Futures 118 (2020) 102530

Contents lists available at ScienceDirect

Futures

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

Essays

Identifying technology evolution pathways using topic variation detection based on patent data: A case study of 3D printing

Zhongzhen Miao[a], Junfei Du[b], Fang Dong[a], Yufei Liu[c,*], Xiaochuan Wang[b]

[a] School of Public Policy and Management, Tsinghua University, Beijing, 100084, China
[b] School of Mechanical Science & Engineering, Huazhong University of Science and Technology, Wuhan, 430074, China
[c] Center for Strategic Studies, Chinese Academy of Engineering, Beijing, 100084, China

ARTICLE INFO

Keywords:
Technology evolution
Topic variation detection
Text mining
Patent time series
3D printing

ABSTRACT

Previous studies on identifying technology evolution pathways have ignored the point in time when an evolution pathway changes. To fill this gap in the literature, we use a novel approach based on text-mining to identify technology evolution pathways. First, the topic model was applied to discover technology topics, and topic variation through time was modeled. Second, critical junctures were detected based on topic variation to generate time segments. Finally, the main topics' changes at different segments were analyzed to identify the pathway, and a visualization of the main topics' trends were presented. To demonstrate the effectiveness of the text-mining approach, we examined 3D printing technologies using 34,090 patents from 1990 to 2017, and the database we used are updated weekly. We found that traditional technologies showed a declining trend, and their practical application technologies related to products is promising. The new approach can be applied to identify technology evolution pathways characterized by critical junctures. These critical junctures are helpful in understanding technological development more clearly, especially in identifying what and when technological changes occur.

1. Introduction

Patent documents contain rich details about research results that are valuable for both innovation analysis and technology trend analysis (Tseng, Lin, & Lin, 2007; Wang, Liu, Zhou, & Wen, 2018; Wang, Urban, Zhou, & Chen, 2018). Hence, patent analysis has been widely used to identify evolution pathways of various technologies, which aims to trace technological development over time (Chang, Wu, & Leu, 2010; Chen, Huang, & Chen, 2013; Ganguli, 2004; Kropsu-Vehkapera, Haapasalo, & Rusanen, 2009; Ma & Porter, 2015; Wang, Liu et al., 2018; Wang, Urban et al., 2018; Wu, 2016; Yoon, 2011; Yoon & Kim, 2011; Zhou, Li, Lema, & Urban, 2015; Zhou, Pan, & Urban, 2018). Much effort has been devoted to these studies as well as different kinds of approaches, such as patent citation analysis, main path analysis, and patent classification analysis. While these approaches have contributed meaningfully to technology evolution pathway analysis, they have failed to assess the points in time that an evolution pathway changes. These time points are critical to analyze and illustrate the technology development process comprehensively, especially for determining when and what technological changes happened. Thus, we refer to these points as *critical junctures* in this paper.

This paper uses text-mining to identify technology evolution pathways. The pathway was identified by detecting critical junctures and analyzing the transition of main topics over time. First, the topic model was used based on patent data to discover topics, and time information was introduced to uncover the topic trends. Second, the topic variation detection method was applied to detect

* Corresponding author at: Center for Strategic Studies, Chinese Academy of Engineering, Xi Cheng, Beijing 100088, China.
E-mail address: liuyufei0418@qq.com (Y. Liu).

https://doi.org/10.1016/j.futures.2020.102530
Received 30 March 2019; Received in revised form 4 February 2020; Accepted 4 February 2020
Available online 07 February 2020
0016-3287/

critical junctures based on topic trends. Then, time segments representing different stages of pathways became clear with the junctures. Third, the transition of the main topics among the segments was analyzed to identify the pathway. Finally, a visualization method was used to demonstrate the pathway and the technology trends. The technology evolution pathway identified through the proposed approach contained critical junctures. This paper explains the text mining approach with the time contribution of the technology evolution pathway. Moreover, the main contribution of our study is application innovation rather than method innovation.

2. Literature review

2.1. Technological evolution

Technological evolution has been examined in several studies. In general, technological evolution is defined as technological changes or technology development over time (Adomavicius, Bockstedt, Gupta, & Kauffman, 2007; Chen et al., 2013; Huang et al., 2017a; Iansiti, 1995). We defined the term "technology evolution pathways" to mean the representation of technology evolution.

Studies conducted to identify technology evolution based on patent data have used bibliometric analysis or topical analysis. Several bibliometric methods can be used to identify traits of patents, including citations, patent classification, and patent semantics. Patent citation analysis enables the trajectory of knowledge to be traced using citation chaining patents (Huenteler, Ossenbrink, Schmidt, & Hoffmann, 2016), as citations are effective in identifying technology evolution. Chen et al. (2013) used citation chaining to map the document network and hierarchical structure needed to generate clusters containing similar documents. The clusters were used to detect and analyze technology evolution. Based on the citation network, Huang, Zhang, Ma, Porter, and Wang (2015); Huang et al. (2017a) and Verspagen (2007) applied the main path analysis to detect the dominant patent trajectory in order to trace technology evolution.

Patent classification can also be used to trace the technology evolution (Wang, Sung, & Huang, 2016; Zhou, Zhang, Porter, Guo, & Zhu, 2014). Wang et al. used International Patent Classifications (IPCs) to trace technological changes in technology fields and in the evolution pathways. Another approach utilizes keywords to trace technology evolution. For example, Yoon and Park (2004) proposed a text-mining approach for patent documents to generate a semantic network that reflected technology trends. Kim, Suh, and Park (2008) proposed a similar approach to develop a semantic network of keywords, and subsequently developed patent maps to illustrate the technology evolution of emerging technologies. These studies paid little attention to the role of timing in technology evolution. Natural language processing (NLP) is also used to trace technology evolution. For example, Ma and Porter (2015) proposed a new framework that used topical analysis to conduct another analysis to obtain a terms and phrases list using VantagePoint's NLP. They processed the results using manual analysis to identify potential innovation pathways and technological opportunities. However, all the above researchers paid little attention to the role of timing in technology evolution.

In recent years, the timing of evolution pathways has attracted attention. The life cycle of technology has been studied to identify the technology evolution pathway of 3D printing (3DP) technologies (Huang et al., 2017b). Huang et al. (2017b) applied the S-curve method to divide the life cycle of 3DP technologies into two periods; then, co-classification analysis, co-word analysis, and main path analysis were performed to analyze the technology evolution pathway. Since the S-curve method is based on the number of patents, changes in technological content, such as topics, were not considered. Variations in technological content can also lead to changes in pathways.

In summary, few studies have focused on points in time when significant technological changes happen. Patent document data, a typical form of time-series data that includes time information and technological content information, is hidden in text data. Thus, critical junctures can be analyzed by using time and text information from patent data.

2.2. Topic modeling

Topic modeling algorithms are statistical methods used to analyze the words of original texts to discover latent semantic topics (Blei, Carin, & Dunson, 2010). Proposed by Blei, Ng, and Jordan (2003); Latent Dirichlet Allocation (LDA) is one of the best known. This unsupervised learning approach estimates the properties of multinomial observations and latent semantic topics hidden in a number of documents and also details how various documents belong to different topics (Blei, 2012).

LDA is frequently employed to discover topics. Xing and Croft (2006) proposed an LDA-based document model within the language modeling framework for ad-hoc retrieval. Similarly, Kim and Oh (2011) used a framework based on LDA to discover important topics and their meaningful structure within online news archives. Yang et al. (2013) introduced a topic expertise model based on LDA to uncover the meaningful structure and trends of important topics and issues hidden within online news archives.

In addition, LDA is a powerful tool for technology management when applied to patent data. For example, Chen, Zhang, Zhang, and Zhu (2015) applied LDA to identify technology topic changes based on patent data. Chen, Zhang, Zhu, and Lu (2017a), they also proposed a technological forecasting approach based on LDA to discover and estimate the trends for specific topics based on text from lots of patent claims. This method has significantly contributed to text mining studies; hence, we selected this as the underlying basis for our framework of topic discovery.

2.3. Topic variation model

As mentioned above, data on timing and technological content are needed to detect critical junctures. Patent data contains information on timing and technological content (Xie & Miyazaki, 2013), so patent topics usually change with the development of the corresponding technology over time. This process represents the technology evolution pathway to a certain degree, and critical junctures can be detected by identifying all changes in patent topics. While few approaches exist for analyzing patent documents' topic changes, existing studies have processed the textual data of news while focusing on the chronological order of each news story's publication (Hu et al., 2011; Wang, Shi, Dong, & Ge, 2013; Yan et al., 2011).

Hu et al. (2011) focused on the critical junctures of news coverage of a developing story and proposed an approach to detect the critical junctures in the narrative based on timing and text data. A critical juncture is a specific point in time when decisive changes occur in the development of a topic. Like news, patent data have text and temporal information that can be processed using similar approaches. The critical junctures can be detected using this method based on data extracted from patent documents. Thus, a solution can be provided to detect critical junctures for the technology evolution pathway.

Important dates or key phases relating to technology evolution pathways can change dramatically and are meaningful. The time points of changes in technology evolution pathways can be detected by pinpointing technological topic changes based on related time-series text documents. The rapid development of information technology provides methods for different academic disciplines (Liu et al., 2019).

As for textual data analysis of textual documents including website comments, news, academic papers, and micro-blogging sites, topic changes have been successfully detected (Cuneyt, Akcora, Bayir, Demirbas, & Ferhatosmanoglu, 2010; Hu et al., 2011; Wang et al., 2013; Yu, Meng, & Yu, 2011).

For instance, Hu et al. (2011) studied automatic generation of critical junctures based on timeline overview for a news topic that paid specific attention to critical junctures, whereas other researches focused on the content of the topic changes. Critical junctures can highlight key phases of a news topic and can be used to map the outline of the whole story. They can also contribute to the development a novel method that identifies critical junctures by modeling topic activities and theme variation using time and textual data.

Moreover, patent documents contain text segments whose topics can represent technological topics and references to timing. Hu et al. (2011) paid attention to identify the time points of topic changes so as to identify the critical junctures of patent topics to track important dates or key phases of technology evolution pathways changes. Our study is based on the work of Hu et al. (2011) to identify important dates that delineate the technology evolution pathway changes, and to produce a topic variation detection model.

3. Methodology

This section explains our approach to identify the technology evolution pathway using the topic variation model. We provide the framework for topic variation detection approach used in this paper with detailed explanations and parameters for each step. Our framework integrates the LDA topic model and topic variation model. The research contribution consists of the combination of existing methods and its application to technology evolution pathway identification.

3.1. Framework

Fig. 1 illustrates the framework of our proposed approach. First, patent data was collected, and the textual data from the patents were cleaned. Second, the topic number was determined according to its perplexity index (Mazin & Junlan, 2009), and the LDA model was applied to discover technological topics from the patent textual data and the distributions of various documents on different topics. Third, the textual data were labeled by date, including the year and month. We used the topic weight to indicate the topic contribution to all documents (Chen, Zhang, Zhu, & Lu, 2017b). According to the distribution, the monthly topic weight of all topics was calculated to generate a monthly topic weight matrix. Then, the topic variation detection approach was utilized to detect the time points when topics changed dramatically, i.e., at critical junctures. Fourth, the time span was divided into segments by using each critical juncture as a milestone to clearly mark different segments that represented stages of technology evolution. We assessed the various time segments to reveal the main topics of each segment and identify the pathway. We then used the visualization method to demonstrate the pathway and the main topics' trends.

3.2. Data collecting and preprocessing

We used the Derwent World Patent Innovation (DWPI) database to retrieve patents because it is a comprehensive database of detailed patent documents and has been used for related studies (Chen et al., 2017a; Kong, Zhou, Liu, & Xue, 2017). The database covers more than 77.6 million patents from 52 patent authorities worldwide and two literature sources. The DWPI database is updated weekly from the authorities.

We downloaded the needed information on patents, including the title of the DWPI (Derwent World Patent Innovation), the abstract of the DWPI, the first claim in English, the first claim of DWPI, and the earliest priority date. As the title, abstract, and claim were reformatted by DWPI to improve the quality of the description, the resulting data were suitable for topic modeling.

If the record of the first patent claim was empty, the first claim in English was used instead. Patents without an earliest priority date or textual data were jettisoned. Subsequently, the patent text documents were sorted according to the earliest priority date to form a time series. The textual data, including titles, abstracts, and claims, contained a large number of stop words, such as *the*, *this*, *that*, along with noise words, punctuations, and HTML fragments. Before putting them through topic modeling with LDA, stop words and noise words were filtered out from the patent text documents.

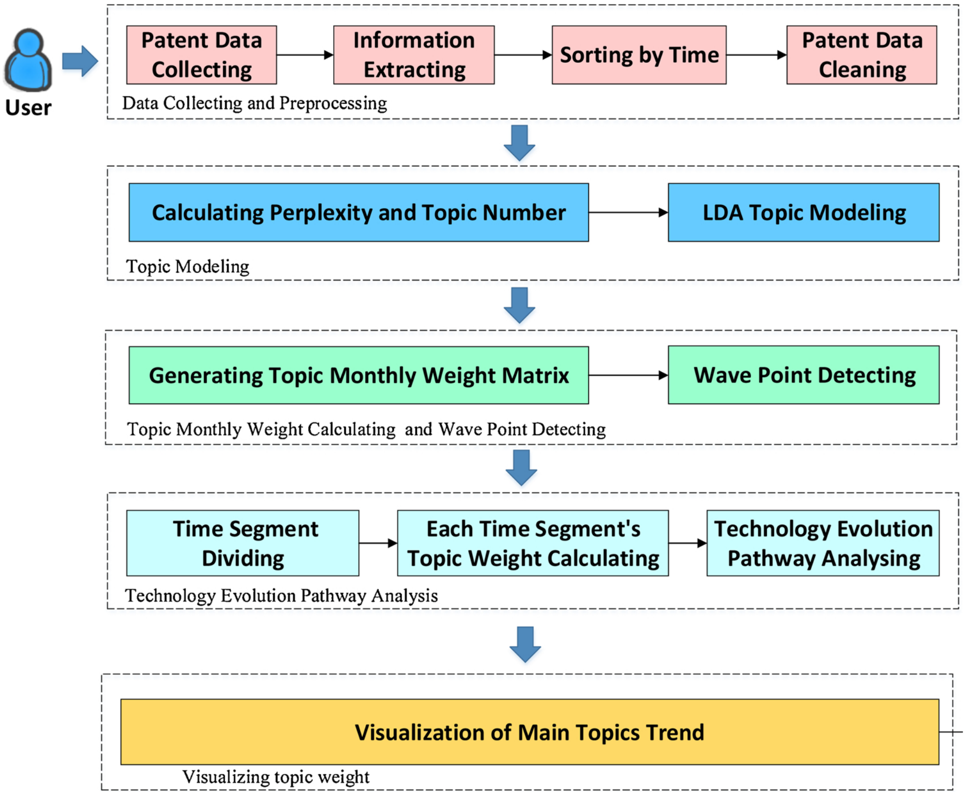

Fig. 1. Framework of our approach.

3.3. Patent topic discovery and topic weight calculation

After collecting and preprocessing the data, the LDA topic model is used on the processed data to extract potential topic information. This approach was used for unsupervised topic modeling, which can be used to extract hidden topics from patent documents to discover technological topics.

The structure and generative process of LDA are explained in Appendix B1. Before topic modeling with LDA, the parameters, β, *iteration*, and K must be determined. Dirichlet hyperparameters α and β significantly influenced the performance of the model and generally had a smoothing effect on the multinomial parameters. Lowering the values of α and β reduced this smoothing effect and resulted in more robust topic associations. We set the parameters at $\alpha = 50/K$; $\beta = 0.01$ and the $iteration = 5000$ based on precedent (Chen et al., 2017a; Chen et al., 2015; Ma & Porter, 2015; Xing & Croft, 2006). The number of topics, K, was determined using the perplexity index, which is in Appendix B2. We then input these factors into the LDA model to discover technological topics and the topic structure.

After the topic structure was identified, a topic weight matrix was used to quantify topic trends and show how topic weight changed over time. A month was used as the time period in order to detect technological content changes based on the discovery of technological topics. The time-series patent documents then had to be grouped by month, as shown below (Chen et al., 2017a):

$$\begin{aligned} P = \{&L_1(p_1, p_2, \ldots, p_{t_1}), L_2(p_{t_1+1}, p_{t_1+2}, \ldots, p_{t_2}), \ldots,\\ &L_i(p_{t_{i-1}+1}, p_{t_{i-1}+2}, \ldots, p_{t_i}), \ldots,\\ &L_s(p_{t_{s-1}+1}, p_{t_{s-1}+2}, \ldots, p_r)\}, \end{aligned} \tag{1}$$

where P is the collection of all patent documents; L is the time period divided by month; L_i indicates the i_{th} month; p is a patent document (the subscript indicates the number of patent documents); t_i denotes the number of the last patent document in L_i; and r is the total number of patent documents in P.

The topic weight vector for each month was then established based on the topic structure. We developed the program of LDA topic model by Java to conduct LDA topic modeling and get results. The “tassign” file, generated by the LDA topic model was part of the topic structure; each row represents an input patent document, and each item represents words in the document and the topic that matched those words. Table 1 gives an example of the “tassign” file. The leftmost column indicates the patent document number, and the other columns denote words in the patent documents. For each item, the number of words is shown to the left of the colon, and the topic number is on the right. For example, “0:18” of patent No.1 indicates that the patent consisted of the word “No.0,” and that the word “No.0” belonged to the topic No.18. Different words could also belong to the same topic, and different patent documents

Table 1
Example of tassign file.

Patent document number	Words in patent document					
1	0:18	1:10	2:3	3:10	4:4	5:18
2	32:9	33:16	34:19	35:3	36:6	37:9
3	187:11	33:0	89:16	188:10	189:14	190:8
4	233:10	234:18	235:17	31:12	236:3	237:14
5	243:3	244:13	245:3	246:5	247:11	248:13
6	276:6	277:14	278:5	279:15	280:11	82:6

could contain some of the same words. Thus, for patent No.6, topic 6 appeared twice. For each patent row, words are ordered according to this patent document rather than the topic number.

When calculating topic weight, the patents of that same month were considered cumulatively, and the monthly topic weight vector was calculated as follows (Hu et al., 2011):

$$weight(z_k,\ t) = \frac{\sum_{j=1}^{d_t} n_{jk}}{\sum_{j=1}^{d_t} N_j}, \tag{2}$$

where z_k is the k_{th} topic; $weight(z_k,\ t)$ is the weight of the kth topic in the t_{th} month; d_t is the amount of patent documents in the t_{th} month; N_j is the amount of words in the j_{th} patent document; n_{jk} is the amount of words in the j_{th} patent document that belonged to the k_{th} topic; $\sum_{j=1}^{d_t} N_j$ is the amount of words in all patent documents in the t_{th} month; and $\sum_{j=1}^{d_t} n_{jk}$ indicates the amount of words belonging to the kth topic among the words of all patent documents in the t_{th} month.

The calculated values of the monthly topic weight $weight(z_k,\ t)$ were then aggregated to form the monthly topic weight matrix. This matrix (W) consisted of T rows and K columns, where T indicates the total number of months in Y years covered by all patents, and $T = 12*Y$. K is the number of topics. In the matrix W, each row represents a month, each column represents a topic, and each item, W_{tk}, represents the weight of the kth topic in the t_{th} month. The sum of the weights of all the topics in each row is 1.

3.4. Topic variation detection

After calculating the monthly topic weight, the topic variation detection approach was applied to identify the time points when technological topics changed dramatically and reached critical junctures. Based on the critical junctures, the evolution pathway was then identified.

A critical juncture was defined as a time point when topics changed more significantly compared to previous and later time points on the path, and time was measured in months. We had to calculate the distance between the topic weight vectors of adjacent time points t and t + 1 to measure the topic variation $TV(t,\ t+1)$. Then, $TV(t,\ t+1)$ was calculated using a standard distance measure: the square root of the Jansen Shannon divergence (Hu et al., 2011). The following equation describes the method:

$$TV(t,t+1)=\sqrt{\frac{1}{2}\sum_{k=1}^{K} weight(z_k,t)\log\frac{weight(z_k,t)}{A_{z_k}} + \frac{1}{2}\sum_{k=1}^{K} weight(z_k,t+1)\log\frac{weight(z_k,t+1)}{A_{z_k}}}, \tag{3}$$

where k denotes the kth topic; K stands for the number of topics; $weight(z_k,\ t)$ is the weight of the kth topic in the t_{th} month; and $A_{z_k} = \frac{1}{2}(weight(z_k,\ t) + weight(z_k,\ t+1))$.

The time point t is a critical juncture if $TV(t\text{-}2,t\text{-}1)$, $TV(t\text{-}1,\ t)$, and $TV(t,t\ +1)$ satisfy the following criteria (Hu et al., 2011):

$$\begin{cases} TV_{(t-1,t)} > 1.5TV_{(t-2,t-1)} \\ TV_{(t-1,t)} > 1.5TV_{(t,t+1)} \end{cases}. \tag{4}$$

This means that a significant topic variation existed at point t compared to previous and subsequent time points. However, these criteria did not apply to the first time point and the last two time points because there was no previous or later time point.

3.5. Identifying technology evolution pathway and visualizing topic trends

In order to observe critical technological changes, the time range was divided into several segments based on critical junctures. We used topic weight changes between different segments to represent technological changes. The method for obtaining the topic weight vector of each segment was based on the "tassign" file. Similarly, the topic weight vector of each time segment was obtained as follows (Hu et al., 2011):

$$weight(z_k,\ TS_i) = \frac{\sum_{j=1}^{d_i} n_{jk}}{\sum_{j=1}^{d_i} N_j}, \tag{5}$$

where z_k is the kth topic; TS abbreviates for time segment; $weight(z_k,\ TS_i)$ is the weight of the kth topic in the i_{th} time segment; d_i is the number of patent documents in the i_{th} time segment; N_j is the number of words in the j_{th} patent document; n_{jk} is the number of words in the j_{th} patent document that belonged to the kth topic; $\sum_{j=1}^{d_i} N_j$ indicates the total number of words in all patent documents in the i_{th} time segment; and $\sum_{j=1}^{d_i} n_{jk}$ indicates the total number of words belonging to the kth topic among the words of all patent documents in the i_{th} time segment.

After the topic weight vector for each time segment was calculated, the pathway was identified by analyzing the main topic transition in different segments. In order to provide more details about the technological topic changes, we used ThemeRiver to visualize the topic changes. The patents' timeline, the target topics' content, and the topic weight were visualized using the river's flow and changing width (Havre, Hetzler, Whitney, & Nowell, 2002).

4. Results and discussion

To test the efficacy of the proposed approach, we used 3DP as a case study because it is a disruptive technology (Petrick & Simpson, 2013) and has significant influence on both manufacturing processes and business models (Jiang, Kleer, & Piller, 2017).

We applied the topic variation detection method to identify the technology evolution pathway. First, we retrieved the patent data of 3DP technology and identified the topics and corresponding topic weights based on the LDA topic model. Next, the patent data were grouped by month to produce the monthly topic weight matrix needed to detect the changes in topic weight. Subsequently, the topic variation detection method was used to detect critical junctures, which were used to mark the boundary between time segments on the technology evolution pathway. Finally, the technology evolution pathway was drawn using the main topics and the topic weight change of the main topics over all time segments. The results of the LDA and the topic variation model were obtained based on a program we developed using Java ourselves.

4.1. Data collecting and preprocessing

We used the Derwent Innovation patent database provided by Clarivate Analytics as the source of our patent data. We invited three experts in China in the field of 3D printing and organized a workshop to develop a search strategy in June 2018. The three participating experts are Prof. Shi Yusheng, vice chairman of World 3D Printing Technology Industry Association, who has engaged in the research and development of 3D printing technologies for over twenty years; Dr. Liu Huailan, associate professor at Huazhong University of Science and Technology, who is an researcher specialized in studying 3D printing industry; and Dr. Liang Jixiang, researcher at Beijing 3D Printing Institute, who has worked in the 3D printing sector for years. During the workshop, the experts discussed the search strategy composed of key words and time period. The key words including technology concepts, application technologies, and 3D printing methods.

As shown in Table 2; key words used in search of patent data have been listed with the corresponding subfields of those words noted in the left column. The search period has been set to 1990–2017 according to suggestions from experts. The starting years was chosen becasue the quality of patent data is relatively poor before 1990 making it unsuitable for LDA modeling. 2017 is chosen as the end year to avoid truncation problems since patent applications usually require 2–3 years before getting granted. Using the search strategy discussed above, we have collected 34,090 patents. Following the data collection, the timing and technological content of patent data are extracted and cleaned for topic modeling.

Table 2
Key words used in search strategy.

Subfield	Key words
Technology concept	3D printing
	Additive manufact
	rapid prototype
Application technology	Stereo Lithography Apparatus
	Bioprinting
	3D mosaic
	Digital brick lay*
3D printing method	Layered manufact*
	Solid freeform fabrication
	Laminated object Manufact
	Digital Light Process
	Selective Laser Sinter
	Fused Deposition Model*
	Direct Laser Fabrication
	Direct Metal Deposition
	Laser-clad-forming technology
	Electron beam selective melt

Table 3
Part of topic monthly weight matrix.

	Topic1	Topic2	Topic3	Topic4	Topic5
2017-1	0.029034	0.119396	0.070747	0.026438	0.067601
2017-2	0.027692	0.138743	0.051868	0.010741	0.068127
2017-3	0.019968	0.125725	0.055234	0.015576	0.073242
2017-4	0.021141	0.116990	0.050643	0.013233	0.068667
2017-5	0.017732	0.127947	0.051550	0.011304	0.080360
2017-6	0.020318	0.128940	0.057057	0.015757	0.069034
2017-7	0.019510	0.134620	0.055477	0.015841	0.074813
2017-8	0.022493	0.134385	0.042823	0.016561	0.075248
2017-9	0.027597	0.136360	0.069781	0.014083	0.069283
2017-10	0.024592	0.165874	0.044752	0.012856	0.084092
2017-11	0.019406	0.106071	0.077509	0.018747	0.057073
2017-12	0.019519	0.09852	0.068102	0.017055	0.067877

4.2. Patent topic discovery and topic weight calculation

After data was preprocessed, the number of topics was determined by the perplexity index. We chose 20 according to the perplexity index in Appendix B. Then, the LDA model was applied to discover technological topics.

The LDA topic model yielded twenty potential topics. Each topic was represented by the top 20 words ranked by their corresponding probabilities. The words for each topic are listed in Table A1 of Appendix A.

To detect the critical junctures, the topic weights from different time points had to be calculated to quantify topic trends. After calculating the topic weight, we established the monthly topic weight matrix of the 20 topics. Table 3 shows a partial monthly topic weight matrix with some of the topics from 2017 as examples.

4.3. Topic variation point detection

Based on the monthly topic weight matrix, the topic variation detection method was used to detect the topic change points as critical junctures. The topic variation value was calculated with Eq. (3), and the results are shown in Fig. 2. The horizontal axis indicates time, and the vertical axis indicates topic variation value. The red dots in Fig. 2 represent detected critical junctures, while yellow dots are adjacent time points of critical junctures. Eight critical junctures were detected, including January 1990, August 1997, September 2001, October 2003, May 2005, May 2006, April 2010, August 2010, and November 2012 (Fig. 4). Taking the eight critical junctures as the boundary, the 1990–2017 period was divided into nine time segments (Tseng et al.). Table 4 illustrates the details.

In order to show the change in technological topics in each TS, the topic weight for each TS was calculated to represent the topic's contribution to the technological field. Table 5 shows the values of topic weight in each TS. The table is color-coded red to blue to indicate topic weights from high to low. The maximum values of the topic weight in each segment is shown in bold in Table 5. The four topics (1, 2, 6, and 7) were the main topics. Then, the evolution pathway of the 3DP technology was collated by chronologically linking the main topics.

4.4. Technology evolution pathway analysis

Fig. 3 depicts the evolution pathway of 3DP technology. The four topics in each segment were the four top ranked topics in that segment, and the font size was determined according to the rank of topic weight. In order to identify the technological names of the

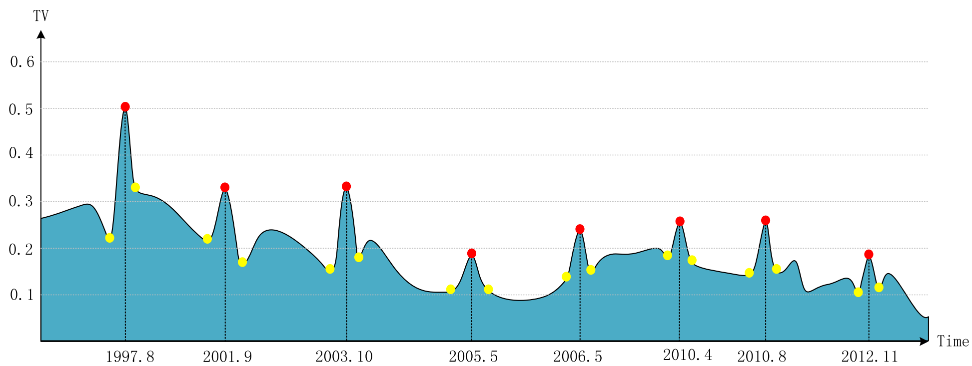

Fig. 2. Topic variation.

Table 4
Time segments.

Time Slot	Beginning	End
TS 1	January 1990	July 1997
TS 2	August 1997	August 2001
TS 3	September 2001	September 2003
TS 4	October 2003	April 2005
TS 5	May 2005	April 2006
TS 6	May 2006	March 2010
TS 7	April 2010	July 2010
TS 8	August 2010	October 2012
TS 9	November 2012	December 2017

Table 5
Topic weights in all time segments.

	TS 1	TS 2	TS 3	TS 4	TS 5	TS 6	TS 7	TS 8	TS 9
Topic1	0.041602	0.050584	**0.139678**	**0.203778**	**0.238329**	**0.178329**	0.100203	**0.104844**	0.028044
Topic2	0.017714	0.012303	0.012316	0.012006	0.017137	0.018993	0.017765	0.021229	**0.081245**
Topic3	0.060380	0.062820	0.047238	0.050133	0.048210	0.050218	0.046043	0.051713	0.056012
Topic4	0.093426	0.109162	0.067488	0.051579	0.057988	0.052623	0.065082	0.065064	0.035836
Topic5	0.032514	0.024273	0.024435	0.026385	0.026970	0.032063	0.037474	0.032259	0.051851
Topic6	**0.163518**	**0.133856**	0.113125	0.097028	0.089530	0.084604	0.086517	0.098671	0.052297
Topic7	0.086696	0.081534	0.060277	0.067268	0.055457	0.073471	**0.128438**	0.076920	0.038850
Topic8	0.017829	0.020586	0.029219	0.033880	0.042892	0.039050	0.030890	0.037064	0.050850
Topic9	0.007631	0.009313	0.011716	0.013743	0.013123	0.016219	0.008591	0.015321	0.055790
Topic10	0.091714	0.096837	0.103805	0.101857	0.069962	0.055941	0.056901	0.064156	0.051763
Topic11	0.024235	0.035380	0.033455	0.022265	0.022154	0.032394	0.046950	0.045640	0.077993
Topic12	0.029878	0.021084	0.021924	0.018782	0.019433	0.026484	0.026724	0.024456	0.038996
Topic13	0.037023	0.030231	0.033104	0.018918	0.017538	0.022480	0.021219	0.035464	0.047984
Topic14	0.018107	0.019821	0.016500	0.018345	0.017382	0.020052	0.020982	0.021127	0.061710
Topic15	0.053604	0.072055	0.084144	0.080756	0.092941	0.088922	0.050857	0.088950	0.046523
Topic16	0.033231	0.055002	0.040532	0.035699	0.030025	0.045243	0.050188	0.055485	0.037412
Topic17	0.039636	0.033187	0.043839	0.051142	0.056773	0.059649	0.069939	0.042832	0.033503
Topic18	0.068381	0.067770	0.066424	0.049232	0.035968	0.045557	0.063787	0.052050	0.055930
Topic19	0.062577	0.050118	0.037629	0.033007	0.033002	0.043039	0.057182	0.046626	0.032333
Topic20	0.020304	0.014085	0.013153	0.014198	0.015185	0.014670	0.014268	0.020129	0.065079

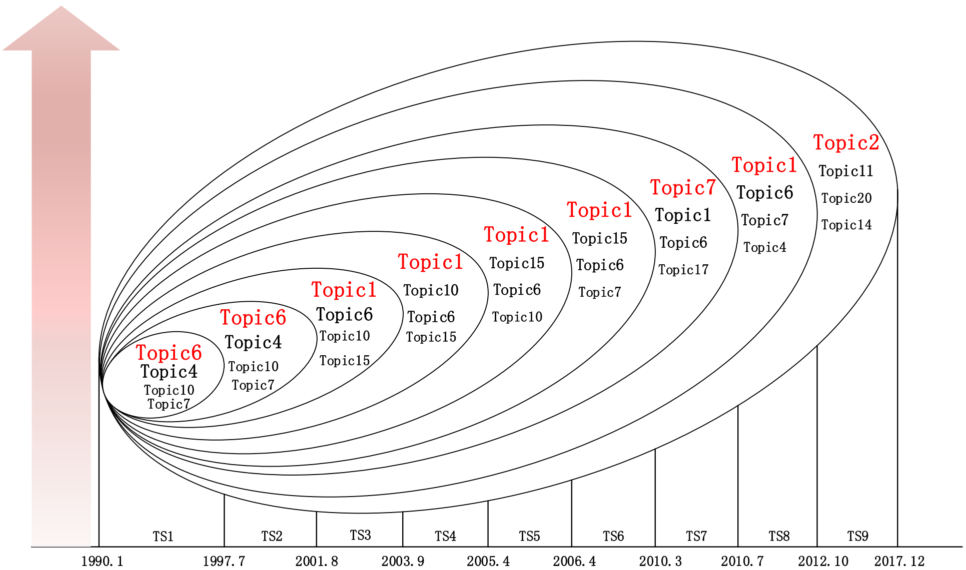

Fig. 3. Technology evolution pathway for 3D printing.

four topics, we asked Dr. Liang and Prof. Liu, as mentioned in Section 4.1, to summarize key words of topics and give technological names. The key words belong to each topic are shown in Table A1 of Appendix A. After thorough discussion with the two experts, we identified the technological names of Topic 1, Topic 2, Topic 6, and Topic 7 as digital light process (DLP), methods and apparatus

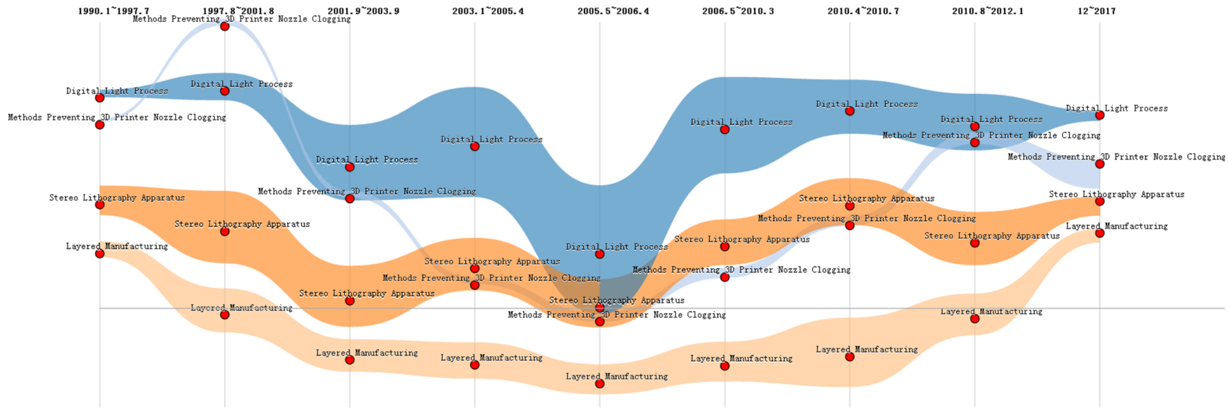

Fig. 4. Topic weight change over time.

preventing 3D printer nozzle clogging (MAP3DPNC), stereo lithography apparatus (SLA), and layered manufacturing (LM), respectively. Topic 6 (SLA) was the most weighted topic of the 3DP technology during TS1 and TS2. However, the most weighted topic eventually became Topic 1 (DLP) during TS3, and during the next three segments, Topic 1 remained the most weighted topic. In TS7, the most weighted topic turned into Topic 7 (LM). By TS8, Topic 1 became the most weighted topic once again. In TS9, Topic 2 (MAP3DPNC) became the most weighted topic. We found that Topic 6 was the main topic during both TS1 and TS2, whereas the topic weight value of Topic 6 changed noticeably during these two segments. Similarly, Topic 1 was the main topic from TS3–TS6, as the topic weight value of Topic 1 changed significantly during these adjacent time segments. Despite the main topics remaining the same during TS1–TS2 and TS3–TS6, topic weights changed during these segments. Thus, whether the main topic changed could not be used as the criterion to assess the topic variation points.

All topics appearing in the pathway of 3DP were main topics, including Topic 1, Topic 2, Topic 6, and Topic 7. Among these topics, Topic 1 was the application of DLP. Topic 2 was an application technology. Topic 7 was a common manufacturing technology of 3DP. Topic 6 was the technology related to the 3DP device. Each of the four topics represented important subfields of 3DP, such as 3DP manufacturing technology, the application of the technology related to 3DP, the application technology, and the device related technology. Thus, the technologies corresponding to the four topics are important technologies in the 3DP field.

To show detailed changes in the main technological topics, the topic weight of the four topics was visualized by ThemeRiver. As shown in Fig. 4, the topic weight of the four corresponding topics to DLP, MAP3DPNC, SLA, and LM was visualized according to the topic weight values in Table 5. TS and topic weight information are provided in the Fig. 4. The red dots indicate topics, and the width of the topic indicates the topic weight. The changes in width from left to right represent the change in topic weight over time.

4.5. Discussion

The topic weight of DLP (digital light process) technology was relatively low in time segment 1 (TS1) and TS2. In TS3, the topic weight of the technology increased rapidly, peaking in the fifth segment, and then gradually decreasing. In TS3 to TS6, the topic weight of the technology was always the largest of the four topics. This indicated that DLP technology developed slowly in the early stage and rapidly in the medium term, and became mainstream in the 3D manufacturing field around 2001. Although its development momentum has slowed down in recent years, it still has a large share of attention and remains a research hotspot. Today, 3D printers using DLP are used widely.

The topic weight of the MAP3DPNC (methods and apparatus preventing 3D printer nozzle clogging) technology remained low from TS1 to TS8 but increased sharply during TS9 to attain the highest topic weight. This showed that the technology started developing later and was not noticed until recently. In the next few years, the technology may develop rapidly. Moreover, nozzle clogging is one of the most common hardware failures of 3D printers, so this technology's ability to prevent 3D printer nozzle clogging is practical and represents an advance in 3D printers. Printers are increasingly being adopted, so the benefits may transfer to related technologies.

The topic weight of the SLA (stereo lithography apparatus) declined slowly in TS1–TS9, whereas the development trend of SLA was steady. SLA was a mature technology in the era of rapid prototyping manufacturing, so it was the center of attention in 1990-2015. With the development of other technologies (e.g., DLP), attention has gradually shifted to newer alternatives.

The topic weight of LM (layered manufacturing) technology stayed at a steady level and fluctuated before 2010; it was at its largest value during the seventh stage. However, the advantage was not obvious compared to other technologies, and it dropped significantly during the latter two stages. This suggested that the concept of LM was proposed in early years and has not occupied a dominant position. Moreover, in recent years, it has followed a slight downward trend. The future development of LM is hard to predict.

Fig. 4 shows the topic weight changes of the four main topics of the 3D printing field in nine different TS. We found that methods and apparatus preventing 3D printer nozzle clogging technology developed slowly during the early stages, but accelerated in recent years. This indicated that the technology has promise and may become key in the 3DP field. Moreover, compared with the other three technologies, this technology was more closely related to the practical application of 3D printers, which indicated that the underlying technologies received more attention. With the wide use of 3D printers, other technologies related to the application of 3D printers may garner more attention. By analyzing the entire topic weight from an overall perspective, we observed that before 2010, there was a large discrepancy in the topic weight of the different topics. However, in recent years, the difference in topic weight has grown smaller. This phenomenon may be due to breakthroughs in a few technologies in the early days, which might explain why the topic weight of these technologies was significantly higher than other topics. As the number of technologies has become more fully developed, differences in the topic weight of the topics have decreased. The attention given to mature technologies in the early stages has also shifted to newer technologies over time.

5. Conclusions

This study presented a new approach for using topic variation detection to identify the technology evolution pathway by isolating its critical junctures. A case study of 3D printing was used to demonstrate its feasibility. We found that technologies that were mature and prominent in early stages experienced a downward trend over time. We also discovered that attention has shifted to newer technologies, and that more technologies have been fully developed. Meanwhile, technology related to the application of 3D printers has attracted more attention, so this avenue of research may be fruitful for further exploration. The method proposed in this paper meaningfully contributes to technology evolution pathway analysis, especially in detecting critical junctures and the evolution pathway. As shown through the case of 3D printing, the technology is also useful for scholars who are interested in 3D printing.

Further analysis of issues that were not covered in this study may be useful. For example, the critical junctures of the technology evolution pathway have been seldom researched, so the results lack a comparative component, and the accuracy was hard to assess as a result. In addition, the critical junctures were detected from a technology-driven perspective, so factors such as product and market demand were not considered in our framework. More factors affecting technology development can be taken into consideration in future research to analyze the technology evolution pathway.

CRediT authorship contribution statement

Zhongzhen Miao: Conceptualization, Methodology, Writing - original draft, Supervision. **Junfei Du:** Data curation, Visualization, Investigation. **Fang Dong:** Data curation, Visualization, Investigation. **Yufei Liu:** Conceptualization, Methodology, Writing - original draft, Supervision, Writing - review & editing. **Xiaochuan Wang:** Software, Validation.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (91646102, L1724034, L1624045 and L1524015), the Chinese Ministry of Education Project of Humanities and Social Sciences (16JDGC011), the Chinese Academy of Engineering's China Knowledge Centre for the Engineering Sciences and Technology Project (CKCEST-2018-2-13), the UK-China Industry Academia Partnership Programme (UK-CIAPP260), the Volvo-Supported Green Economy and Sustainable Development at Tsinghua University (20153000181), and the Tsinghua Initiative Research Project (2016THZW).

Appendix A

Table A1
The top 20 ranked words of 20 topics.

Topic1	Topic2	Topic3	Topic4	Topic5	Topic6	Topic7
light	plate	model	manufacturing	body	portion	layer
image	provided	method	additive	main	build	surface
display	block	step	component	shell	object	substrate
source	side	mold	method	wall	support	forming
optical	upper	process	process	side	system	film
projection	connected	involves	material	outer	apparatus	structure
projector	base	data	components	structure	comprising	method
digital	bottom	dimensional	manufactured	cavity	surface	resin
color	fixing	point	producing	provided	fluid	liquid
lens	frame	mould	structure	cover	assembly	formed
processing	lower	scanning	involves	cylinder	configured	pattern
beam	fixed	manufacturing	production	core	energy	coating
screen	supporting	processing	element	formed	view	curing
lamp	connecting	based	workpiece	surface	shows	sheet
apparatus	groove	design	tool	hole	plurality	glass
system	mounting	dimension	building	ring	claim	form
mirror	printer	software	included	opening	layers	layers

(continued on next page)

Table A1 (*continued*)

Topic1	Topic2	Topic3	Topic4	Topic5	Topic6	Topic7
crystal	hole	obtaining	shows	hollow	element	transparent
signal	sliding	obtain	independent	sealing	member	comprises
video	left	structure	rapid	front	form	area

Topic8	Topic9	Topic10	Topic11	Topic12	Topic13	Topic14
device	motor	polymer	agent	part	printing	nozzle
control	shaft	composition	parts	section	material	heating
system	connected	resin	material	slice	dimensional	pipe
module	wire	comprises	acid	shape	printer	connected
connected	rotating	preferred	obtain	multiple	printed	heat
sensor	driving	comprising	adding	screw	print	chamber
controller	wheel	group	mixing	shaped	drawing	device
temperature	screw	acid	solution	gear	modeling	feeding
printer	feeding	compound	temperature	type	product	cooling
processing	machine	preferably	water	piece	extrusion	tube
signal	transmission	particles	powder	hole	fused	head
controlling	mechanism	thermoplastic	stirring	shaft	deposition	provided
position	roller	components	drying	welding	shows	tank
comprises	gear	monomer	composite	multi	color	channel
output	belt	weight	mixture	cutting	comprises	outlet
detecting	provided	component	comprises	mesh	multi	printer
input	printer	solvent	step	groove	view	water
time	fixed	acrylate	nano	edge	materials	inlet
machine	bearing	curable	weight	formed	based	port
interface	drive	selected	preferred	metal	solid	liquid

Topic15	Topic16	Topic17	Topic18	Topic19	Topic20
data	bone	unit	powder	fiber	print
object	cell	power	metal	molding	head
information	comprises	circuit	laser	filament	platform
dimensional	tissue	supply	ceramic	product	axis
system	cells	electrode	alloy	container	mechanism
method	structure	conductive	forming	composite	direction
based	implant	high	sintering	plastic	printer
user	method	magnetic	melting	steel	guide
image	porous	element	high	comprises	moving
computer	biological	current	temperature	water	rail
printer	medical	electric	sand	paper	frame
apparatus	comprising	board	aluminum	injection	table
generating	patient	field	titanium	molded	device
file	matrix	integrated	comprises	fabric	movement
shows	artificial	voltage	method	made	vertical
drawing	hydrogel	pressure	selective	food	lifting
model	scaffold	electronic	particles	reinforced	working
processor	solution	coil	vacuum	concrete	machine
memory	vessel	electrical	copper	reinforcing	horizontal
included	growth	discharge	process	method	slide

Appendix B

1. The structure and generative process of LDA

LDA contains three Bayesian structural layers: topics, words, and documents. It represents documents as random mixtures of multiple latent topics, where each topic is a distribution of words (Blei et al., 2003).

The illustration of the LDA model is shown in Fig. B1. The variable K is the number of topics, M is the number of patent documents, and Nm is the total number of words in patent m. Moreover, $z_{m,n}$ denotes the topic assignment of the nth word in the m_{th} patent, and $w_{m,n}$ denotes the nth word in the m_{th} patent. In addition, $\vec{\vartheta}_m$ indicates the distribution across topics for patent m, and $\vec{\varphi}_k$ indicates the distribution over the vocabularies of topic k. The hyperparameters α and β were used to determine the amount of smoothing applied to the topic distributions for each patent document and the word distributions for each topic. The shadow circle was a visible node, while the others were latent nodes. Only the words in the corpus were observable (Blei, 2012; Blei, Ng, & Jordan, 2003; Chen, Zhang, Zhu, & Lu, 2017, a; Heinrich, 2005; Steyvers & Griffiths, 2007).

In the LDA topic model process, words in a patent document were derived using the probability distribution of topics generated by the probability distribution of the document. In a generated patent document, the probability of the occurrence of each word was calculated using the following equation (Blei et al., 2010):

$$p(w_n \mid d_m) = \sum_{k=1}^{K} p(w_n \mid z_k)^{*}p(z_k \mid d_m), \tag{B.1}$$

where w_n is the nth word; z_k is the kth topic; d_m is the m_{th} patent document; $p(w_n|z_k)$ indicates the probability of word w_n in topic z_k; $p(z_k|d_m)$ is the probability of topic z_k in patent document d_m; $p(w_n|z_k)$ is calculated by $\vec{\varphi}_k$; and $p(z_k|d_m)$ is calculated by $\vec{\vartheta}_m$.

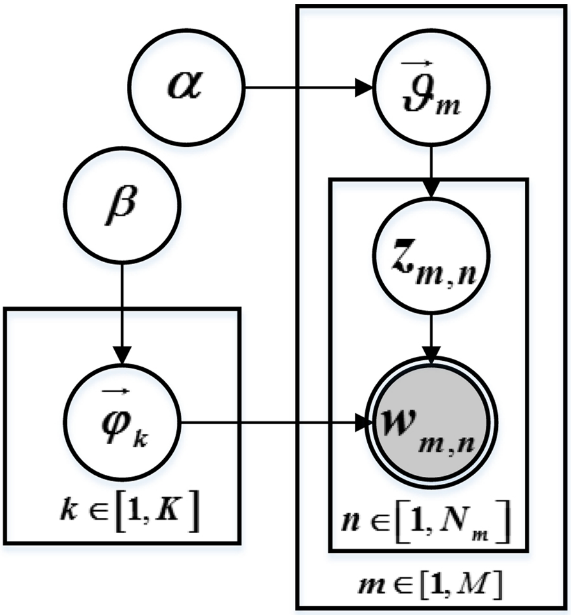

Fig. B1. Graphical model representation of LDA.

2. *The perplexity index of LDA*

In this paper, the number of topics, K, was determined using the perplexity index, which indicates the degree of uncertainty of a document belonging to a topic and can be used to evaluate LDA topic models (Mazin & Junlan, 2009). In order to select the best K value, for all textual documents, different topic numbers were selected to train the LDA topic model, and the perplexity was calculated individually. The number of topics corresponding to the smaller perplexity was selected as the optimal topic number.

Perplexity was calculated using the following equation (Mazin & Junlan, 2009):

$$perplexity = \exp\left\{-\left(\sum_{m=1}^{M}\sum_{n=1}^{N_m}\log\left(\sum_{k=1}^{K}p(w_n|z_k)p(z_k|d_m)\right)\right)\Big/\left(\sum_{m=1}^{M}N_m\right)\right\}, \tag{B.2}$$

where M is the number of documents; N_m is the number of words in document m; K is the number of topics; $p(w_n|z_k)$ indicates the probability of word w_n in topic z_k; and $p(z_k|d_m)$ is the probability of topic z_k in patent document d_m.

In this paper, K was set to 1, 10, 20, 30, 40, 50, 60, 70, 80, 90, and 100, and the perplexity corresponding to each K was calculated. Finally, the number of topics corresponding to the smaller perplexity was selected as the optimal topic number.

The curve that the calculated perplexity changed over time is shown in Fig. 3. The horizontal axis represents the number of topics, and the vertical axis represents the perplexity value. When the number of topics reached 20, the curve tended to be smooth, and overfitting could occur. Thus, we set the number of topics to 20 (Fig. B2).

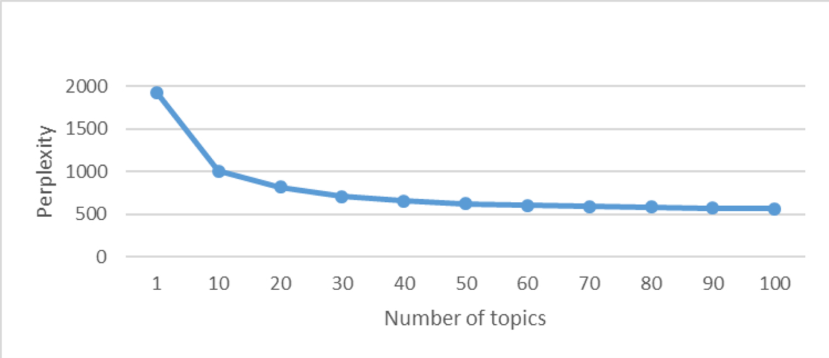

Fig. B2. Perplexity index of different number of topics.

References

Adomavicius, G., Bockstedt, J. C., Gupta, A., & Kauffman, R. J. (2007). Technology roles and paths of influence in an ecosystem model of technology evolution. *Information Technology and Management, 8*(2), 185–202.

Blei, D. M. (2012). Probabilistic topic models. *Communications of the ACM, 55*(4), 77.

Blei, D., Carin, L., & Dunson, D. (2010). Probabilistic topic models. *IEEE Signal Processing Magazine, 27*(6), 55–65.

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). *Latent dirichlet allocation: JMLR.oRg*.

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

Chen, H., Zhang, Y., Zhang, G., & Zhu, D. (2015). Modeling technological topic changes in patent claims. *Paper Presented at the Portland International Conference on Management of Engineering and Technology*.

Chen, S. H., Huang, M. H., & Chen, D. Z. (2013). Exploring technology evolution and transition characteristics of leading countries: A case of fuel cell field. *Advanced Engineering Informatics, 27*(3), 366–377.

Chen, H., Zhang, G., Zhu, D., & Lu, J. (2017a). Topic-based technological forecasting based on patent data: A case study of Australian patents from 2000 to 2014. *Technological Forecasting and Social Change, 119*, 39–52.

Chen, H., Zhang, G., Zhu, D., & Lu, J. (2017b). Topic-based technological forecasting based on patent data: A case study of Australian patents from 2000 to 2014. *Technological Forecasting and Social Change, 119*.

Cuneyt, G., Akcora, M., Bayir, A., Demirbas, M., & Ferhatosmanoglu, H. (2010). Identifying breakpoints in public opinion. *Paper Presented at the Workshop on Social Media Analytics*.

Ganguli, P. (2004). Patents and patent information in 1979 and 2004: A perspective from India. *World Patent Information, 26*(1), 61–62.

Havre, S., Hetzler, E., Whitney, P., & Nowell, L. (2002). ThemeRiver: Visualizing thematic changes in large document collections. *IEEE Transactions on Visualization and Computer Graphics, 8*(1), 9–20.

Heinrich, G. http://www.arbylon.net/publications/text-est.pdf.2005.

Hu, P., Huang, M., Xu, P., Li, W., Usadi, A. K., & Zhu, X. (2011). Generating breakpoint-based timeline overview for news topic retrospection. *Global Change Biology Bioenergy, 7*(4), 260–269.

Huang, Y., Zhang, Y., Ma, J., Porter, A., & Wang, X. (2015). *Tracing technology evolution pathways by combining tech mining and patent citation analysis*.

Huang, Y., Zhu, F., Guo, Y., Porter, A. L., Zhang, Y., & Zhu, D. (2017a). Exploring technology evolution pathways to facilitate technology management: A study of dye-sensitized solar cells (DSSCs). *Paper Presented at the Portland International Conference on Management of Engineering and Technology*.

Huang, Y., Zhu, F., Guo, Y., Porter, A. L., Zhang, Y., & Zhu, D. (2017b). Exploring technology evolution pathways to facilitate technology management: A study of dye-sensitized solar cells (DSSCs). *Paper Presented at the Portland International Conference on Management of Engineering & Technology*.

Huenteler, J., Ossenbrink, J., Schmidt, T. S., & Hoffmann, V. H. (2016). How a product's design hierarchy shapes the evolution of technological knowledge—Evidence from patent-citation networks in wind power. *Research Policy, 45*(6), 1195–1217.

Iansiti, M. (1995). Technology integration: Managing technological evolution in a complex environment. *Research Policy, 24*(4), 521–542.

Jiang, R., Kleer, R., & Piller, F. T. (2017). Predicting the future of additive manufacturing: A Delphi study on economic and societal implications of 3D printing for 2030. *Technological Forecasting and Social Change, 117*, 84–97.

Kim, D., & Oh, A. (2011). Topic chains for understanding a News Corpus. *Paper Presented at the International Conference on Computational Linguistics and Intelligent Text Processing*.

Kim, Y. G., Suh, J. H., & Park, S. C. (2008). Visualization of patent analysis for emerging technology. *Expert Systems with Applications, 34*(3), 1804–1812.

Kong, D., Zhou, Y., Liu, Y., & Xue, L. (2017). Using the data mining method to assess the innovation gap: A case of industrial robotics in a catching-up country. *Technological Forecasting and Social Change, 119*, 80–97.

Kropsu-Vehkapera, H., Haapasalo, H., & Rusanen, J.-P. (2009). Analysis of technology management functions in finnish high tech companies. *The Open Management Journal, 2*(1), 1–10.

Liu, Y., Zhou, Y., Liu, X., Dong, F., Wang, C., & Wang, Z. (2019). Wasserstein GAN-Based small-sample augmentation for new-generation artificial intelligence: A case study of cancer-staging data in biology. *Engineering, 5*(1), 156–163.

Ma, J., & Porter, A. L. (2015). Analyzing patent topical information to identify technology pathways and potential opportunities. *Scientometrics, 102*(1), 811–827.

Mazin, G., & Junlan, F. (2009). *Speech and language processing*.

Petrick, I. J., & Simpson, T. W. (2013). 3D printing disrupts manufacturing: How economies of one create new rules of competition. *Research-Technology Management, 56*(6), 12–16.

Steyvers, M., & Griffiths, T. (2007). Probabilistic Topic Models. In T. Landauer, S. D. McNamara, & W. Kintsch (Eds.). *Latent Semantic Analysis: A Road to Meaning.* Laurence Erlbaum.

Tseng, Y.-H., Lin, C.-J., & Lin, Y.-I. (2007). Text mining techniques for patent analysis. *Information Processing & Management, 43*(5), 1216–1247.

Verspagen, B. (2007). Mapping technological trajectories as patent citation networks: A study on the history of fuel cell research. *Advances in Complex Systems, 10*(1), 93–115.

Wang, C. C., Sung, H. Y., & Huang, M. H. (2016). Technological evolution seen from the USPC reclassifications. *Scientometrics, 107*(2), 537–553.

Wang, D., Shi, F., Dong, W., & Ge, Y. (2013). *Detecting opinion drift from chinese web comments based on sentiment distribution computing.* .

Wang, B., Liu, Y., Zhou, Y., & Wen, Z. (2018). Emerging nanogenerator technology in China: A review and forecast using integrating bibliometrics, patent analysis and technology roadmapping methods. *Nano Energy, 46*, 322–330.

Wang, Y., Urban, F., Zhou, Y., & Chen, L. (2018). Comparing the technology trajectories of solar PV and solar water heaters in China: Using a patent Lens. *Sustainability, 10*(11).
Wu, C. C. (2016). Constructing a weighted keyword-based patent network approach to identify technological trends and evolution in a field of green energy: A case of biofuels. *Quality & Quantity, 50*(1), 213–235.
Xie, Z., & Miyazaki, K. (2013). Evaluating the effectiveness of keyword search strategy for patent identification. *World Patent Information, 35*, 20–30.
Xing, W., & Croft, W. B. (2006). LDA-based document models for ad-hoc retrieval. *Paper Presented at the International ACM SIGIR Conference on Research and Development in Information Retrieval*.
Yan, R., Wan, X., Otterbacher, J., Kong, L., Li, X., & Zhang, Y. (2011). Evolutionary timeline summarization: A balanced optimization framework via iterative substitution. *Paper Presented at the International ACM SIGIR Conference on Research and Development in Information Retrieval*.
Yang, L., Qiu, M., Gottipati, S., Zhu, F., Jiang, J., Sun, H., et al. (2013). CQArank:Jointly model topics and expertise in community question answering. *Paper Presented at the ACM International Conference on Conference on Information & Knowledge Management*.
Yoon, J. (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. (2011). An automated method for identifying TRIZ evolution trends from patents. *Expert Systems with Applications, 38*(12), 15540–15548.
Yoon, B., & Park, Y. (2004). A text-mining-based patent network: Analytical tool for high-technology trend. *The Journal of High Technology Management Research, 15*(1), 37–50.
Yu, J., Meng, W., & Yu, C. (2011). Topic sentiment change analysis. *Machine Learning & Data Mining in Pattern Recognition, 6871*, 443–457.
Zhou, X., Zhang, Y., Porter, A. L., Guo, Y., & Zhu, D. (2014). A patent analysis method to trace technology evolutionary pathways. *Scientometrics, 100*(3), 705–721.
Zhou, Y., Li, X., Lema, R., & Urban, F. (2015). Comparing the knowledge bases of wind turbine firms in Asia and Europe: Patent trajectories, networks, and globalisation. *Science & Public Policy, 43*(4), 476–491.
Zhou, Y., Pan, M., & Urban, F. (2018). Comparing the international knowledge flow of China's wind and solar photovoltaic (PV) industries: Patent analysis and implications for sustainable development. *Sustainability, 10*(6), 1883.