The current issue and full text archive of this journal is available on Emerald Insight at:
https://www.emerald.com/insight/1741-038X.htm

Received 21 March 2020
Revised 3 July 2020
3 August 2020
Accepted 4 August 2020

A topic-based patent analytics approach for exploring technological trends in smart manufacturing

Juite Wang and Chih-Chi Hsu
Graduate Institute of Technology Management, National Chung Hsing University, Taichung, Taiwan

Abstract

Purpose – Smart manufacturing can lead to disruptive changes in production technologies and business models in the manufacturing industry. This paper aims to identify technological topics in smart manufacturing by using patent data, investigating technological trends and exploring potential opportunities.

Design/methodology/approach – The latent Dirichlet allocation (LDA) topic modeling technique was used to extract latent technological topics, and the generalized linear mixed model (GLMM) was used to analyze the relative emergence levels of the topics. Topic value and topic competitive analyses were developed to evaluate each topic's potential value and identify technological positions of competing firms, respectively.

Findings – A total of 14 topics were extracted from the collected patent data and several fast growth and high-value topics were identified, such as smart connection, cyber-physical systems (CPSs), manufacturing data analytics and powder bed fusion additive manufacturing. Several leading firms apply broad R&D emphasis across a variety of technological topics, while others focus on a few technological topics.

Practical implications – The developed methodology can help firms identify important technological topics in smart manufacturing for making their R&D investment decisions. Firms can select appropriate technology strategies depending on the topic's emergence position in the topic strategy matrix.

Originality/value – Previous research studies have not analyzed the maturity levels of technological topics. The topic-based patent analytics approach can complement previous studies. In addition, this study provides a multi-valuation framework for exploring technological opportunities, thus providing valuable information that supports a more robust understanding of the technology landscape of smart manufacturing.

Keywords Industry 4.0, Smart manufacturing, Technology strategy, Topic modeling, Patent analytics

Paper type Research paper

1. Introduction

Advances in information technologies and the growing demand for customized products with shorter product life cycles have driven the emergence of smart manufacturing in recent years (Brettel *et al.*, 2014; Kagermann *et al.*, 2013). Smart manufacturing causes not only disruptive changes in production technologies in the manufacturing industry but also creates enormous opportunities for new business models, products and services (Pereira and Romero, 2017; Stock and Seliger, 2016; Thoben *et al.*, 2017). Therefore, it is important for governments and industrial firms to analyze technological trends and explore potential technological opportunities at an early stage to formulate effective technology strategies.

Most firms identify technological opportunities based on expert judgment formed by scrutinizing and evaluating recent technological developments. However, experts may become less reliable when technical and market data are increasing rapidly and the technology life cycle

This research is supported by grant no. MOST 108–2221-E-005–078 from the Ministry of Science and Technology of the Republic of China (Taiwan).

Journal of Manufacturing Technology Management
Vol. 32 No. 1, 2021
pp. 110-135

1741-038X
DOI 10.1108/JMTM-03-2020-0106

is shrinking. With advances in information technology, computer-based approaches have been used to complement the expert-based approach (Porter and Cunningham, 2004).

Early research studies on the computer-based technological opportunity analysis (TOA) mainly analyzed structured bibliographical information in research publications or patent documents, which is easier to access and analyze for activities, patterns and gaps in technology development (Daim *et al.*, 2006; Ernst, 2003). However, unstructured patent information may contain more important and valuable information that is useful for mining technological opportunities. Therefore, numerous approaches applying the techniques of data and text mining and statistics have been developed for analyzing technological opportunities using research publications and patent documents (Abbas *et al.*, 2014; Bonino *et al.*, 2010; Madani and Weber, 2016).

This paper develops a patent analytics approach based on the latent Dirichlet allocation (LDA) for identifying emerging technological topics from patent data to explore potential technological opportunities in smart manufacturing. The LDA topic modeling technique (Blei *et al.*, 2003), an unsupervised machine learning approach, is applied to extract latent technological topics from the collected patent corpus. Although several studies have used the LDA with linear regression or time series techniques to forecast the trend of technological topics (Chen *et al.*, 2017; Suominen *et al.*, 2017), their approaches assumed that various topics are mutually independent. This assumption is not valid in the LDA topic analysis because the LDA inherently assumes that each document is represented as a mixture of various latent topics and each topic has some degree of correlation with other topics (Bolker *et al.*, 2009; Gibbons *et al.*, 2010). Therefore, this research applies the generalized linear mixed model (GLMM) (McCulloch *et al.*, 2008) as an alternate for analyzing the relative emergence levels of various technological topics, which is useful in the topic trend analysis in an emerging technological field such as smart manufacturing. A topic value analysis is developed based on the three patent indicators to assess the potential impact of each topic, while a topic competitive analysis is also used to identify technological positions of competing firms in each topic. Our research's results can help firms understand various latent topics in smart manufacturing, their relative emergence levels and their impacts and competitive landscapes in developing technology strategies for smart manufacturing.

The paper is organized as follows: Section 2 reviews the literature of the topic-based patent analytics approach and smart manufacturing. Section 3 presents the developed LDA-based patent analytics approach. The research study's results are presented in Section 4 and discussed in Section 5. Section 6 concludes the paper.

2. Literature review

2.1 Literature review of smart manufacturing

Smart manufacturing is described as a production-oriented cyber-physical system (CPS) (Lee *et al.*, 2015a; Monostori *et al.*, 2016) that integrates production systems, warehouses, logistics and even environmental and social requirements to create the digitization of the automated manufacturing environment (Kagermann *et al.*, 2013; Oesterreich and Teuteberg, 2016). The ultimate goal is to establish the global value creation networks, allowing efficiency and productivity improvements among firms across the whole value chains. Its enabling technologies include CPSs, the Internet of Things (IoT), big data, machine learning, cloud computing, autonomous robotics, additive manufacturing, embedded and agent-based systems and virtual/augmented reality (Ahuett-Garza and Kurfess, 2018; Chen *et al.*, 2018; Vaidya *et al.*, 2018).

Smart manufacturing has attracted attention in the recent literature (Kamble *et al.*, 2018). Since smart manufacturing research is still in its early stages, most studies have used conceptual approaches (Kamble *et al.*, 2018; Kusiak, 2018), literature reviews (Alcácer and

Cruz-Machado, 2019; Oztemel and Gursev, 2020; Wagire *et al.*, 2020), case studies (Muller *et al.*, 2018; Thoben *et al.*, 2017), expert interviews (Lu and Weng, 2018), industrial surveys (Zheng *et al.*, 2019) or experiments (Wang *et al.*, 2016) for discussing and testing the concepts, theories and applications. Among these, the case study approach has been widely used by researchers. To the best of our knowledge, only a handful of papers have used patent data for analyzing the technology landscape in smart manufacturing (Trappey *et al.*, 2016). Despite its contributions, previous research has not analyzed the trends or maturity level of smart manufacturing technologies, which is important in making technology investment decisions (Porter *et al.*, 2011). To address this research gap, this research develops the LDA topic modeling approach for identifying latent technological topics from patent data and applies the GLMM statistical model for analyzing relative emergence levels across a range of technological topics in smart manufacturing.

2.2 Literature review of the topic-based patent analytics approach

Patent documents, which are published by a government office to describe an invention providing a solution to a specific problem in the technological domain, are key intellectual property, comprising reliable technical information that reflects advances in technological development (Ernst, 2003). A patent document consists of two kinds of technical information: (1) structured information such as patent number, publication date, inventor and classifications and (2) unstructured information, such as the abstract, claims and description. Analyzing patent data can help R&D engineers, firms and organizations reveal important technical details, monitor and forecast technology trends, understand technology landscapes, stimulate novel industrial solutions and even make vital technological investment decisions.

Traditional patent analysis literature has primarily used the bibliometric analysis for analyzing structured bibliographical information and citations of patents to understand activities, trends and gaps in technologies (Daim *et al.*, 2006). Advances in data science, however, have enabled researchers to apply text mining techniques that extract important textual features from technical documents (e.g. publications and patents) for the TOA (Ma *et al.*, 2013), emerging topic detection (Porter *et al.*, 2019; Rotolo *et al.*, 2015; Small *et al.*, 2014), technology trend (Chen *et al.*, 2017) and technology assessment and impact analyses (Choi *et al.*, 2007), technology roadmapping (Bildosola *et al.*, 2017) and technology and innovation policy studies (Ma *et al.*, 2017). Please refer to Abbas *et al.* (2014), Bonino *et al.* (2010) and Madani and Weber (2016) for a detailed literature review.

Since decision makers usually deal with hundreds or thousands of patent documents, finding the important topic structure that is hidden in the patent corpus is a cumbersome task. The technological topic analysis, which is used to categorize one technology field into various subfields, is an important task in the technology trend analysis, emerging technology, hot spots and intellectual structure identification. Previous research studies on patent analytics have used International Patent Classification (IPC) codes for technology segmentation. However, IPC codes cannot offer more detailed information about technical attributes or the knowledge described in patent documents, which can provide more useful information (Wang *et al.*, 2014). Another inherent limitation of the IPC for patent analysis is that the IPC's classification principle is functionally oriented rather than application oriented, which means that the IPC does not match well with industrial technological applications (Venugopalan and Rai, 2015). Moreover, since the IPC has a lengthy classification update cycle and cannot respond to the rapidity of current technological change, it is not the most effective approach for the technological topic analysis in the latest technological areas such as smart manufacturing.

To overcome some of the abovementioned limitations, several approaches based on the bag-of-words model have been used in identifying technological topics from a set of patent documents, where each document is represented as the bag of its words. The co-word analysis, a technique that analyzes the co-occurrences of keywords between various patent documents, has been applied to define technological topics and identify relationships and interactions between the topics as well as emerging technological trends (Engelsman and van Raan, 1994; Huang, 2017). In addition, based on the main idea that words with similar meaning will occur in similar documents, the LDA was developed to discover various latent topics in a set of documents in which each document can involve a set of topics that are assigned to it via the LDA. Since existing patent classification systems have certain limitations, an automatic patent classification approach based on the LDA was developed for identifying technological topics from patent data sets (Venugopalan and Rai, 2015; Wang *et al.*, 2014). The results indicate that LDA can effectively identify main technological areas and the topic probability distribution of a patent provides a reduced representation of the knowledge content in a patent. It has been applied in the patent competitive intelligence analysis (Wang *et al.*, 2014), patent classification (Venugopalan and Rai, 2015), technology forecasting (Chen *et al.*, 2017), identifying technological convergence (Lee *et al.*, 2015b) and organizational knowledge profiling (Hu *et al.*, 2014; Suominen *et al.*, 2017).

In sum, various studies in the patent mining literature have applied LDA to discover the hidden topic structure in the patent corpus and have shown that LDA can effectively classify main technological topics. The linear regression and time series techniques have been employed for the trend analysis of individual technological topics (Chen *et al.*, 2017; Suominen *et al.*, 2017). However, their trend analysis approaches assumed that various topics are mutually independent, which is not valid for LDA, where each topic is assumed to have some degree of correlation with other topics (Gibbons *et al.*, 2010). This research applies the GLMM (McCulloch *et al.*, 2008), which is a powerful class of statistical models, as an alternate approach for analyzing emerging technological topics to overcome the abovementionedt limitation. In addition, the developed methodology can analyze relative emergence levels of various technological topics, which is useful in a new technological domain, such as smart manufacturing.

3. Data and the methodology

3.1 Data

This research uses patent data related to smart manufacturing for analyzing underlying technological topics and determines their relative emergence levels, impacts and competitive landscapes. The United States Patent and Trademark Office (USPTO), European Patent Office (EPO) and World Intellectual Property Organization (WIPO) patent databases are used as the major source of patent documents. Since smart manufacturing is an integrated concept that involves a broad range of technological fields, this research applies the broad query strategy to search the title, abstract and claim to retrieve patent data that have a wide coverage of various technological categories in smart manufacturing. First, keywords related to core smart manufacturing concepts are used, such as “Industry 4.0,” “smart manufacturing,” “smart factory,” “Industrial Internet of Things,” “Industrial Internet,” “cyber-physical production” and “cyber-physical logistics.” Second, based on the advanced technologies used in smart manufacturing (Ahuett-Garza and Kurfess, 2018; Chen *et al.*, 2018; Vaidya *et al.*, 2018) and keywords suggested in previous studies (Trappey *et al.*, 2016), technology-related keywords in smart manufacturing, such as “cyber-physical system,” “artificial intelligence,” “machine learning,” “additive manufacturing,” “augmented reality,” “virtual reality,” “cloud computing,” “prognostics and health management,” “data science,” “Internet of things,” “embedded systems,” “sensor network” and “robotics,” are used. In addition, manufacturing-related keywords such as “manufacturing,” “production,” “machine” and “factory” are used to restrict the collected patent documents within the manufacturing industry. The patent publication period is limited from Jan 2006 to Dec 2018 because the first patent incorporating the smart manufacturing concept was issued in 2006.

Initially, 6,450 patent documents were collected. After carefully screening out duplicate patents that belonged to the same patent family and unrelated patents that had IPC codes irrelevant to the manufacturing field, 5,521 patent documents were used for the subsequent analysis.

3.2 The methodology

The proposed patent analytics approach (Figure 1) applies the LDA topic modeling technique (Blei *et al.*, 2003) to extract latent technological topics from smart manufacturing-related patent data. Then, the topic value analysis is used to estimate the potential value of each topic based on the three patent indicators, while the topic competitive analysis is performed to understand technological positions of various competing firms in a topic. Next, the GLMM (Bolker *et al.*, 2009) is used to evaluate the relative emergence level of each topic. Finally, the results of the previous steps can be used to assist firms in selecting promising technological topics for formulating technology strategies.

3.2.1 Step 1: patent document preprocessing. Unlike research papers published in academic journals, patent documents have no defined keywords. Text mining has been used in the literature of patent analytics to extract key text features from patent documents (Tseng *et al.*, 2007) because using experts for defining keywords is expensive or unavailable. Single words are often used, but they are often too general in meaning or ambiguous to represent a technological concept (Wang *et al.*, 2014). Since technical terminologies in patent documents are mostly noun phrases, e.g. “artificial intelligence,” “data mining” and “Internet of things” (Wang *et al.*, 2014), this research applies the noun phrase extraction technique in the area of natural language processing (NLP) (Manning *et al.*, 1999) to retrieve key text features from the title and abstract of a patent document in the collected patent corpus where the technical disclosure of a patent is usually summarized in these fields.

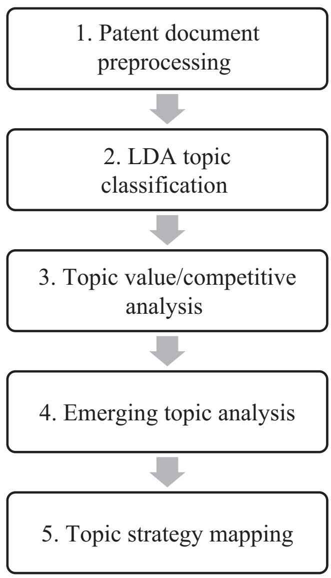

Figure 1.
The topic-based patent analytics methodology

The noun phrase extraction technique basically involves two steps: parts of speech (POS) tagging and defining noun-phase rules. POS tagging is used to label a word in a patent document as a particular part of speech such as noun, verb, article, adjective, pronoun or adverb. Noun-phrase extraction rules are then defined to extract noun phrases from patent documents:

(1) The length of a noun phrase: After conducting a pilot test, the length of the noun phrase is set to a minimum of 1 and a maximum of 3.

(2) The noun-phrase pattern: $(A|N)^*N(P+D^*(A|N)^*N)^*$, where N represents noun or proper noun, A represents adjective, P represents preposition, D represents determiner, "+" represents one or more than one time, "|" is an or operator and "*" represents 0 or more times.

The identified noun phrases represent a short summary of the information contained within a patent document. In addition, lemmatization, which removes inflectional endings only and returns the word to its base or dictionary form depending on the context, is jointly used with the POS tagging to improve overall accuracy (Straka *et al.*, 2016). The *R* package "UDPipe" for tokenization, tagging, lemmatization and dependency parsing is used for implementing noun-phrase extraction.

After the noun phrases are determined, several preprocessing methods are used. First, high-frequency words (e.g. invention, specification and claim) and general academic words (e.g. research, data and document) in the patent document are also removed. In addition, a synonym is replaced by a unique term or phrase. For example, the IoT, Industrial Internet of Things (IIoT) and "industrial Internet of Things" are replaced with "Internet of Things." Then, every phrase in the noun-phrase list is carefully examined to make sure that only meaningful phrases are used in subsequent analyses.

Finally, the document-phrase matrix is constructed from the collected patent corpus, where each patent document is characterized as a weighting vector of phrases encountered in that document (Yoon and Park, 2005). The term frequency–inverse document frequency (TF–IDF) (Berry and Kogan, 2010) is used to identify more essential, pervasive phrases in the collected patent corpus. TF–IDF is calculated as follows:

$$\mathrm{TF}-\mathrm{IDF}(t,d) = \mathrm{TF}(t,d) \times \mathrm{IDF}(t,D), \tag{1}$$

where the TF (t, d) is the number of occurrences of term t in document $d \in D$, while the IDF (t, D) measures the rareness of term t that occurs across the corpus D of documents. The IDF is calculated as follows:

$$\mathrm{IDF}(t,D) = \log \frac{N}{|\{d \in D: t \in d\}|} \tag{2}$$

where N is the total number of documents in D and the denominator is the number of documents containing term t. IDF would decrease the weight for commonly used terms while increasing the weight for terms that are rarely used in the set of patent documents. Hence, TF–IDF tends to remove terms or phrases that are common in the document set. This research keeps the phrases with TF–IDF value higher than the median value.

Note that although NLP is used to locate important smart manufacturing terms or phrases, noise data that are not related to technical terms in smart manufacturing can still exist and influence the validity of the results. Several iterations of the abovementioned processing methods have been performed for removing unnecessary data. Meaningful terms will be re-entered into the document-phrase matrix as required based on expert judgment.

3.2.2 Step 2: the LDA topic classification. This research applies the LDA, an unsupervised machine learning model, to identify underlying topic structures based on latent relationships of technological terms or phrases extracted from the collected patent corpus. The LDA assumes that each document is represented as a mixture of various latent topics, where each topic is characterized by a distribution over words. Given a corpus D consisting of M documents with each document $d \in D$ having N_d words, LDA models patent corpus D according to the following generative process:

(1) For $k = 1$ to K, where K is the total number of topics

Sample parameters for word distribution of each topic

$\phi(k) \sim$ Dirichlet (β)

(2) For $d = 1$ to D, where D is the number of documents

Sample parameters for topic distribution of each document

$\theta_d \sim$ Dirichlet (α)

For $w = 1$ to N_d, where N_d is the number of words in document $d \in D$

(1) Select the topic z_i for word w: $z_i \sim$ multinomial (θ_d)

(2) Select word based on topic z's word distribution: $w_i \sim$ multinomial $(\phi(z_i))$

In the above generative process, the words in the documents are the only observed variables, while latent variables (ϕ and θ_d) are determined in the inference process, given the values of hyper parameters (α, β and K). To derive latent variables based on a set of documents, the LDA inference process can be performed by the variational expectation-maximization (VEM) algorithm (Blei *et al.*, 2003) or Gibbs sampling (Griffiths and Steyvers, 2004) to improve the posterior distributions of both the topic representations of all the documents (θ_d) and word distributions of all the topics ($\phi(z_i)$) produced in the generation process. The results of the inference can identify the latent topic structure in a large collection of patent documents. Because of its computational efficiency, this research applies Gibbs sampling to estimate a sequence of dependent random variables that best approximates the true posterior. The Gibbs sampler is a Markov chain Monte Carlo (MCMC) algorithm for obtaining a sequence of observations which are approximated from a specified multivariate probability distribution when direct sampling is difficult. The document-topic matrix (θ) and the topic-phrase matrix (φ) are generated after the inference process, where the element (d, k) of the document-topic matrix represents the topic proportion for topic k in document d and the element (k, w) of the topic-phrase matrix represents the probability of a term or phrase related to topic k.

Another critical issue in the LDA topic analysis is to determine appropriate values for the hyperparameters (α, β and K) which should be selected in advance. When α (β) is large, each document (words) is likely to contain many various topics (Blei *et al.*, 2003). Given the number of topics K, the literature recommends $50/K$ and 0.1 for α and β, respectively (Griffiths and Steyvers, 2004). However, the appropriate value of K is not easy to determine. This research applies a grid search to find the appropriate values of α, β and K, where the perplexity analysis (Blei *et al.*, 2003), often used to evaluate the quality of the LDA generative model, is applied. Perplexity (Brown *et al.*, 1992) is a measure defined in information theory to assess how well the obtained LDA model can predict the test documents. A low perplexity indicates greater prediction of an obtained LDA model. The tenfold cross-validation method is used with the grid search.

3.2.3 Step 3: topic value and topic competitive analyses. The topic value analysis estimates the potential value of each technological topic identified in step 2. Each patent document is a mixture of a number of topics on the basis of the Dirichlet probability distribution, where the relationships between documents and topics are represented as probabilities that can be obtained from the document-topic matrix (θ). The potential value of each topic can be determined using the document-topic matrix and the three important patent indicators, forward citations, claim counts and patent family size, that are often used in the literature to assess a patent's quality or impacts.

(1) Topic citation rate: Forward citation of a patent is the number of citations the patent receives from later patent applications. When a patent is cited in many different patents, it indicates that it has technological and economic value for future technological development. This measure has been widely applied as an indicator in evaluating the value of a patent (Reitzig, 2004). Further, since the number of forward citations is affected by publication year, this research adjusted this number by dividing the average number of forward citations for all patents from the same patent year to ameliorate the problem of more recent patents having fewer forward citations (Breitzman and Thomas, 2015). This research calculates the average forward citation count for each topic. The average forward citation rate of topic i is calculated as follows:

$$tfc_i = \sum_{i=1}^{n} \theta_{ji} \cdot c_j \tag{3}$$

where θ_{ji} is the entry value of the document-topic matrix, that is, the strength between topic i and patent document j and c_j is the adjusted citation count of patent document j.

(2) Topic claim rate: The claims of a patent define the scope of the protection conferred by a patent. This has been regarded as a possible operationalization of a patent's "breadth." A larger number of independent claims imply higher economic value (Reitzig, 2004). Similarly, the claim rate for each topic is calculated as follows:

$$tc_i = \sum_{j=1}^{n} \theta_{ji} \cdot e_j \tag{4}$$

where e_j is the claim count of patent document j.

(3) Topic family size rate: Patent family refers to a collection of applications for patents that are filed in several countries, which are related to each other by one or several common priority filings. The value of a patent is related to the geographic scope of its patent protection and large international patent families have been found to be particularly valuable. Applicants might be willing to accept additional costs and delays of extending protection to other countries only if they deem it worthwhile (Harhoff *et al.*, 2003). This research calculates the topic family size rate which is as follows:

$$tfs_i = \sum_{j=1}^{n} \theta_{ji} \cdot h_j \tag{5}$$

where h_j is the family size of patent document j.

Then, the total topic value is the sum of the normalized values of the abovementioned three topic value indicators.

The topic competitive analysis is used to identify technological positions of competing firms for the selected technological topic to better understand R&D emphasis and the capabilities of those firms. Similar to topic value estimation, the relationship between topic i and firm k is calculated as follows:

$$tfm_{ik} = \sum_{j=1}^{n} \theta_{ji} \cdot df_{jk} \tag{6}$$

where df_{jk} is equal to 1 if firm k is the applicant of patent document j, otherwise it is equal to 0. According to the topic–firm matrix tfm, the topic–firm network can be drawn to show the relationship between topics and firms. The link thickness and the distance between a topic and a firm characterize the relational strength between them. The thicker a link and the closer a topic and a firm are, the greater the R&D emphasis the firm places on the topic. The topic–firm network may be useful to firms in understanding the competitive landscape of a technological topic and to explore potential collaboration opportunities if two firms have complementary research interests.

3.2.4 Step 4: the emerging topic analysis. The GLMM is used to analyze the relative emergence level of each topic based on the amount of published patent documents and the growth rate of each topic. Griffiths and Steyvers (2004) used linear regression for analyzing academic publications over time to assess which topics are hot (positive slope) versus cold (negative slope). However, the residuals or errors in linear regression models are assumed to be normally and independently distributed (Neter *et al.*, 1989). In addition, time series techniques have been used for estimating topic trends and their approaches also assumed that various topics are mutually independent (Chen *et al.*, 2017; Suominen *et al.*, 2017). The abovementioned assumption is not valid in the LDA topic analysis because the observed outcomes are correlated with residuals to some degree and there is data dependency across the topics (Gibbons *et al.*, 2010).

The GLMM (McCulloch *et al.*, 2008), a class of statistical models that considers both fixed and random predicator variables, is used in this research to analyze the number of patent documents per topic over time, when topics are not assumed to be completely independent. Mixed (fixed and random) effects models are useful when data have more than one source of random variability. For example, both within-topic and across-topic variabilities over time must be considered in the topic trend analysis. This approach has been popularly used in ecological and evolutionary studies in recent years (Bolker *et al.*, 2009). The expected number of patent documents y_i conditional on topic i over time is modeled by Poisson's GLMM:

$$\log(E(y_i)) = (\alpha + \omega_{i1}) + (\beta + \omega_{i2})t + e_i \tag{7}$$

where t is time (in year), α and β are the fixed intercept and slope, respectively, ω_{i1}, ω_{i2} are random intercept and slope, respectively, whose variances are estimated by the model and e_i is the error normally distributed with mean 0 and a constant variance. Thus, the total number of published patent documents increases or decreases approximately linearly as a function of years with fixed intercept (the mean number of patent documents) and slopes (the mean growth rate of patent documents). The number of patent documents assigned to each topic appears to vary in the slopes and intercepts, namely, random intercepts and slopes. Topics with positive random intercepts can be interpreted as having a higher than average number of published patents in the time period, while topics with positive random slopes can be interpreted as having a higher than average growth in published patents.

3.2.5 Step 5: topic strategy mapping. According to the random intercept and random slope of each topic, the identified topics can be positioned on a two-dimensional topic strategy matrix (Figure 2) to visualize the relative emergence levels of various technological topics, where the x-axis is the random intercept characterizing the relative amount of filed patents and the y-axis is the random slope characterizing the relative patent growth rate.

Quadrant I contains topics that have both the number of filed patents and patent growth rate higher than the average number and growth rate of published patents, respectively, where the topics located in this quadrant can be interpreted as competitive emerging topics (Berkowitz, 1993). Quadrant II comprises topics having patent growth rates higher than the average but a number of patents lower than the average, where each topic can be considered as a newly emerging topic. Conversely, topics with growth rates lower than the average are rather static and positioned in quadrants III and IV. Quadrant III comprises topics having a number of filed patents lower than the average, where the potential of each topic in this quadrant is uncertain and requires further analysis. Quadrant IV includes topics having the number of patents higher than the average, where topics in this quadrant can be viewed as more developed topics than topics in other quadrants. The topic strategy matrix can be used to help governments and firms in selecting promising technological topics and formulating technology strategies for smart manufacturing.

4. The data analysis and results

4.1 Major characteristics of the collected patent corpus

First, the major characteristics of the 5,521 patent documents collected in this study are discussed. As shown in Figure 3, the annual number of patents issued in smart manufacturing slowly increased from 2006 to 2013 and grew rapidly after 2014. The six leading countries were the USA (2,285), Germany (392), Japan (223), China (145), the UK (130) and France (107). US entities filed more patents than all other countries combined. The top-five applicants were General Electric (242), HP development (168), Siemens (146), Stratasys

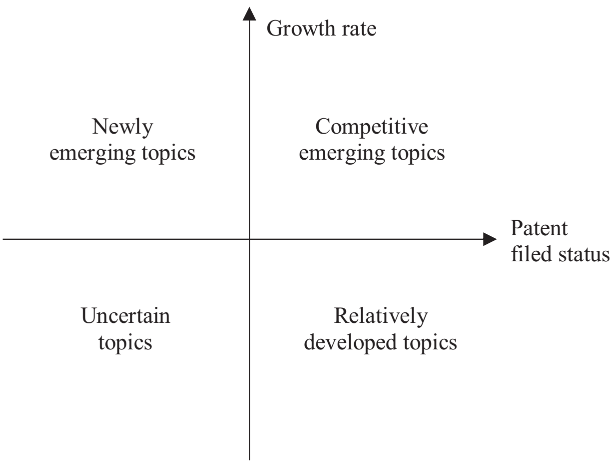

Figure 2. The topic strategy matrix

(125) and United Technologies (98). Most of the technological leaders were multinational companies that researched, developed and manufactured products in numerous areas, including industrial automation, aerospace and electronics. In addition, the top-ten IPC codes were B29C67 (1,165), B22F3 (864), B29C64 (804), B33Y10 (441), G05B19 (378), G06F17 (352), B33Y30 (322), H04L29 (261), B33Y50 (175) and G06F19 (153). Most IPC codes were related to additive manufacturing, such as techniques for shaping of material in a plastic state (B29C67), additive deposition manufacturing techniques, additive agglomeration or additive layering techniques (B29C64), processes of additive manufacturing (B33Y10), apparatus of additive manufacturing (B33Y30) and data processing for additive manufacturing (B33Y50). G05B19 includes inventions related to industrial control systems used to operate and/or automate industrial processes efficiently according to data from remote sensors. Since modern manufacturing processes produce an enormous amount of data, G06F17 and G06F19 are inventions of methods for processing and analyzing data for specific functions such as process improvement, fault detection and classification, reduction in waste and more informed decision-making. H04L29 involves inventions related to arrangements, apparatus, circuits or systems for manufacturing data communication and networking for real-time access, holistic security and totally integrated automation.

After patent document preprocessing, 4,194 patent documents with 5,074 phrases were finally selected and used for subsequent analyses. Statistical software *R* was used for the data analysis in this study.

4.2 The latent Dirichlet allocation topic classification

The hyperparameters α, β and K of each variable were chosen using a grid search, where the search space is defined as $\alpha \in [0.05, 0.1, 0.5, 1, 5, 10]$, $\beta \in [0.05, 0.1, 0.5, 1, 5, 10]$ and $K \in [10, 30]$. Tenfold cross validation, which is often used in the supervised learning, was used to evaluate to the perplexity performance of an obtained LDA model. The training data set was used to generate an LDA model and the test data set was used to evaluate the perplexity of the LDA model. We then narrowed down the search space to $\alpha \in [0.1, 1.0]$, $\beta \in [0.05, 0.15]$ and $K \in [10, 30]$. The best value was found (Figure 4) where $\alpha = 0.5$, $\beta = 0.12$ and $K = 14$.

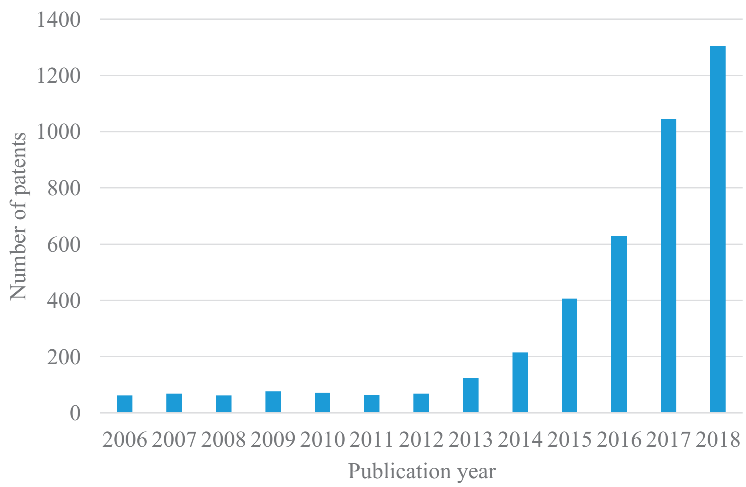

Figure 3.
Growth trend of published patents in smart manufacturing

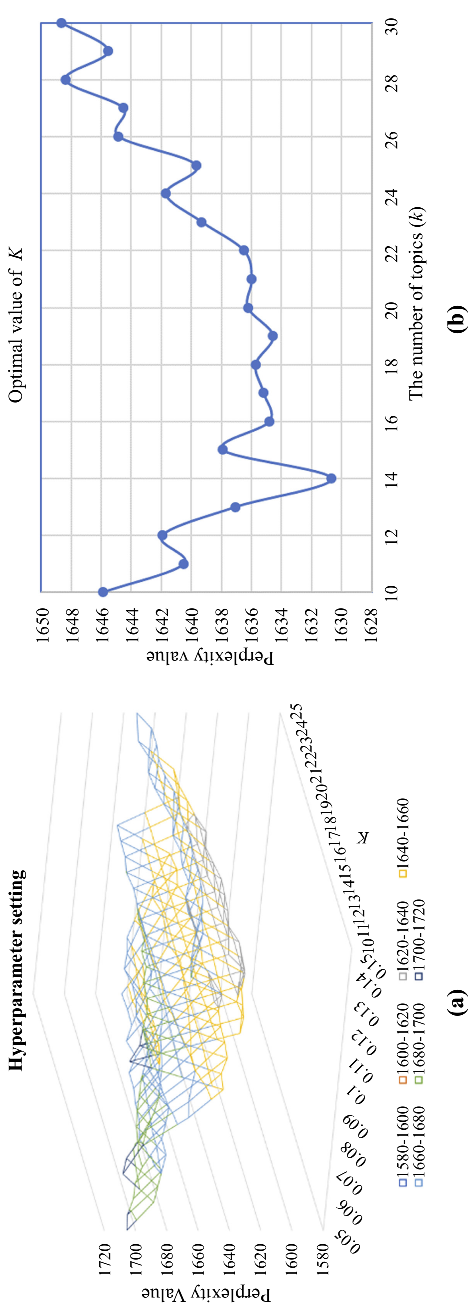

Figure 4.
Computational results of hyperparameter setting for the latent Dirichlet allocation

The topic-phrase matrix (φ) generated from LDA is shown in Table A1, where ten high-ranked phrases are ordered according to their probabilities related to each topic. Each topic was labeled based on the high-ranked phrases and patent documents that were highly correlated with the topic from the document-topic matrix (θ). A brief description of each topic and a few highly correlated patent documents are listed in Table A2. Also, two smart manufacturing experts with more than five years of industry experience in manufacturing were invited to examine and refine the labeling results.

To further explain the 14 topics, we classified these topics into six categories as shown in Table 1 based on the main purposes of smart manufacturing technologies defined by Frank *et al.* (2019): vertical integration, virtualization, automation, traceability and flexibility. A manufacturing firm's vertical integration begins with data collection from thousands of sensors distributed throughout various physical production assets (e.g. machines, storage systems and parts) on the shop floor. The data gathered are then transmitted via the industrial network and control system (e.g. supervisory control and data acquisition [SCADA]) for production control to monitor and improve manufacturing performance on the shop floor. The manufacturing execution system (MES) receives and processes data for delivering the production status to the enterprise resource planning system (ERP). The production orders may also be deployed from the ERP to the MES and production tasks may be assigned to individual machines through the industrial network and control system. Therefore, vertical integration can include sensing technologies for accurate and reliable measurement data acquisition and processing by various types of equipment and processes (topics 1 and 3). Industrial network management and control technologies connect devices, machines, equipment and information systems with improved speed, interoperability, reliability and security (topics 1, 4 and 10) while monitoring and controlling those physical industrial assets and processes to ensure production quality (topics 4 and 8). In addition, cloud computing technologies can be used to realize the manufacturing-as-a-service concept that optimizes manufacturing resource sharing to achieve coherence and economies of the scale (topic 9) (Tao and Qi, 2017).

Virtualization aims to create a digital twin which is a real-time virtual representation of physical production assets, processes and systems that can help optimize manufacturing performance (Lee *et al.*, 2015a). The digital twin captures and aggregates operational and environmental data collected from sensors distributed throughout manufacturing processes on the shop floor to understand the current behavioral status of machines, work in progress and environmental conditions within the factory (topics 4 and 9). Analytics techniques can be used to analyze the collected data for detecting problems earlier, predict outcomes with a higher degree of accuracy and monitor and improve operational performance (topics 5 and 6).

Categories	Technological topics
Vertical integration	Topic 1 (smart connection), topic 3 (smart sensing), topic 4 (CPSs), topic 8 (remote monitoring and control), topic 9 (cloud-based services) and topic 10 (industrial Internet network management)
Virtualization	Topic 4 (CPSs), topic 5 (fault detection and classification), topic 6 (manufacturing data analytics) and topic 9 (cloud-based services)
Automation	Topic 5 (fault detection and classification), topic 7 (industrial automation and robotics) and topic 8 (remote monitoring and control)
Traceability	Topic 1 (smart connection) and topic 2 (production traceability)
Flexibility	Topic 11 (powder bed fusion), topic 12 (material extrusion), topic13 (additive manufacturing for products and apparatus) and topic 14 (additive manufacturing for tools and mold making)

Table 1.
Categories of 14 technological topics

Smart manufacturing technologies can also promote factory automation to perform routine tasks and minimize human involvement for cost reduction while improving quality, reliability and production yield. Smart robots can replace humans in simple and structured tasks within closed areas while collaborating with humans for more complicated tasks in open spaces using smart sensor human–machine interfaces (topic 7). In addition, advanced analytical tools can be used to reduce and avoid machine downtime from unexpected disruptions by analyzing data collected from sensors, detecting and even predicting machine defects beforehand via remote monitoring and control of industrial devices, machines and processes (topics 5 and 8).

With rising quality standards and greater product liability, traceability is becoming increasingly important in the manufacturing industry for tracking individual parts, sub-assemblies and products in real time throughout the manufacturing process, from raw materials entering the factory to final products being delivered to the customer. It can provide manufacturers with real-time visibility into their operations to facilitate root-cause analysis and continuous improvement of quality and efficiency (Yao *et al.*, 2018). In addition, using traceability technologies (e.g. radio-frequency identification [RFID] and beacons) can help parts, products, machines and workers operate and interact with each other, thus enabling a more flexible and efficient manufacturing of customized products (topics 1 and 2).

Additive manufacturing also can improve manufacturing flexibility, allowing manufacturers to respond to changing requirements in uncertain and competitive global markets (Yao *et al.*, 2018). In contrast to conventional manufacturing approaches that remove material from a stock item, additive manufacturing produces a three-dimensional-shaped structure by successively adding material layer by layer from a computer-aided design model. It can enhance flexibility by (1) manufacturing products “on-demand” for low volume production in response to customer orders, (2) producing parts with complex geometries, (3) fabricating parts without tooling, (4) exploiting process variables for efficient production and (5) producing tools or molds for subtractive processes (Yao *et al.*, 2018). Several additive manufacturing methods have been developed, such as powder bed fusion and material extrusion (topics 11 and 12). Although additive manufacturing technologies have been developed for more than three decades, the industrial applications are still relatively in its infancy. Numerous inventions have been patented in electronics, aerospace, automotive and medical industries (topics 13 and 14).

4.3 Topic value and topic competitive analyses

Table 2 shows the results of the topic value analysis based on the three topic indicators as shown in Eqs. (3)–(5). The top-six technological topics that have topic values greater than the

Topic number	Topic label	Value	Topic number	Topic label	Value
1	Smart connection	2.568	8	Remote monitoring and control	0.531
2	Production traceability	0.575	9	Cloud-based services	0.789
3	Smart sensing	1.546	10	Industrial Internet network management	0.059
4	Cyber-physical systems	1.043	11	Power bed fusion	1.038
5	Fault detection and classification	0.531	12	Material extrusion	0.753
6	Manufacturing data analytics	1.577	13	AM products and apparatus	1.039
7	Manufacturing automation and robotics	1.242	14	AM for tool/mold making	0.891

Table 2.
The topic value analysis

average (1.164) include, in descending order, the following: "smart connection" (topic 1), followed by "manufacturing data analytics" (topic 5), "smart sensing" (topic 3), "manufacturing automation and robotics" (topic 7), "cyber-physical systems" (topic 4) and "power-bed fusion additive manufacturing" (topic 9).

Figure 5 shows the topic–firm network which depicts the relationships between topics and leading firms, where the thickness of a link represents the strength of the relationship. We found that GE and Siemens, located in the central position of the network, were the key players in smart manufacturing. GE owned the largest number of smart manufacturing patents and was a technological leader for almost all topics. Their R&D emphasized several topics, including CPSs (topic 4), remote monitoring and control (topic 8) and cloud-based service technologies (topic 9), especially in additive manufacturing (topics 11, 12, 13 and 14). Similar to GE, Siemens had also invested in various smart manufacturing topics but strongly emphasized smart sensing technologies (topic 3), CPSs (topic 4) and manufacturing automation and robotics (topic 7), especially in additive manufacturing (topics 16, 17, 18 and 19).

Unlike GE and Siemens, which promoted broad R&D emphases across a variety of smart manufacturing technologies, HP, Stratasys, United Technologies Corporation (UTC) and Arcam applied an R&D-focused strategy on additive manufacturing (topics 16, 17, 18 and 19). Similarly, HP pioneered powder-bed fusion technologies (topic 11), while Stratasys was the leader in material extrusion (topic 12) and developed additive manufacturing systems for various industries such as automobile, aerospace and electronics (topic 13). Fanuc Corporation pursued machine learning and big data analytics techniques for total factory control and predictive maintenance of manufacturing equipment (topic 6), while Rohm developed smart sensing technologies for detecting environmental/physical conditions and next-generation sensor networks for the semiconductor industry (topic 3). Kyland invested in communication network devices and equipment for the industrial Internet (topic 10).

The remaining major players focused on a few technological topics. For example, Strong Force IoT emphasized its investment in smart connection (topic 1) and manufacturing automation and robotics (topic 7), while Taiwan Semiconductor Manufacturing Company (TSMC) developed fault detection and classification techniques (e.g. virtual metrology) for predicting wafer quality based on machine parameters and sensor data in the production

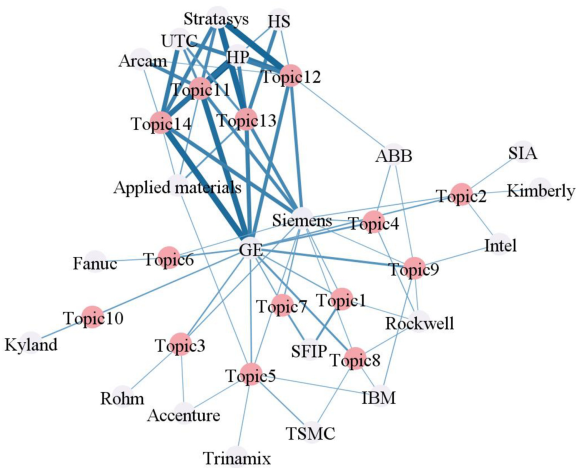

Figure 5.
The topic–firm network

equipment for semiconductor manufacturing (topics 5 and 8). Rockwell Automation Technologies worked on smart connection (topic 1), CPSs (topic 4), remote monitoring and control (topic 8) and cloud-based service technologies (topic 9). International Business Machines (IBM) Corporation also pursued fault detection and classification (topic 5) and cloud-based service technologies (topic 9).

4.4 Emerging topic analysis and topic strategy mapping

The GLMM approach was applied to analyze the relative emergence levels of the 14 topics. The lme4 package for *R*, which provides functions to fit and analyze GLMMs, was used for the statistical analysis. Statistically significant results were obtained (intercept: standard error = 0.06002, z-value = 34.90, p-value < 0.001; slope: standard error = 0.06792, z-value = 10.85, p-value < 0.001), where the 14 topics were positioned on the topic strategy matrix, as shown in Figure 6. A detailed analysis is presented in Section 4.4.

5. Discussion

As shown in the topic strategy matrix, topic 2 (production traceability technologies), topic 6 (manufacturing data analytics) and topic 12 (material extrusion technologies) are located in quadrant I, crowded with both filed patents and higher growth rates than the other topics. This indicates that these topics were competitive technological areas in the industry and that it was hard to avoid the intellectual properties of others, especially for technological leaders in these topical areas such as GE, Siemens, HP, Stratasys and Fanuc. However, because of the high growth rate, there may still be opportunities for firms to develop their own novel or improved inventions. Among these three topics, topic 6 had a higher value than other topics.

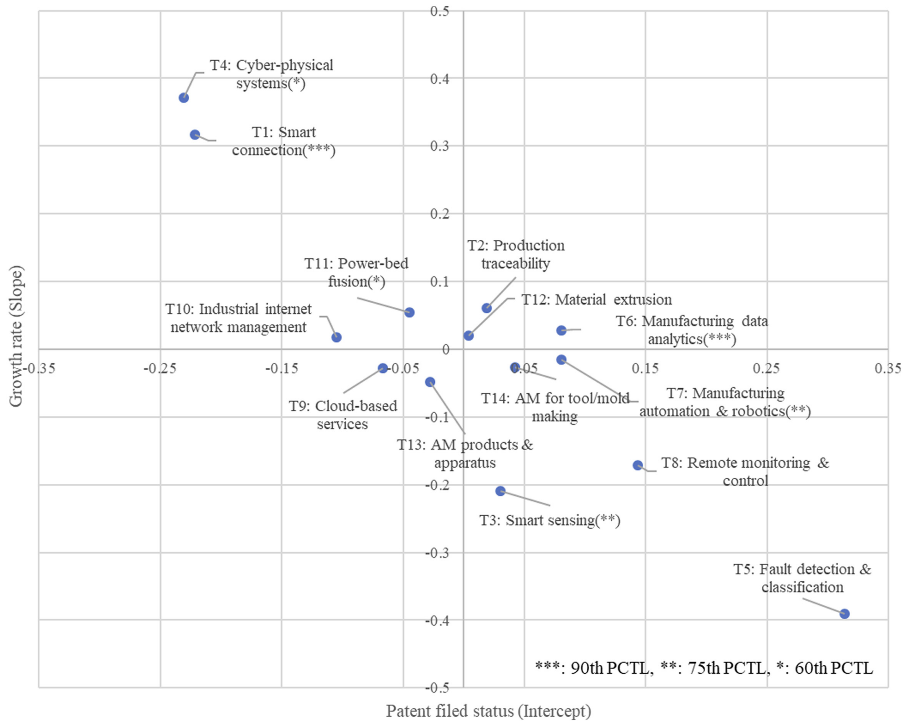

Figure 6.
The topic strategy matrix for smart manufacturing

A licensing or cross-licensing strategy might also be used to avoid infringement litigation from other firms while reducing R&D investment costs and development lead time for earlier market entry time.

Topics in quadrant II have few patents filed but with high growth rates and possess great investment opportunities. These included topic 1 (smart connection), topic 4 (CPSs), topic 10 (industrial Internet network management) and topic 11 (powder bed fusion additive manufacturing), where topics 1, 4 and 11 had higher values. The topics located in this quadrant were at the initial growth stage and might become profitable in the future if developed and protected properly. Therefore, firms intending to enter these fields must develop technologies superior to those of the competing firms, such as GE, Strong Force IoT Portfolio, Siemens and HP, the leading firms for these topics. Meanwhile, they must file patents that cover their technologies broadly and deeply to protect their technological assets in the long run.

Quadrant III involves topics with few patents filed and low growth rates and which may still be uncertain and more difficult to interpret and react to. Also, two topics are positioned in this quadrant: topic 9 (cloud-based services) and topic 13 (additive manufacturing products and apparatus). Compared to industries such as health care, finance, retail and education, cloud computing has been adopted less in the manufacturing industry. With the emerging trend of smart manufacturing, cloud computing is being increasingly adopted by manufacturing firms to virtualize manufacturing resources as on-demand services (Tao and Qi, 2017). Therefore, topic 9 (cloud-based services) has shown potential technological and market feasibility and has entered the growth stage. Yet, topic 13 (additive manufacturing products and apparatus) may face challenging technological barriers and great market uncertainty and may require a long period of development to take-off. Though additive manufacturing technologies have already demonstrated successful applications from rapid prototyping to manufacturing component parts, wide adoption of these technologies in the aerospace, electronics and biomedical industries still faces technical challenges such as limited materials, dimensional inaccuracy, inconsistent quality and high cost (Ngo *et al.*, 2018).

To fully explore potential technological opportunities in quadrant III, firms can select potential technological topics in quadrant III and apply the real options concept to develop flexible technology investment strategies for capturing opportunities while avoiding risks (Smit and Trigeorgis, 2007; Wang *et al.*, 2015). Firms can start with an initial investment in technology development projects in these topic fields. If the project experiences technical success, it creates options for substantially higher investments in the continuing project with relatively higher expected market value. If the project fails to achieve technical feasibility, then there is no need to further invest on the project, which limits downside risk.

Quadrant IV contains topic 3 (smart sensing technologies), topic 5 (fault detection and classification), topic 7 (manufacturing automation and robotics) and topic 8 (remote monitoring and control), crowded with patents filed but with lower growth rates than the average. This implies that these technological topics may have already entered the mature technology development stage. The topics in this quadrant are highly competitive but have only weak profitability. Firms mulling entrance into the topic fields in this quadrant should be aware that these technologies are unlikely to be the foundation of competitive advantages in the future. Based on their technological position, firms may either consider licensing technologies from firms or developing niche inventions if they need to enter these technological areas.

6. Conclusions, practical implications and limitations

This paper developed a topic-based patent analytics methodology to investigate technological trends and explore potential opportunities in smart manufacturing using patent data.

The LDA topic modeling technique was applied to extract latent technological topics. The topic value analysis was then developed based on the three patent indicators to evaluate each topic's potential value, while the topic competitive analysis was used to identify technological positions of competing firms in each topic. Then, the GLMM was used to evaluate the relative emergence levels of these topics and the topic strategy matrix was constructed to understand the relative emergence positions of various technological topics. The abovementioned analytical results are useful in shaping technology strategies for smart manufacturing.

6.1 Contributions

The academic contributions of this paper are as follows: first, to the best of our knowledge, only a handful of papers have used patent data for analyzing the technology landscape in smart manufacturing (Trappey *et al.*, 2016). Despite their contributions, previous research studies have not analyzed the trends or maturity levels of various technological topics, which is important in making technology investment decisions (Porter *et al.*, 2011). This research provides a topic-based patent analytics approach that can complement previous studies. Second, previous studies have applied the latent semantic analysis (LSA), a topic modeling approach based on the singular value decomposition, for analyzing academic publications in smart manufacturing (Wagire *et al.*, 2020). This study is the first to integrate the LDA with the GLMM to analyze patent data for identifying latent technological topics in smart manufacturing and visualizing their relative emergence levels in terms of the topic strategy matrix. Finally, this study provides a multi-valuation framework to explore technological opportunities from the emergence position, potential impact and technological positions of various competing firms for each topic, thus providing valuable information for comprehending the technology landscape of smart manufacturing.

The practical contribution of this paper is to help R&D managers identify major smart manufacturing technological topics from patent data and evaluate each topic from the three perspectives: relative emergence positions, value impacts and competitive landscape. A total of 14 topics were extracted. Among them, technological topics, such as smart connection, CPSs, manufacturing data analytics and power bed fusion additive manufacturing, were considered as highly emerging and highly valuable topics. Several major competing firms for each topic were recognized and their R&D strategies were also discussed. The research findings can help firms formulate their technology strategies in smart manufacturing.

6.2 Practical implications

We offer several practical implications for R&D managers. First, because of the higher technological and market uncertainties for emerging technologies, making technology investment decisions to acquire important technology assets is critical for technology-based firms to achieve long-term market success. The methodology developed herein can help firms identify various technological topics in smart manufacturing and understand the technology and competitive landscapes for making their R&D investment decisions. As discussed in Section 5, firms must select appropriate technology strategies based on the topic's emergence position on the topic strategy matrix. Firms may consider investing in topics in quadrants I and II since these have relatively fast growth potential and high-value impacts, such as smart connection technologies, CPSs, manufacturing data analytics and powder bed fusion additive manufacturing. In addition to internal R&D for acquiring required technology assets, the topic competitive analysis can assist R&D managers in understanding technological positions of various competing firms. They should avoid patent infringement and select appropriate technology acquisition strategies to obtain technological knowledge from other firms, such as licensing, joint R&D, mergers and acquisitions (M&A) and R&D alliance (Santamaría *et al.*, 2009).

In addition, special care should be taken when investing in the topics located in quadrants III and IV. R&D managers may consider applying the real-option logic with "wait-and-see" flexibility in their evaluation of the uncertain technological topics in quadrant III. Firms can defer investment until technology and market feasibility is clear. For topics in quadrant IV, firms should carefully assess the costs and benefits of their R&D investments and acquisition strategies on those topics because these technologies are unlikely to be the foundation of competitive advantages, but they may provide technological infrastructure or bases for integrating other technologies or advancing next-generation smart manufacturing technologies.

6.3 Limitations and future research directions

Although three major patent databases that cover most patent documents published globally have been used in this research, non-English patent documents that have been published but are not included in the three patent databases were not included in this study. Future research can consider using a multilingual NLP tool to process patent data for a more comprehensive topic analysis of smart manufacturing. In addition, most topic analysis methods, such as LDA, apply the vector space model (VSM) based on terms or keywords to represent the patent corpus, resulting in a lack of sentence-level semantics and information loss. Such problems can affect the accuracy of technological topic identification. Future research may apply advanced NLP techniques, such as Word2Vec, a word embedding technique based on deep learning, to replace the traditional VSM used in LDA (Zhang *et al.*, 2018).

References

Abbas, A., Zhang, L. and Khan, S.U. (2014), "A literature review on the state-of-the-art in patent analysis", *World Patent Information*, Vol. 37, pp. 3-13.

Ahuett-Garza, H. and Kurfess, T. (2018), "A brief discussion on the trends of habilitating technologies for Industry 4.0 and Smart manufacturing", *Manufacturing Letters*, Vol. 15, pp. 60-63.

Alcácer, V. and Cruz-Machado, V. (2019), "Scanning the Industry 4.0: a literature review on technologies for manufacturing systems", *Engineering Science and Technology, an International Journal*, Vol. 22 No. 3, pp. 899-919.

Berkowitz, L.J.R.-T.M. (1993), "Getting the most from your patents", *Research-Technology Management*, Vol. 36 No. 2, pp. 26-31.

Berry, M.W. and Kogan, J. (2010), *Text Mining: Applications and Theory*, John Wiley & Sons, Chichester.

Bildosola, I., Río-Bélver, R.M., Garechana, G. and Cilleruelo, E. (2017), "TeknoRoadmap: an approach for depicting emerging technologies", *Technological Forecasting and Social Change*, Vol. 117, pp. 25-37.

Blei, D.M., Ng, A.Y. and Jordan, M.I. (2003), "Latent dirichlet allocation", *Journal of Machine Learning Research*, Vol. 3, January, pp. 993-1022.

Bolker, B.M., Brooks, M.E., Clark, C.J., Geange, S.W., Poulsen, J.R., Stevens, M.H.H. and White, J.-S.S. (2009), "Generalized linear mixed models: a practical guide for ecology and evolution", *Trends in Ecology and Evolution*, Vol. 24 No. 3, pp. 127-135.

Bonino, D., Ciaramella, A. and Corno, F. (2010), "Review of the state-of-the-art in patent information and forthcoming evolutions in intelligent patent informatics", *World Patent Information*, Vol. 32 No. 1, pp. 30-38.

Breitzman, A. and Thomas, P. (2015), "The emerging clusters model: a tool for identifying emerging technologies across multiple patent systems", *Research Policy*, Vol. 44 No. 1, pp. 195-205.

Brettel, M., Friederichsen, N., Keller, M. and Rosenberg, M. (2014), "How virtualization, decentralization and network building change the manufacturing landscape: an Industry 4.0 Perspective", *International Journal of Mechanical, Industrial Science and Engineering*, Vol. 8 No. 1, pp. 37-44.

Brown, P.F., Pietra, V.J.D., Mercer, R.L., Pietra, S.A.D. and Lai, J.C. (1992), “An estimate of an upper bound for the entropy of English”, *Computational Linguistics*, Vol. 18 No. 1, pp. 31-40.

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

Chen, B., Wan, J., Shu, L., Li, P., Mukherjee, M. and Yin, B. (2018), “Smart factory of Industry 4.0: key technologies, application case, and challenges”, *IEEE Access*, Vol. 6, pp. 6505-6519.

Choi, C., Kim, S. and Park, Y. (2007), “A patent-based cross impact analysis for quantitative estimation of technological impact: the case of information and communication technology”, *Technological Forecasting and Social Change*, Vol. 74 No. 8, pp. 1296-1314.

Daim, T.U., Rueda, G., Martin, H. and Gerdsri, P. (2006), “Forecasting emerging technologies: use of bibliometrics and patent analysis”, *Technological Forecasting and Social Change*, Vol. 73 No. 8, pp. 981-1012.

Engelsman, E.C. and van Raan, A.F.J.R.P. (1994), “A patent-based cartography of technology”, *Research Policy*, Vol. 23 No. 1, pp. 1-26.

Ernst, H. (2003), “Patent information for strategic technology management”, *World Patent Information*, Vol. 25 No. 3, pp. 233-242.

Frank, A.G., Dalenogare, L.S. and Ayala, N.F. (2019), “Industry 4.0 technologies: implementation patterns in manufacturing companies”, *International Journal of Production Economics*, Vol. 210, pp. 15-26.

Gibbons, R.D., Hedeker, D. and DuToit, S. (2010), “Advances in analysis of longitudinal data”, *Annual Review of Clinical Psychology*, Vol. 6, pp. 79-107.

Griffiths, T.L. and Steyvers, M. (2004), “Finding scientific topics”, *Proceedings of the National Academy of Sciences*, Vol. 101, Supplement 1, pp. 5228-5235.

Harhoff, D., Scherer, F.M. and Vopel, K. (2003), “Citations, family size, opposition and the value of patent rights”, *Research Policy*, Vol. 32 No. 8, pp. 1343-1363.

Hu, Z., Fang, S. and Liang, T. (2014), “Empirical study of constructing a knowledge organization system of patent documents using topic modeling”, *Scientometrics*, Vol. 100 No. 3, pp. 787-799.

Huang, J.Y. (2017), “An analysis of the intellectual structure of the cloud patents of SaaS”, *Technology Analysis and Strategic Management*, Vol. 29 No. 8, pp. 917-931.

Kagermann, H., Helbig, J., Hellinger, A. and Wahlster, W. (2013), “Recommendations for implementing the strategic initiative INDUSTRIE 4.0: securing the future of German manufacturing industry, Final Report of the Industrie 4.0 Working Group, Forschungsunion.

Kamble, S.S., Gunasekaran, A. and Gawankar, S.A. (2018), “Sustainable Industry 4.0 framework: a systematic literature review identifying the current trends and future perspectives”, *Process Safety and Environmental Protection*, Vol. 117, pp. 408-425.

Kusiak, A. (2018), “Smart manufacturing”, *International Journal of Production Research*, Vol. 56 Nos 1-2, pp. 508-517.

Lee, J., Bagheri, B. and Kao, H.-A. (2015a), “A cyber-physical systems architecture for Industry 4.0-based manufacturing systems”, *Manufacturing Letters*, Vol. 3, pp. 18-23.

Lee, W.S., Han, E.J. and Sohn, S.Y. (2015b), “Predicting the pattern of technology convergence using big-data technology on large-scale triadic patents”, *Technological Forecasting and Social Change*, Vol. 100, pp. 317-329.

Lu, H.-P. and Weng, C.-I. (2018), “Smart manufacturing technology, market maturity analysis and technology roadmap in the computer and electronic product manufacturing industry”, *Technological Forecasting and Social Change*, Vol. 133, pp. 85-94.

Ma, T., Porter, A.L., Guo, Y., Ready, J., Xu, C. and Gao, L. (2013), “A technology opportunities analysis model: applied to dye-sensitised solar cells for China”, *Technology Analysis and Strategic Management*, Vol. 26 No. 1, pp. 87-104.

Ma, T.T., Zhang, Y., Huang, L., Shang, L.N., Wang, K.R., Yu, H.Z. and Zhu, D.H. (2017), "Text mining to gain technical intelligence for acquired target selection: a case study for China's computer numerical control machine tools industry", *Technological Forecasting and Social Change*, Vol. 116, pp. 162-180.

Madani, F. and Weber, C. (2016), "The evolution of patent mining: applying bibliometrics analysis and keyword network analysis", *World Patent Information*, Vol. 46, pp. 32-48.

Manning, C.D., Manning, C.D. and Schütze, H. (1999), *Foundations of Statistical Natural Language Processing*, MIT press, Cambridge, Massachusetts.

McCulloch, C.E., Searle, S.R. and Neuhaus, J.M. (2018), *Generalized, Linear, and Mixed Models*, Wiley, Hoboken, New Jersey.

Monostori, L., Kádár, B., Bauernhansl, T., Kondoh, S., Kumara, S., Reinhart, G., Sauer, O., Schuh, G., Sihn, W. and Ueda, K. (2016), "Cyber-physical systems in manufacturing", *CIRP Annals*, Vol. 65 No. 2, pp. 621-641.

Muller, J.M., Buliga, O. and Voigt, K.I. (2018), "Fortune favors the prepared: how SMEs approach business model innovations in Industry 4.0", *Technological Forecasting and Social Change*, Vol. 132, pp. 2-17.

Neter, J., Wasserman, W. and Kutner, M.H. (1989), *Applied Linear Regression Models*, Irwin, Homewood, Illinois.

Ngo, T.D., Kashani, A., Imbalzano, G., Nguyen, K.T. and Hui, D.J.C.P.B.E. (2018), "Additive manufacturing (3D printing): a review of materials", *Methods, Applications and Challenges*, Vol. 143, pp. 172-196.

Oesterreich, T.D. and Teuteberg, F.J.C.i.i. (2016), "Understanding the implications of digitisation and automation in the context of Industry 4.0", *A Triangulation Approach and Elements of a Research Agenda for the Construction Industry*, Vol. 83, pp. 121-139.

Oztemel, E. and Gursev, S. (2020), "Literature review of Industry 4.0 and related technologies", *Journal of Intelligent Manufacturing*, Vol. 31 No. 1, pp. 127-182.

Pereira, A. and Romero, F. (2017), "A review of the meanings and the implications of the Industry 4.0 concept", *Procedia Manufacturing*, Vol. 13, pp. 1206-1214.

Porter, A.L. and Cunningham, S.W. (2004), *Tech Mining: Exploiting New Technologies for Competitive Advantage*, John Wiley & Sons, Chichester.

Porter, A.L., Cunningham, S.W., Banks, J., Roper, A.T., Mason, T.W. and Rossini, F.A. (2011), *Forecasting and Management of Technology*, 2nd ed., Wiley, Chichester.

Porter, A.L., Garner, J., Carley, S.F. and Newman, N.C. (2019), "Emergence scoring to identify Frontier R&D topics and key players", *Technological Forecasting and Social Change*, Vol. 146, pp. 628-643.

Reitzig, M. (2004), "Improving patent valuations for management purposes - validating new indicators by analyzing application rationales", *Research Policy*, Vol. 33 Nos 6-7, pp. 939-957.

Rotolo, D., Hicks, D. and Martin, B.R. (2015), "What is an emerging technology?", *Research Policy*, Vol. 44 No. 10, pp. 1827-1843.

Santamaría, L., Nieto, M.J. and Barge-Gil, A. (2009), "Beyond formal R&D: taking advantage of other sources of innovation in low-and medium-technology industries", *Research Policy*, Vol. 38 No. 3, pp. 507-517.

Small, H., Boyack, K.W. and Klavans, R. (2014), "Identifying emerging topics in science and technology", *Research Policy*, Vol. 43 No. 8, pp. 1450-1467.

Smit, H.T.J. and Trigeorgis, L. (2007), "Strategic options and games in analysing dynamic technology investments", *Long Range Planning*, Vol. 40 No. 1, pp. 84-114.

Stock, T. and Seliger, G. (2016), "Opportunities of sustainable manufacturing in Industry 4.0", *Procedia CIRP*, Vol. 40, pp. 536-541.

Straka, M., Hajič, J. and Straková, J. (2016), "UDPipe: trainable pipeline for processing CoNLL-U files performing tokenization, morphological analysis, POS tagging and parsing", *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*, European Language Resources Association (ELRA), pp. 4290-4297.

Suominen, A., Toivanen, H. and Seppänen, M. (2017), "Firms' knowledge profiles: mapping patent data with unsupervised learning", *Technological Forecasting and Social Change*, Vol. 115, pp. 131-142.

Tao, F. and Qi, Q. (2017), "New IT driven service-oriented smart manufacturing: framework and characteristics", *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, Vol. 49 No. 1, pp. 81-91.

Thoben, K.-D., Wiesner, S. and Wuest, T. (2017), "Industrie 4.0 and smart manufacturing–a review of research issues and application examples", *International Journal of Automation Technology*, Vol. 11 No. 1, pp. 4-16.

Trappey, A.J., Trappey, C.V., Govindarajan, U.H., Sun, J.J. and Chuang, A.C. (2016), "A review of technology standards and patent portfolios for enabling cyber-physical systems in advanced manufacturing", *IEEE Access*, Vol. 4, pp. 7356-7382.

Tseng, Y.-H., Lin, C.-J. and Lin, Y.-I. (2007), "Text mining techniques for patent analysis", *Information Processing and Management*, Vol. 43 No. 5, pp. 1216-1247.

Vaidya, S., Ambad, P. and Bhosle, S. (2018), "Industry 4.0 – a glimpse", *Procedia Manufacturing*, Vol. 20, pp. 233-238.

Venugopalan, S. and Rai, V. (2015), "Topic based classification and pattern identification in patents", *Technological Forecasting and Social Change*, Vol. 94, pp. 236-250.

Wagire, A.A., Rathore, A. and Jain, R. (2020), "Analysis and synthesis of Industry 4.0 research landscape: using latent semantic analysis approach", *Journal of Manufacturing Technology Management*, Vol. 31 No. 1, pp. 31-51.

Wang, B., Liu, S.B., Ding, K., Liu, Z.Y. and Xu, J. (2014), "Identifying technological topics and institution-topic distribution probability for patent competitive intelligence analysis: a case study in LTE technology", *Scientometrics*, Vol. 101 No. 1, pp. 685-704.

Wang, J., Wang, C.Y. and Wu, C.-Y. (2015), "A real options framework for R&D planning in technology-based firms", *Journal of Engineering and Technology Management*, Vol. 35, pp. 93-114.

Wang, S., Wan, J., Li, D. and Zhang, C. (2016), "Implementing smart factory of industrie 4.0: an outlook", *International Journal of Distributed Sensor Networks*, Vol. 12 No. 1, p. 3159805.

Yao, X., Zhang, J., Li, Y. and Zhang, C. (2018), "Towards flexible RFID event-driven integrated manufacturing for make-to-order production", *International Journal of Computer Integrated Manufacturing*, Vol. 31 No. 3, pp. 228-242.

Yoon, B. and Park, Y. (2005), "A systematic approach for identifying technology opportunities: keyword-based morphology analysis", *Technological Forecasting and Social Change*, Vol. 72 No. 2, pp. 145-160.

Zhang, Y., Lu, J., Liu, F., Liu, Q., Porter, A., Chen, H. and Zhang, G. (2018), "Does deep learning help topic extraction? A kernel k-means clustering method with word embedding", *Journal of Informetrics*, Vol. 12 No. 4, pp. 1099-1117.

Zheng, T., Ardolino, M., Bacchetti, A., Perona, M. and Zanardini, M. (2019), "The impacts of Industry 4.0: a descriptive survey in the Italian manufacturing sector", *Journal of Manufacturing Technology Management*, Advance online publication, doi: 10.1108/JMTM-08-2018-0269.

Appendix

Topic labels and descriptions for smart manufacturing

The label and corresponding phrases and brief description for each topic are described in Tables A1 and A2, respectively.

Table A1. Topic labels and corresponding phrases

Topic 1 Smart connection technologies	Topic 2 Production traceability technologies	Topic 3 Smart sensing technologies	Topic 4 Cyber-physical systems	Topic 5 Fault detection & classification	Topic 6 Manufacturing data analytics	Topic 7 Manufacturing automation and robotics
Local data collection (0.037)	Risk (0.022)	Electrode (0.013)	Internet of things (0.104)	Fault detection (0.025)	Machine learning (0.056)	Data analysis (0.057)
Crosspoint switch (0.028)	Location information (0.014)	Station (0.013)	Gateway (0.032)	Operating system (0.015)	Industrial machine (0.022)	Cloud numerical control (0.018)
Multiple output (0.028)	Industrial communication flow (0.011)	Solid electrolyte layer (0.009)	Field device (0.03)	Virtual metrology (0.014)	Artificial intelligence (0.016)	Schedule (0.017)
User equipment (0.027)	Supply chain (0.011)	Self-diagnosis (0.007)	Cloud platform (0.025)	Control module (0.013)	Label (0.015)	Production line (0.015)
Internet of things (0.022)	Electronic device (0.01)	Analytical unit (0.006)	Cloud server (0.019)	Abnormality (0.013)	Measurement data (0.013)	Network infrastructure (0.014)
Computing cloud (0.011)	Data packet (0.009)	Particle sensor (0.006)	Embedded system (0.016)	Anomaly (0.012)	Big data (0.011)	Robot (0.012)
Data acquisition module (0.008)	Smart tag (0.009)	State machine (0.006)	Public key (0.012)	Recipe (0.012)	Predictive model (0.007)	Internet of things (0.01)
Detection device (0.007)	Industrial robot (0.008)	Predictive maintenance (0.006)	Receiver device (0.01)	Advanced process control (0.012)	Image data (0.006)	Simulator (0.009)
Sensor device (0.007)	Electronic tag (0.006)	Gas sensor (0.005)	Data processing system (0.009)	Diagnostic (0.011)	Classifier (0.006)	Quality control (0.008)
Beacon (0.007)	Tag read (0.004)	Sensor network (0.005)	Operation information (0.008)	Metrology data (0.01)	Neural network (0.006)	Resource allocation (0.007)

(*continued*)

Topic 8 Remote monitoring and control	Topic 9 Cloud-based service technologies	Topic 10 Industrial internet network management	Topic 11 Powder-bed fusion technologies	Topic 12 Material extrusion technologies	Topic 13 Products and apparatus made by additive manufacturing	Topic 14 Additive manufacturing for tool and mold making
Defect (0.034)	Control application (0.024)	Bus controller (0.033)	Three-dimensional article (0.027)	Filament (0.03)	Platen (0.013)	Mold (0.031)
Time series data (0.03)	Query (0.022)	Bus terminal (0.018)	Lens (0.011)	Housing (0.028)	Rotor (0.013)	Shell (0.018)
Input data (0.016)	Cache (0.022)	Industrial internet field (0.014)	Fixture (0.01)	Reservoir (0.015)	Feed material (0.01)	Build chamber (0.018)
Output data (0.016)	Cloud service (0.018)	Time slice (0.012)	Curable material (0.009)	Preform (0.014)	Stator (0.009)	Internal passage (0.015)
Process condition (0.013)	Data stream (0.011)	Configuration information (0.011)	Layer of build (0.008)	Feedstock (0.009)	Winding (0.009)	Slice (0.012)
Sensor network (0.012)	Machine to machine (0.01)	Best master clock (0.006)	Metallic powder (0.007)	Solvent (0.008)	Film (0.009)	Fibre (0.011)
Industrial asset (0.012)	Summary report (0.01)	Remote terminal (0.006)	Cartridge (0.007)	Matrix material (0.008)	Roller (0.009)	Cell (0.01)
Hot spot (0.008)	Beacon badge (0.009)	Service oriented automation (0.006)	Ophthalmic lens (0.007)	Column (0.007)	Blade (0.008)	Liquid metal (0.01)
Test data (0.008)	Cloud service control (0.006)	Cryptographic key (0.005)	Printhead (0.007)	Deposition head (0.006)	Circuit board (0.008)	Lattice structure (0.008)
Industrial control device (0.007)	Service architecture (0.006)	Heterogeneous field device (0.005)	Enclosure (0.007)	Extrusion head (0.006)	Coil (0.008)	Aperture (0.007)

Table A1.

Topic	Description	Representative patents and probabilities
1. Smart connection technologies	Methods and systems related to managing and acquiring data made available in real time and transmitting them with specific network protocols	US20180048713A1(0.827) WO2018203923A1(0.786) US20180255377A(0.750)
2. Production traceability technologies	Techniques related to real-time tracking manufacturing parts, inventory and assets and monitoring status of jobs or work orders	EP3136316A1(0.833) US20140240249A1(0.763) US20040100383A1(0.773)
3. Smart sensing technologies	Techniques related to sensor and measurement technologies to gather data and monitor actions through the production environment, including self-evaluation of the validity of measurement data	US20110030451A1(0.875) US20170299543A1(0.813) WO2015166751A1(0.795)
4. Cyber-physical manufacturing technology	Systems and methods to facilitate implementation of a cyber-physical system for assessing and/or predicting manufacturing problems	WO2018025477A1(0.845) US20180054376A1(0.795) US20160321081A1(0.763)
5. Fault detection and classification	Techniques related to detection and classification of faults for evaluating manufacturing quality and health status of industrial assets and improving manufacturing operations based on collected data, especially in process control	US20180143248A1(0.804) US20070100487A1(0.775) US20170220008A1(0.775)
6. Manufacturing data analytics	Techniques related to using artificial intelligence and big data analytics to build model for improving manufacturing performance, such as predicting outcomes and analyzing operational risks and security	US20180322234A1(0.786) US20160004794A1(0.763) US20170157767A1(0.750)
7. Industrial automation and robotics	Techniques, systems or apparatus to facilitate factory automation and adapted for organizing, planning and scheduling industrial assets	EP2783812A2(0.888) US20080300705A1(0.804) US20180121815A1(0.795)
8. Remote monitoring and control	Methods and systems for monitoring and controlling locally or remotely for improving manufacturing quality and performance to the unpredictable environmental changes	EP3260849A1(0.845) EP3343421A1(0.827) US20080300709 A1(0.810)
9. Cloud-based service technologies	Methods and systems related to using cloud computing technologies to manage, allocate and control manufacturing resources and capabilities	EP3343372A1(0.882) EP3346381A1(0.871) EP3057007A1(0.833)
10. Industrial Internet network management	Techniques related to network-specific arrangements or communication protocols for supporting connecting resources of the IoT and network security in the manufacturing environment	EP3261296A1(0.895) EP3261275A1(0.882) EP3261298A1(0.859)
11. Power bed fusion technologies	Techniques related to power bed fusion additive manufacturing, such as processes, apparatus, material and data acquisition and processing	EP3254784A1(0.940) EP3064295A1(0.932) EP3153253A1(0.906)

Table A2.
Topic description

(*continued*)

Topic	Description	Representative patents and probabilities
12. Material extrusion technologies	Techniques related to processes, apparatus, data acquisition/processing and adaptive material for material extrusion additive manufacturing processes	US20180250737A1(0.932) EP3081364A1(0.921) US20170050363A1(0.917)
13. Products and apparatus made by additive manufacturing	Techniques related to products and apparatus made by additive manufacturing in various industries such as electronics, auto and aerospace	WO2017179748A1(0.958) EP2858076A1(0.956) EP2754516A2(0.95)
14. Additive manufacturing for tool and mold making	Techniques related to making tools and molds using additive manufacturing for cost reduction and rapid manufacturing	EP3342572A1(0.948) EP2716390A2(0.946) WO2018182686A1(0.929)

Table A2.

Corresponding author
Juite Wang can be contacted at: rdwang@nchu.edu.tw