine
130	Personalized customization	Software, modeling, personalized, dimension, patients, laser district, constituency of melting, finite element, 3d reconstruction, scanner, facemask, Personalized customization, medical, digital model, reverse engineering, lesions

#### Table 5

Terms of example technology topics in declining stage.

Technology topic		High probability terms
039	Plastic material	Plastic, thermoplastic, ammonia ester, polyurethane, packing, polyester, water-based, isocyanate, inorganic
		thermoplastic material, Moore, thermoplastic elasticity, diisocyanate, acid ester
055	Mechanical arm	Mechanical, mechanical arm, microspheres,
		manipulator, mechanical properties, mechanical
		processing, central control, terrain, arm, mechanical strength numping
112	Sensing device	Sensor, temperature sensor, crystal lattice, humidity.
		ranging sensor, measurement system, lattice structure,
		panel, sensor installation, humidity sensor, sensor
		module, sensitivity, liquid level sensor, laser
		displacement, velocity sensor, sensor connection
123	Material stock	Material stock, processor, container, transporting,
	device	transporting pipe, scanner, ranging, box body, storage
		bin, central processing, display, material storage tank,
		laser ranging, control valve, gantry, transporting
		organization, clay, storage bin
136	Photocuring	Resin, photocuring, photocuring resin, resin material,
	resin	photocuring material, resin composite, resin curing,
		photocuring printing, polymerization retarder

recent years, research on the intersections between 3D printing and related disciplines has been a hot trend, among which bio-printing and robot printing appear to be the most promising prospects. In our case, Topics 061, 091 are in the emerging stage, Topic 121 is in the growing stage, and Topic 067 is in the upgrading stage, and fit with the trend toward integrating 3D printing with biology. Looking deeper into bioprinting the most general topic "biomaterials" is in the updating stage; "artificial blood vessels" and "bioglass" are in the emerging stage, while "gel material" is currently developing at a rapid pace. Topic 117 "unmanned printing" is consistent with the trend toward robot printing. The wide application of artificial intelligence and the liberation of human supervision are two great trends of 3D printing in the future. Another trend in 3D printing is metal printing. Topics 041, 086, 092 are in the growing stage, and Topics 039, 136 are in the declining stage, and are consistent with this trend. At present, the main materials of 3D printing are ABS (acrylonitrile butadiene styrene plastic) and PLA (polylactic acid). Both have the disadvantage of low strength. Although improved hard materials have emerged in recent years, they are still essentially plastic, and their uptake has been limited. Therefore, the decline of plastic 3D printing is occurring alongside the growth of metal 3D printing. Today, metal 3D printing is possible with alloys based in aluminum, titanium, cobalt-chromium, stainless steel, iron-nickel, and others. However, there are still problems to be overcome, such as low material density and poor surface accuracy, so metal 3D printing techniques are still being developed and improved. Overall, our results are highly consistent with the actual current trends in 3D printing technology.

However, there are some limitations to this research. First, the extent to which a patent contributes to the overall level of a technology tends to vary with its importance. In this study, we treated each patent equally. Second, some invalidated patents may have contained noise, despite our attempts to only include invalidated patents caused by direct action on the part of the applicant in the sample, some patents may have been waived due to mergers, transfers, or the establishment of new subsidiaries, which would reflect the corporation's own strategies, not a general trend in the industry. Lastly, some technologies could not be identified from the results of the topic model. The meanings of some high probability terms in a topic were too broad or too confusing to correspond to a certain technology. This is a limitation of the model.

In future research, we plan to assign weights according to a set of patent indicators to more accurately estimate the importance of each patent given the technology under consideration. Identifying patents that companies have abandoned as a result of changes in business strategy would also minimize a bias, and help to make the results more accurate. Further, we may explore using the full text of the patents in the future so as to meet the higher data demands of natural language processing. Importantly, this method still relies on experts to clear the corpus and topic list manually. Therefore, effort will be spent on automating aspects of the framework through machine learning methods. This final key direction is a method of effectively identifying core invalidated patents that could present future technology opportunities.

#### CRediT authorship contribution statement

Xiaotong Han: . Donghua Zhu: Conceptualization, Software, Resources, Supervision, Project administration, Funding acquisition. Ming Lei: Validation, Investigation. Tugrul Daim: Writing - review & editing.

# Acknowledgments

This work is supported by the General Program of the National Natural Science Foundation of China (Grant nos. 71673024 and 71774012). The findings and observations in this paper are those of the authors and do not necessarily reflect the views of the supporters. The authors would like to thank colleagues from Beijing Institute of Technology, the China Patent Information Center.

# References

- Blei, D.M., Ng, A.Y., Jordan, M.I., 2003. Latent dirichlet allocation. J. Mach. Learn. Res. 3, 993–1022.
- Carlsson, B., Stankiewicz, R., 1991. On the nature, function and composition of technological systems. J. Evol. Econ. 1 (2), 93–118.
- 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. Daim, T., 2012. Patent analysis of wind energy technology using the patent alert system. World Patent Inf. 34 (1), 37–47.
- Daim, T., Grueda, G., Martin, H., Gerdsri, P., 2006. Forecasting emerging technologies: use of bibliometrics and patent analysis. Technol. Forecast. Soc. Change 73 (8), 981–1012.
- Ernst, H., 1997. The use of patent data for technological forecasting: the diffusion of CNC-technology in the machine tool industry. Small Bus. Econ. 9 (4), 361–381.
- Foster, R.N., 1986. Working the S-curve: assessing technological threats. Res Manag. 29 (4), 17–20.
- Fu, Y., 2005. Strategy for development and utilization of invalid Patent. R & D Manag. 17 (3), 96–100.
- Geum, Y., Jeon, J., Seol, H., 2013. Identifying technological opportunities using the novelty detection technique: a case of laser technology in semiconductor manufacturing. Technol. Anal. Strateg. Manag. 25 (1), 1–22.

#### X. Han et al.

- Gibson, E., Blommestein, K., Kim, J., Daim, T., Garces, E., 2017. Forecasting the electric transformation in transportation. Technol. Anal. Strateg. Manag. 29 (10), 1103–1120.
- Gonçalves Pereira, C., Lavoie, J., Garces, E., Basso, F., Dabić, M., Porto, G., Daim, T., 2019. Assessment of technologies: forecasting of emerging therapeutic monoclonal antibodies patents based on a decision model. Technol. Forecast. Soc. Change 139, 185–199.
- Griffiths, T.L., Steyvers, M., 2004. Colloquium paper: mapping knowledge domains: finding scientific topics. Proc. Natl. Acad. Sci. U.S.A. 101, 5228. Suppl 1.
- Gu, D., Yang, P., 2010. Development and utilization for information of the public known and used patent technology. Libr. Inf. Serv.
- Heinrich, G. 2008. Parameter estimation for text analysis. Technical Report. Hsu, F.C., 2006. Technology and knowledge document cluster analysis for enterprise
- R&D strategic planning. Int. J. Technol. Manag. 36 (4), 336–353.II, J.A.P, 2006. How companies can preserve market dominance after patents expire. Long Range Plann. 39 (1), 71–87.
- Jin, G., Jeong, Y., Yoon, B., 2015. Technology-driven roadmaps for identifying new product/market opportunities: use of text mining and quality function deployment. Adv. Eng. Inform. 29 (1), 126–138.
- Kim, B., Gazzola, G., Lee, J.M., Kim, D., Kim, K., Jeong, M.K., 2014. Inter-cluster connectivity analysis for technology opportunity discovery. Scientometrics 98 (3), 1811–1825.
- Kim, B., Gazzola, G., Yang, J., Lee, J.M., Coh, B.Y., Jeong, M., Jeong, Y.S., 2017. Twophase edge outlier detection method for technology opportunity discovery. Scientometrics 113 (1), 1–16.
- Kim, C., Jeon, J.H., Kim, M.S., 2015a. Identification and management of opportunities for technology-based services: a patent-based portfolio approach. Innov. Manag. Policy Pract. 17 (2), 232–249.
- Kim, G.J., Sang, S.P., Jang, D.S., 2015b. Technology forecasting using topic-based patent analysis. J. Sci. Ind. Res. 74 (5), 265–270.
- Lai, Y., Zhu, D., Liu, Y., 2007. Study on the analysis of patent legal-status information and its empirical research. J. Inf.
- Lee, SooMee, 2012. Why are most of Korean patents invalid? Law Stud. 15, 31-56.
- Lee, C., Kang, B., Shin, J., 2015. Novelty-focused patent mapping for technology opportunity analysis. Technol. Forecast. Soc. Change 90 (B), 355–365.
- Lee, M., Lee, S., 2017. Identifying new business opportunities from competitor intelligence: an integrated use of patent and trademark databases. Technol. Forecast. Soc. Change 119.
- Lee, S., Yoon, B., Park, Y., 2009. An approach to discovering new technology opportunities: keyword-based patent map approach. Technovation 29 (6–7), 481–497.
- Lee, W.S., Sohn, S.Y., 2017. Identifying emerging trends of financial business method patents. Sustainability 9 (9), 1670.
- Lee, Y., Kim, S.Y., Song, I., Park, Y., Shin, J., 2014. Technology opportunity identification customized to the technological capability of SMEs through two-stage patent analysis. Scientometrics 100 (1), 227–244.
- Leslie, C.R., 2006. The anticompetitive effects of unenforced invalid patents. Minn. Law Rev. 91 (1), 101–183.
- Li, S., Garces, E., Daim, T., 2019. Technology forecasting by analogy based on social network analysis: the case of autonomous vehicles. Technol. Forecast. Soc. Change 148.
- Lin, X., Xie, Q., Daim, T., Huang, L., 2019. Forecasting technology trends using text mining of the gaps between science and technology: the case of perovskite solar cell technology. Technol. Forecast. Soc. Change (146), 432–449.
- Ma, J., Porter, A.L., 2015. Analyzing patent topical information to identify technology pathways and potential opportunities. Scientometrics 102 (1), 811–827.
- Ma, T.T., Porter, A.L., Guo, Y., Ready, J., Xu, C., Gao, L.D., 2014. A technology opportunities analysis model: applied to dye-sensitised solar cells for China. Technol. Anal. Strateg. Manag. 26 (1), 87–104.
- Madani, F., Daim, T., Weng, C., 2017. Smart building technology network analysis: applying core periphery structure analysis. Int. J. Manag. Sci. Eng. Manag. 12 (1), 1–11.

- Meyers, P.W., 2004. Non-linear learning in large technological firms: period four implies chaos. Res. Policy 19 (2), 97–115.
- Momeni, A., Rost, K., 2016. Identification and monitoring of possible disruptive technologies by patent-development paths and topic modeling. Technol. Forecast. Soc. Change 104, 16–29.
- Park, Y., Yoon, J., 2017. Application technology opportunity discovery from technology portfolios: use of patent classification and collaborative filtering. Technol. Forecast. Soc. Change 118, 170–183.
- Penrose, E.T., Wolfe, J.N., 2009. The theory of the growth of the firm. Manag. Decis. Econ. 2 (3), 192–193.
- Porter, A.L., Detampela, M.J., 1995. TECHNOLOGY OPPORTUNITIES ANALYSIS. Technol. Forecast. Soc. Change 49 (3), 237–255.
- Schumpeter, J.A., 1911. The theory of economic development. J. Polit. Econ. 1 (2), 170–172.
- Song, K., Kim, K.S., Lee, S., 2017. Discovering New Technology Opportunities Based on Patents: Text-Mining and F-Term Analysis. Technovation, pp. 60–61, 1-14.
- Wang, B., Liu, S., Ding, K., Liu, Z., Xu, J., 2014. Identifying technological topics and institution-topic distribution probability for patent competitive intelligence analysis: a case study in LTE technology. Scientometrics 101 (1), 685–704.
- Wang, X.F., Ma, P.P., Huang, Y., Guo, J.F., Zhu, D.H., Porter, A.L., Wang, Z.N., 2017. Combining SAO semantic analysis and morphology analysis to identify technology opportunities. Scientometrics 111 (1), 3–24.
- Xing, S., 2013. Analysis of the concept of ineffective patent. Sci. Technol. Manag. Res. 33 (3), 156–160.
- Yoon, B., Park, Y., 2005. A systematic approach for identifying technology opportunities: keyword-based morphology analysis. Technol. Forecast. Soc. Change 72 (2), 145–160.
- Yoon, B., Park, I., Coh, B.Y., 2014. Exploring technological opportunities by linking technology and products: application of morphology analysis and text mining. Technol. Forecast. Soc. Change 86, 287–303.
- Yoon, J., Park, H., Kim, K., 2013. Identifying technological competition trends for R&D planning using dynamic patent maps: sAO-based content analysis. Scientometrics 94 (1), 313–331.
- Yoon, J., Park, H., Seo, W., Lee, J.M., Coh, B.Y., Kim, J., 2015. Technology opportunity discovery (TOD) from existing technologies and products: a function-based TOD framework. Technol. Forecast. Soc. Change 100, 153–167.
- Zhu, D., Porter, A.L., 2002. Automated extraction and visualization of information for technological intelligence and forecasting. Technol. Forecast. Soc. Change 69 (5), 495–506.
- Zhu, D., Porter, A., Cunningham, S., Carlisie, J., Nayak, A., 1999. A process for mining science & technology documents databases, illustrated for the case of "knowledge discovery and data mining". Inf. Sci. 28 (1), 7–14.

Xiaotong Han is a PHD candidate of the School of Management and Economics, Beijing Institute of Technology. Her current specialty is technology innovation management and tech mining.

Donghua Zhu is a Professor of the School of Management and Economics, Beijing Institute of Technology and was a Senior Visiting Scholar at School of Public Policy, Georgia Institute of Technology. His-current specialty is data mining, technology assessment and forecasting.

Ming Lei is a PHD candidate of the School of Management and Economics, Beijing Institute of Technology. His-current specialty is technology innovation management and tech mining.

Tugrul Daim is a Professor of Engineering and Technology Management, Portland State University and an editor in Chief IEEE Transactions on Engineering Management. Hiscurrent specialty is technology forecasting.