Technological Forecasting & Social Change 172 (2021) 121064

Contents lists available at ScienceDirect

Technological Forecasting & Social Change

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

Exploring the technology emergence related to artificial intelligence: A perspective of coupling analyses

Munan Li *, Wenshu Wang, Keyu Zhou

School of Business Administration, South China University of Technology, Guangzhou, China

ARTICLE INFO

Keywords:
Artificial intelligence
Coupling analyses
Topic mining
LDA (Latent Dirichlet Allocation)
Coupling strength
Coupling velocity

ABSTRACT

With the spillover of the relevant knowledge on AI (artificial intelligence), an increasing number of scientific topics in the field of AI are meeting many opportunities, including transformation and applications in a multitude of areas; therefore, exploring the TE (technology emergence) and TO (technology opportunities) related to AI highlights the meaning and value. To further visualize the TE or underlying TO on AI, a perspective of coupling analyses and a computing framework on coupling relationships between publications and patents are proposed. AI has become a complicated interdisciplinarity field, and increasingly different categories are involved in the domain of AI; therefore, identifying the relevant TE or TO related to AI has become a relatively intractable but valuable problem, and our work presented in this paper can broaden the vision and bring some insights into AI development. Moreover, the proposed indicators of the coupling strength and coupling velocity and their computing methods can provide new insights or perspectives for exploring the TE or TO of a specific topic and can also enrich the relevant methodologies for technical opportunity analysis and coupling analysis between publications and patents.

1. Introduction

TE (technology emergence) and TOA (technology opportunities analysis) have attracted increasing attention from academia and industry (Ma et al., 2014; Carley et al., 2018; Li et al., 2021). From a conceptual perspective, TE can have some coherence with terminologies such as ET (emerging technology) and TM (tech mining) (Porter and Newman, 2011; Guo et al., 2012). Rotolo et al. (2015) proposed a conceptual framework of ET that defines several features regarding the emergence of novel technologies: radical novelty, relatively fast growth, coherence, prominent impact, uncertainty and ambiguity. Zhang et al. (2016) also believed that recognizing the emergence of new fields is complex and uncertain and requires the mobilization of various sources of intelligence for policy making or strategic decisions. Burmaoglu et al. (2019) argued that TE demonstrates qualitative novelty, synergy, trend irregularity, high functionality, and continuity. To position a TE or identify an ET, patent data are often used as the critical data source, and then the relevant patent analysis techniques, NLP (natural language processing), text mining and topic modeling based on patent texts are some of the mainstream approaches (Venugopalan and Rai, 2015; Zhang et al., 2016; Suominen et al., 2019).

Regarding whether TE is equal to or very similar to ET from the perspective of conceptualization, some debates remain. For instance, Li et al. (2021) argued that TE can be defined as a self-organized phenomenon that is highly correlated with technology development; patenting behavior and research activities can be considered two critical indicators for identifying or measuring TE. Regarding TOA or TOD (technology opportunities discovery), diversified indicators and analytical or discovery methods are the focus of relevant studies, and highly corresponding topics include patent analysis, semantic analysis, text mining and morphology analysis (Yoon and Kim, 2012; Yoon et al., 2014; Wang et al., 2017).

Additionally, in recent decades, AI (artificial intelligence) has become one of the most popular discourses; the relevant academic and industrial activities regarding AI have also experienced rapid growth. Furthermore, the event of the AlphaGo computer program beating a top professional player attracted significant attention from the public in 2016; deep neural network technology related to AlphaGo also instantly became the focus of academia and industry (LeCun et al., 2015; Silver et al., 2016).

With respect to AI and the relationships among AI, machine intelligence, machine learning and deep learning, acknowledged definitions and a general consensus are still lacking. Brunette et al. (2009) reviewed the history of the AI field and concluded that limited agreement with

* Corresponding author.

https://doi.org/10.1016/j.techfore.2021.121064
Received 29 March 2020; Received in revised form 17 July 2021; Accepted 20 July 2021
Available online 31 July 2021

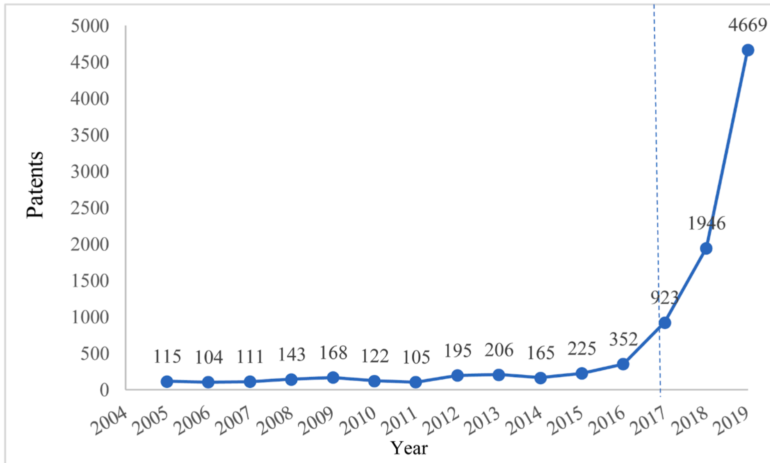

Fig. 1. A simple query of the filed patents related to AI in the DII (Derwent Innovations Index) database[11]

respect to the knowledge boundary or field of AI has been achieved. While reviewing the methods for IPA (intellectual property analytics), Aristodemou & Tietze (2018) argued that AI, machine learning and deep learning represent different techniques or at least different stages of IPA. However, while reviewing the latest studies during the past five years, many scholars and publications prefer to attribute machine learning and deep learning to the general scope/field of AI, particularly for some areas of engineering (Borges et al., 2021). While mapping the evolution and geography of a general-purpose technology, Klinger et al. (2021) argued that deep learning is one of the core techniques in the AI field.

In addition to the relevant technologies for deep learning, the development of theories, algorithms and technologies related to AI has become an important means for major economies to maintain or obtain competitive advantages in the upcoming decades, and AI has also become an important component of national S&T (science and technology) and industrial policies. Many countries worldwide have increased their investment in AI research and development, and the relevant research and technological transformations, applications and commercialization have also been vigorously presented. However, specific studies on AI TE or TOA are still rare, especially how to visualize TE or TO through coupling analyses between AI publications and AI patents, which still leaves much room for exploration.

With the development of computer technology, communications and the Internet, in the late 1990s, AI entered a rapid developmental period (Simon and Munankata, 1997; Bench-Capon and Dunne, 2007; Gosavi, 2009). Some scientists and researchers have begun to consider that relevant AI technologies and applications can have a significant impact on the traditional modes of economic growth and social governance; moreover, AI can also bring some grand challenges to humanity (Gosavi, 2009; Daugherty et al., 2019). To examine the increase in filed patents relevant to AI, the steep growth that has occurred since 2017 and one simple query for AI patents are shown in Fig. 1. The results demonstrate that the technological transformations and applications related to AI have also begun to experience rapid growth.

Regarding the rapid growth of AI patents in the DII (derwent innovations index) database, some interesting questions related to the TOA or TE are raised from the observed phenomenon: (1) which AI subtopics can be associated with TE or TOA; (2) what are the coupling relationships between AI studies and AI applications; and (3) which AI subtopics have the potential for technological transformations and applications in the future?

To explore the above questions, traditional techniques, including coupling analyses between publications and patents, topic modeling and several new indicators, are integrated into an integrated framework, and subsequent empirical analyses are expected to produce the findings or implications for policy making. Furthermore, we attempt to enrich the established knowledge base of TOA and TE to some extent.

2. Literature review

In terms of identifying or evaluating a TE or the potential TO (technology opportunities), techniques such as bibliometrics, patent analysis, technology roadmapping, and text mining are all traditional and effective approaches (Porter, 1999; Daim et al., 2006; Li, 2015; Li et al., 2018). For example, Wang et al. (2015) proposed an analytical SAO (subject-action-object) method to identify technology development trends. Bruck et al. (2016) proposed a technique to recognize ET trends through a class-selective study of citations in the US PCN (patent citation network). Avila-Robinson and Miyazaki (2013)3 discussed the proxy effect of a scientific knowledge base on discerning a TE or TO. Many other studies also exist.

To enrich the collection of technical intelligence, Zhang et al. (2016) proposed technology roadmapping. To probe the technological development trend of 3D printing, Huang et al. (2017) proposed a hybrid method including coclassification, coword and main path analyses. Love-Koh (2018) combined a pragmatic literature search with expert interviews to probe the future impact of precision medicine. Carley et al. (2018) designed a software script to generate a family of emergence indicators for a topic of interest based on the publication or patent text. In addition to traditional patent analysis, Li et al. (2019) introduced Twitter data mining to enhance the identification and monitoring of the development trends of emerging technologies. Regarding the recognition of the frontier of R&D (research & development) and the key players, a technique for emergence scoring was proposed (Porter et al., 2019).

To identify or assess a TE or TO, publications and patents are the primary data sources, and the coupling relationships between publications and patents on the same topic are significant (Venugopalan and Rai, 2015); therefore, relevant studies on TE, TM, publications and patents are often integrated to probe or analyze TE and discover TO. In recent studies, Kwon et al. (2019) argued that a paper containing technologically emergent ideas is positively and strongly associated with its future citation impact in specific areas. In terms of identifying emerging technologies in their early stages, the corresponding scientific studies and knowledge bases can play much more important roles (Zhou et al., 2019). Therefore, to identify, probe or assess a TE for a specific topic, coupling the analyses of publications and patents can be a very valuable and effective means of doing so.

In terms of TE or other relevant issues with respect to AI, Abbas et al. (2019) analyzed the application prospects and agendas of some relevant

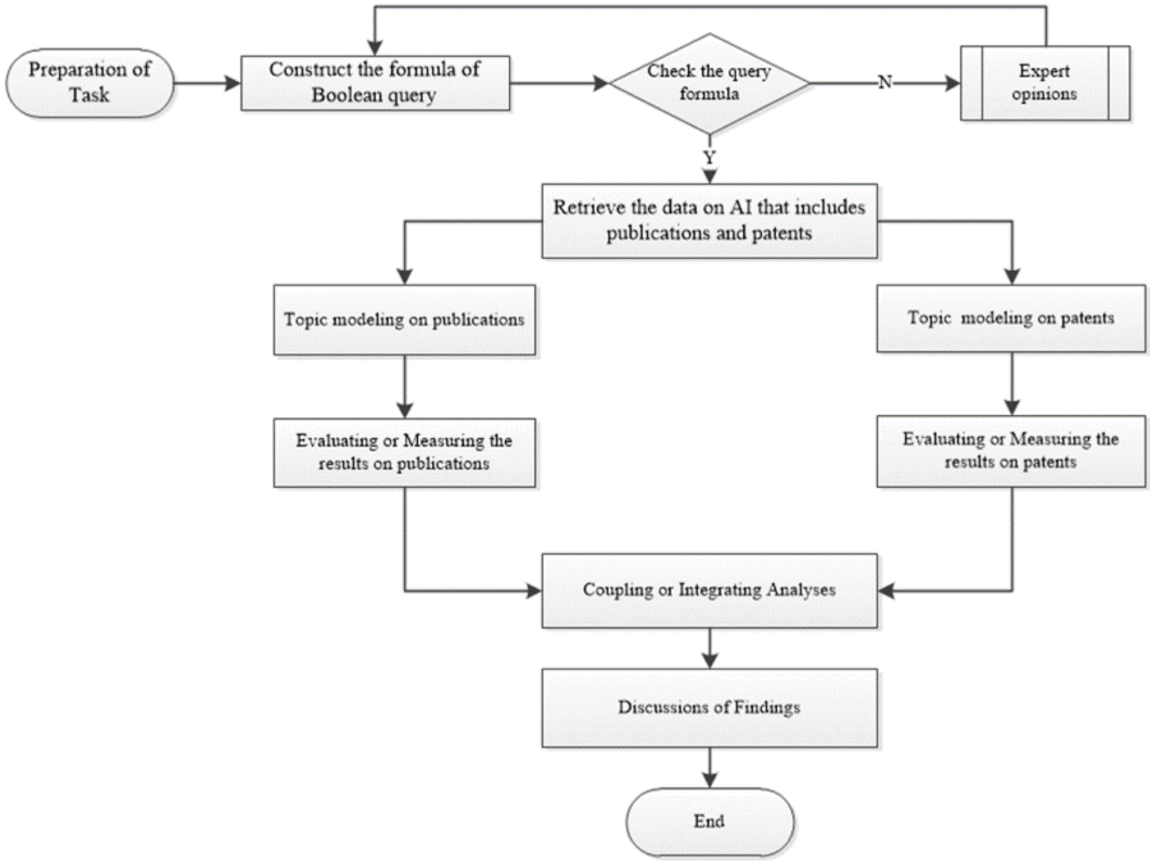

Fig. 2. The paradigm of the research flow in this paper.

AI technologies in the field of information security. Agrawal et al. (2019) argued that AI technology applications can have some social impacts on labor substitution or other corresponding issues. Although an increasing number of relevant AI-based technologies have emerged in the past five years, relevant studies on TE, technology forecasting, and TM in the field of AI are still very rare and inadequate. In addition, the relevant studies on the coupling relationships between publications and patents are also insufficient; in particular, very few quantitative indicators and measurement methods exist.

Recently, techniques for topic modeling and machine learning have been gradually introduced to technological and competitive intelligence analyses. For example, Suominen & Toivanen (2016) compared unsupervised learning with human-assigned subject classification while mapping science topics. Wang et al. (2020) proposed a method based on the latent Dirichlet allocation (LDA) model to evaluate the competitiveness of enterprise technology. Therefore, topic modeling based on LDA or other similar algorithms has the potential to be an important technique for relevant studies on TE, especially for coupling analyses between publications and patents.

3. Research design

In recent decades, relevant publications and patents on AI have exhibited rapid growth, and an increasing number of different categories and research areas can be related to AI; therefore, to extract diversified studies or subtopics, topic modeling can be an appropriate approach. Regarding topic modeling, LDA is currently a popular and mature algorithm (Blei et al., 2003). LDA is a Bayesian model that consists of three layers, a subject, a document and a topic, and is fully based on the Bayesian reasoning mechanism, so it has a good knowledge interpretation capacity (Blei, 2012; Bastani et al., 2019).

However, LDA only provides topic model extraction for text data and does not consider the specific mechanisms or explanatory abilities of the relationships between different topics (Blei, 2012; Suominen and Toivanen, 2016; Guo et al., 2017). Clearly, AI has evolved into a field with comprehensive and interdisciplinary S&T (science & technology) topics. Effectively presenting the development of AI requires an integrated analytical framework. The following is a schematic diagram of the research idea of this paper.

There are several key technical points in Fig. 2:(1) is the collection of publications and patents related to AI. Considering that AI has evolved into a typical multidisciplinary and cross-subject field, the retrieval mode is constructed in a hybrid way; (2) one window represents a decade, so each window can be roughly divided into three periods, which can be controversial or compromised; (3) the LDA algorithm is conducted for topic modeling of the collected publications and patents in different periods; (4) the results of topic modeling on AI publications and patents are evaluated; (5) the latent relationships between publication topics and patent topics are analyzed; and (6) the relevant issues of TE or TO with respect to AI are discussed.

The basic LDA model is a three-layer Bayesian probability model that contains words, topics and documents. LDA can extract understandable and relatively stable potential semantic structures (Blei et al., 2003). The basic principle of LDA is to define the documents as D, the potential topics as T, and the total number of feature words as W. Then, the probability of the occurrence of a certain feature word in a certain document can be expressed as in formula (1).

$$P(\omega_i) = \sum_{j=1}^{T} P(\omega_i | z_i = j) P(z_i = j) \tag{1}$$

[1] The Boolean query is the simple formula: TS="artificial intelligence".

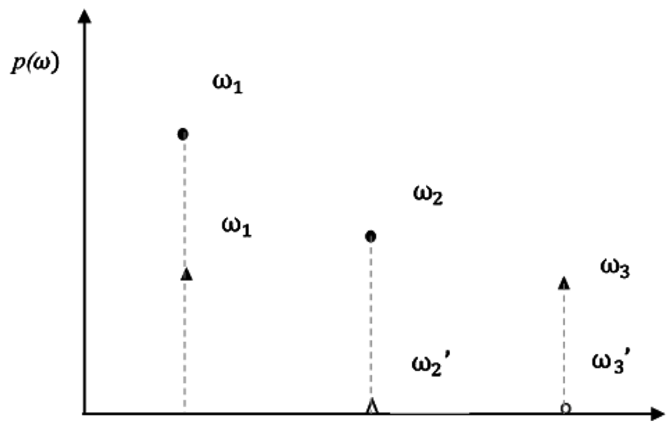

Fig. 3. The reflection points of the terms in Euclidian space.

In formula (1), z_i represents a potential topic containing the characteristic word $\omega_i, P(z_i = j)$ represents the probability that ω_i belongs to topic j, and $P(\omega_i|z_i = j)$ represents the probability that topic j belongs to the document containing the key word ω_i.

Here, we expand these probabilities to their general forms, with theta $\theta^{d}_{z=j}$ and $\phi^{z=j}_{\omega}$; $\theta^{d}_{z=j}$ designed to indicate the document-topic probability distribution and $\phi^{z=j}_{\omega}$ used to indicate the topic-term probability distribution, so θ and ϕ are the core parameters of the LDA model. Furthermore, $\theta^{(d)}$ obeys the Dirichlet distribution with a parameter alpha, $\phi^{(z)}$ is governed by the Dirichlet distribution with a parameter beta, so the overfitting problem can be partly avoided. The prior probability assumptions of θ and ϕ being discrete (alpha) and discrete (beta), respectively, are expressed in Eq. (2).

$$\begin{aligned} &z_i|\theta^{(d_i)} \sim Discrete\left(\theta^{(d_i)}\right), \theta \sim Discrete(\alpha) \\ &\omega_i|z_i, \phi^{(d_i)} \sim Discrete(\phi^{(z_i)}), \phi \sim Discrete(\beta) \end{aligned} \tag{2}$$

In formula (2), α represents the frequency of a topic initially extracted from the document, and β represents the frequencies of terms (featured words) initially extracted from the topic. It is generally believed that different topics and terms are used in basically the same way, so α and β can be set to the same value before running the LDA model.

The running process of the LDA model is:

ü For $Discrete(\alpha)$, sample $\theta^{(d)}$, i.e., the probability distribution of the topic in the document.
ü For $Discrete(\beta)$, sample, $\phi^{(z)}$, i.e., the probability distribution of feature words following the given theme.
ü For θ and ϕ, sample $z_i = P(z_i|\theta)$ and $\omega_i = P(\omega_i|z_i, \phi)$.
ü Repeat the process until each word in the document has been examined.

Determining the optimal number of topics is a difficult problem in topic modeling and a prerequisite to achieving optimal topic clustering results. Currently, the popular reference indexes are perplexity and coherence (Blei, 2012; Stevens et al., 2012; Hu et al., 2014).

In NLP (natural language processing), an indicator of perplexity is used to measure the quality of trained language models and evaluate the generalization abilities of the models. It describes the reciprocal of the set mean of the occurrence probabilities of all words in the test set, and its calculation method is shown in formula (3) (Blei et al., 2003; Blei, 2012).

In general, the lower the perplexity is, the better the generalization ability of the model and the better its adaptability to new samples (Blei et al., 2003). The LDA algorithm can be trained by successively increasing the number of topics, and the perplexity of the model can be observed to make decisions. When the number of topics is higher than a certain threshold value, the model's perplexity continues to decrease and tends to be flat, and the optimal topic number can be determined near the chosen value (Aletras et al., 2017; Shin et al., 2017).

$$\text{Perplexity} = \exp\left\{ -\left(\sum_{m-1}^{M} \sum_{n-1}^{N_m} \log\left(\sum_{k-1}^{K} p(\omega_n|)p(z_k|d_m) \right) \right) \Big/ \left(\sum_{m-1}^{M} N_m \right) \right\} \tag{3}$$

An indicator known as coherence can be used to measure the relevance of topics by calculating the degree of correlation between the highest-scoring feature words in a topic, and it can facilitate dividing the topic into understandable categories (Hu et al., 2014; Shin et al., 2017). The computation of coherence is shown in formula (4). The larger the coherence score is, the better the clustering effect of the model will be. Therefore, the smallest number of clusters that can significantly enhance the coherence of the model is selected as the optimal number of clusters (Mimno et al., 2011; Stevens et al., 2012).

$$\text{Coherence} = \sum_{i<j} score(\omega_i, \omega_j, \varepsilon) \tag{4}$$

Due to some disputes regarding the selection of the optimal topic number for model perplexity (Mimno et al., 2011; Baumer et al., 2017; Arora et al., 2018), the indicator of model coherence is the rule with higher priority for determining the optimal topic number.

To describe and compute the distances between different topics, the basic idea of the Euclidean distance, which refers to the distance between two points in n-dimensional Euclidian space or the natural length of a vector, is used.

$$\begin{aligned} d(x,y) &= \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \ldots + (x_n - y_n)^2} \\ &= \sqrt{\textstyle\sum_{i=1}^{n} (x_i - y_i)^2} \end{aligned} \tag{5}$$

Based on the principle of Euclidean distance, the distance between any two topics in different periods is defined. First, the terms or feature words in any two topics can be assumed to be points in a two-dimensional plane, and the horizontal coordinate of each point represents its attribute; that is, the same feature words have the same horizontal coordinate, and the vertical coordinates represent their probabilities in their respective topics for determining the position of each feature word in the plane.

Second, a point with the same abscissa and zero ordinate can be added for each point without the same feature word in another topic; that is, a feature word with the same attribute but zero probability is added so that the feature words in the two themes form a one-to-one correspondence in the plane.

Finally, the distance of each group or set of characteristic words can be obtained through the distance formula between two points in a two-dimensional plane, and the distance between two topics can be expressed as the sum of all the distances of the characteristic words, as shown in Fig. 3 and Eq. (6).

Basically, while these terms can be reflected in the specific Euclidian space, the distance between two points (terms) can be calculated by formula (6).

$$d_{(i,j)} = \sum \sqrt{\left(p_{subtopic_i}(\omega) - p_{subtopic_j}(\omega)\right)^2}, d_{i,j} \in [0, m] \tag{6}$$

In addition, if the feature words only appear in one of the topics, this distance is the sum of the probabilities of all the feature words. Simply, the above formula can be described as follows: assume ω_1 and ω_2 belong to the first topic, and ω_1 and ω_3 belong to the second topic; then, the total distance between them can be expressed as in Eq. (9).

$$d_{1,2} = p_{subtopic_1}(\omega_2) + p_{subtopic_2}(\omega_3) + \left|p_{subtopic_1}(\omega_1) - p_{subtopic_2}(\omega_1)\right| \tag{7}$$

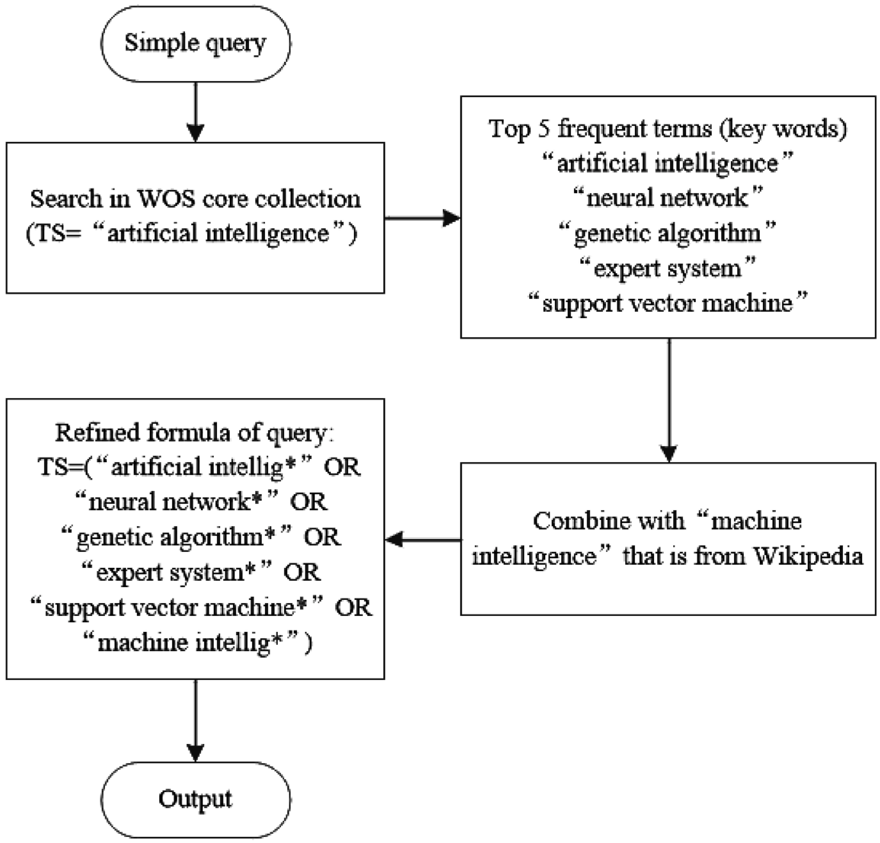

Fig. 4. The flow of finding the search formula on AI publications and patents.

To further describe the possible coupling relationships between publications and patents, two indicators are proposed: they are defined as the CS (coupling strength) and CV (coupling velocity), and they are shown in formula (8) and formula (9), respectively.

Within the same macrotopic of science or technology, e.g., AI, additive manufacturing and nanomedicine, subtopics can present significant differences between publications and patents; therefore, it is still unclear how to measure the possible coupling relationships between subtopic *j* in publications and various subtopic j in patents. In other words, how many subtopics in patent texts have relationships with the specified subtopic *j* in publications? Furthermore, each topic output from LDA or other similar models for topic modeling contains some terms whose memberships are also in the form of probabilities, i.e., for any term *i*, it belongs to subtopic *j* only under a certain probability. Here, a relatively simple formula is proposed to depict the possible coupling between subtopic *j* of the publication text and the patent text under the same macrotopic, and it is shown in formula (8).

$$CS_{subtopic_j} = \sum_{\omega=1}^{k} p_j(\omega) m_\omega \tag{8}$$

In formula (8), $CS_{subtopic_j}$ represents the coupling (strength) of subtopic *j* in the specified publication text with the related patent text under the same macrotopic; $p_j(\omega)$ is the probability of term *k* belonging to subtopic *j*; and m_k is the total number of subtopics in the patent text that contain term *k*.

To describe the CV of subtopic *j* in a specific publication text, the philosophy of half-life calculations is conducted to form the computational method, i.e., when term *k* that belongs to subtopic *j* occurs in the publication text and patent text during the same period (e.g., year, month, week), the CV is defined as 1.0, and the CV is 0.5, while term *k* only occurs in the patents during the second period. By parity of reasoning, while term k only occurs during the n^{th} period, the velocity is 0.5 to the *(n-1)*th power, so the computational formula is as shown in Eq. (9).

$$CS_{sub-topic_j} = \sum_{t=1}^{T} \sum_{k=1}^{N} P_j(k) \times M_t(k) \times (1/2)^{(t-1)} \tag{9}$$

In formula (9), the symbol T is the number of periods, N is the number of terms in subtopic j, $P_j(k)$ is the probability of the kth term belonging to subtopic *j*, and $M_t(k)$ represents the number of patent subtopics including term *k* during period *t*.

4. Empirical analyses

4.1. Data source and retrieval rules

AI has become a highly interdisciplinary topic. How to accurately retrieve highly relevant publications and patents on a specific topic is not yet easy. To extend this type of simple query, the proper search formula is generated by a hybrid approach and shown in Fig. 4.

Regarding Fig. 4, the keyword "artificial intelligence" is first utilized to retrieve the original data, based on which the highly frequent terms extracted by the coword analysis are considered candidate keywords of the next query. Second, the search formula is reconstructed by combining the most frequent terms with the definitions of AI in Wikipedia. Finally, we determine that the retrieval formula of AI is TS= ("artificial intelligence*" OR "neural network*" OR "genetic algorithm*" OR "expert system*" OR "support vector machine*" OR "machine intelligence*"), and the retrieval scope includes the WOS (Web of Science) Core Collections and the DII (Derwent Innovations Index).

From 1990 to 2017, more than 220 thousand publications and more than 29 thousand patents are retrieved, and these are used to construct

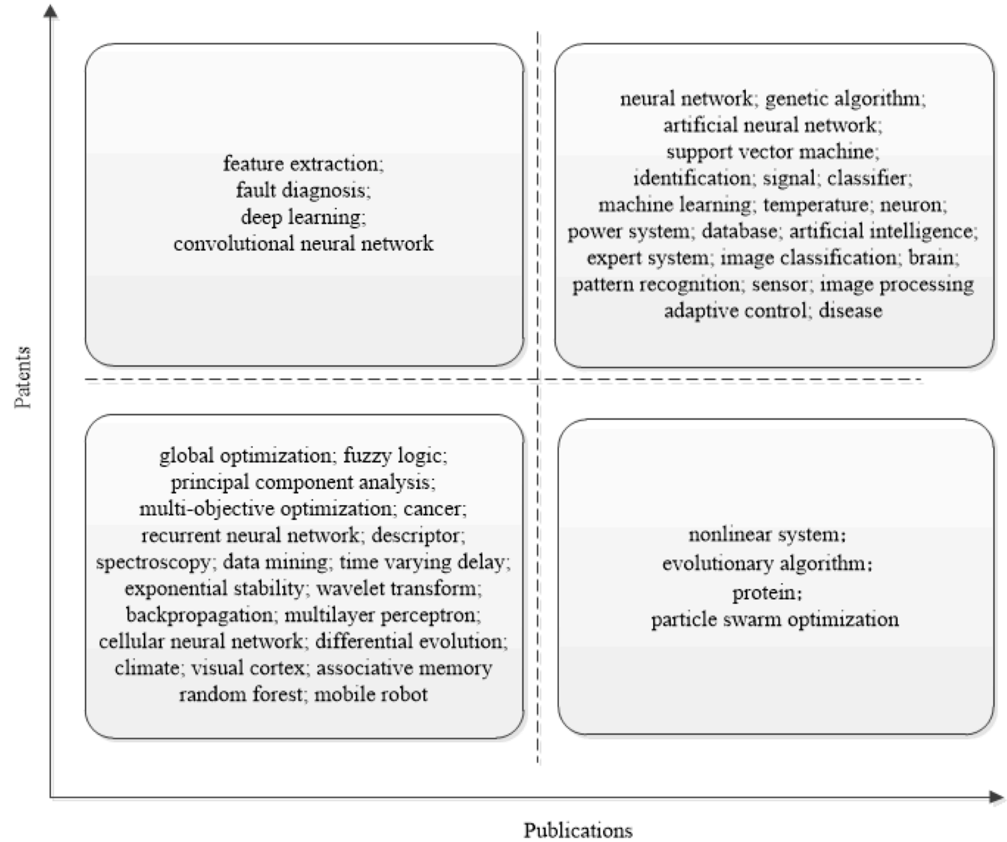

Fig. 5. Simple coupling of the top 50 most frequent terms in the selected AI publications.

the data source for the topic modeling and coupling analyses in the following sections. However, the above formula for searching AI articles and patents is a compromised solution that can be controversial. Finding an accurate formula to search for AI articles and patents is still very complicated. Because the connotations and boundaries of AI are dynamic and evolving, systematic methods for constructing search strategies are also preferred, similar to the search strategies used when the term ‘big data’ was created (Huang et al., 2015).

4.2. Coupling analysis of highly frequent terms between publications and patents

Based on the above query formula for AI, the publication and patent datasets can be retrieved. To analyze the coupling relationships of highly frequent terms in the selected publication text with the corresponding patent text, the top 50 most frequent terms, excluding the general words in publications, are extracted; the relevant patents are retrieved based on these terms, and the results are shown in Table 1.

According to the results presented in Table 1, the top 50 most frequent terms in the selected publication text can be divided into four categories: (1) more publications and more patents, (2) more publications and fewer patents, (3) fewer publications and more patents, and (4) fewer publications and fewer patents. Here, the median values of publications and patents are utilized to evaluate the “more” and “fewer” groupings. Certainly, this evaluation method is very rough; however, this division of these highly frequent terms can also present some interesting information, which is shown in Fig. 5.

Based on the simple coupling of AI publications with AI patents, some additional information can also be collected; for example, terms such as *neural networks, genetic algorithms, support vector* machines, and *expert systems* are not only popular in AI studies but also very popular in AI patents, and such relevant issues, including *random forests, mobile robots*, and *differential evolution,* still have considerable room for both scientific and technological exploration. However, the outputs presented in Fig. 5 are based only on the search formula depicted in Fig. 4, which can be controversial, and another study may draw the opposite results based on different query strategies. The connotation and knowledge boundaries of AI continue to dynamically develop and evolve.

Terms including *feature extraction, deep learning* and *convolutional neural networks (CNNs)* are categorized as having relatively fewer papers and more patents, which can mean that the developmental technology or commercialization surrounding these terms is still active. To further obtain more detailed information on AI studies and AI applications, the LDA algorithm and the proposed indicators in the above content are utilized to implement the following empirical analyses.

4.3. Topic modeling on AI publications and patents

The WOS core collection spans 1990 to 2017, and the time frame can be roughly divided into three periods. The three periods of data are 1990–1999 (the first period), 2000–2009 (the second period) and 2010–2017 (the third period). For the data preprocessing step, traditional techniques such as word segmentation, stop word removal, stem extraction, word form reduction, and term matrix construction for text data are conducted to guarantee the topic modeling performance in the following steps.

The computer hardware environment used in the topic modeling experiment includes the following components. The CPU is an Intel (R) Core i7 at 3.5 Ghz, and the system has 16 GB of RAM. The software environment is Windows 10. Python 3.7 software is used to train and run

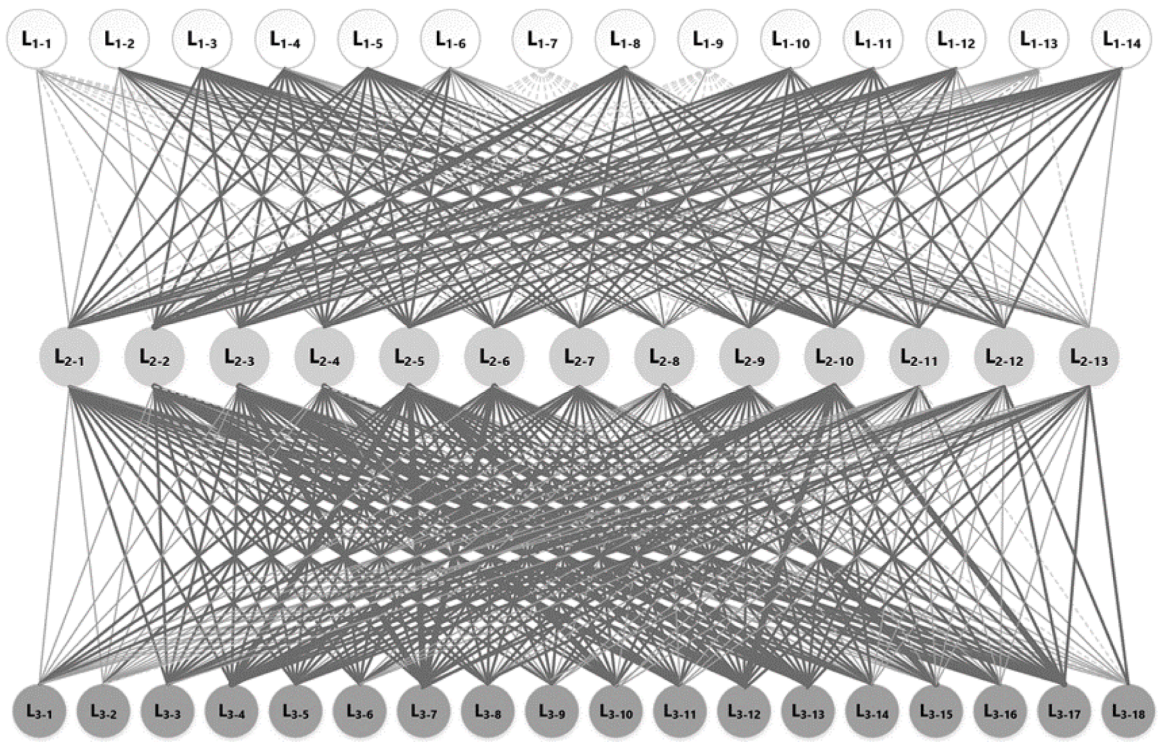

Fig. 6. The connection strengths between different subtopics in the publication texts on AI.

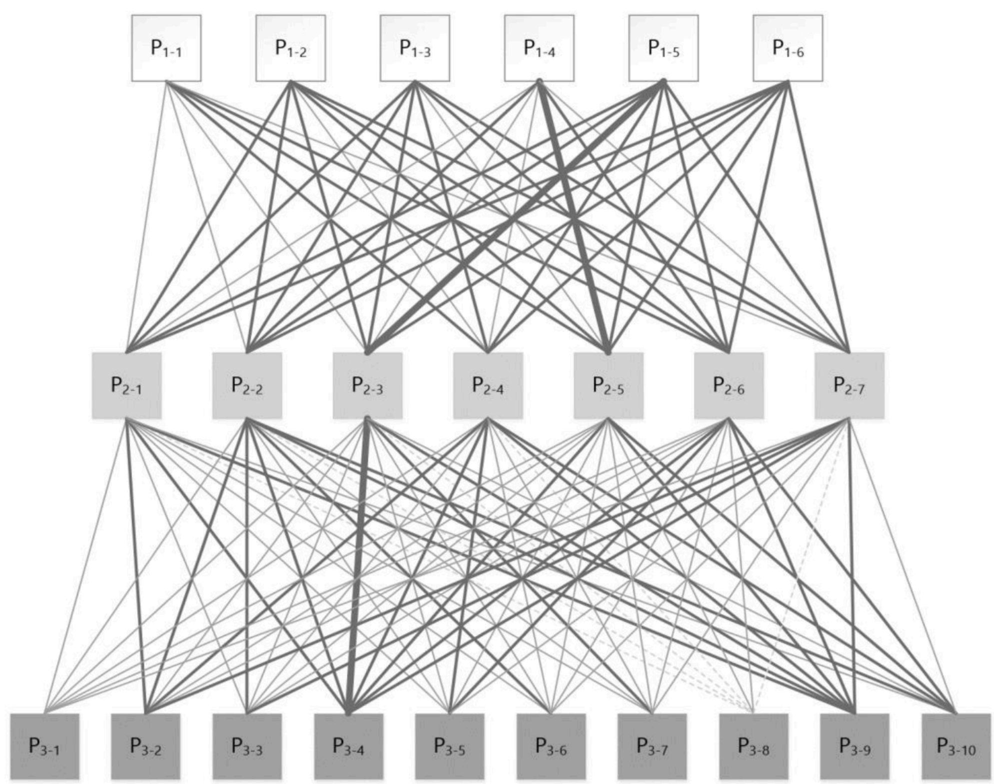

Fig. 7. The connection strengths between different subtopics in the patent texts on AI.

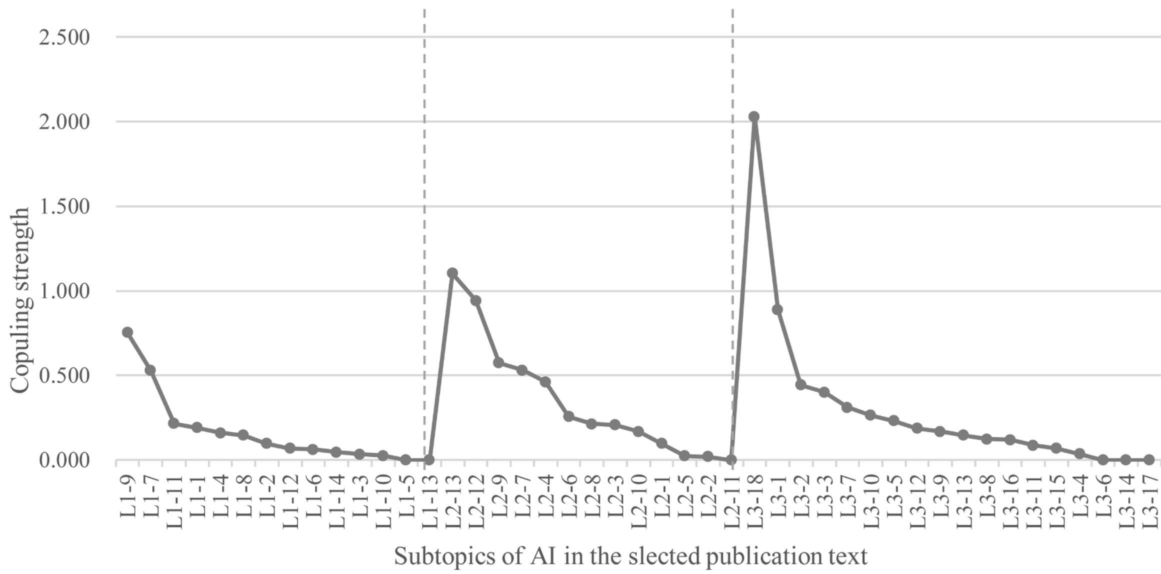

Fig. 8. The coupling strengths of the AI subtopics in the selected publication texts.

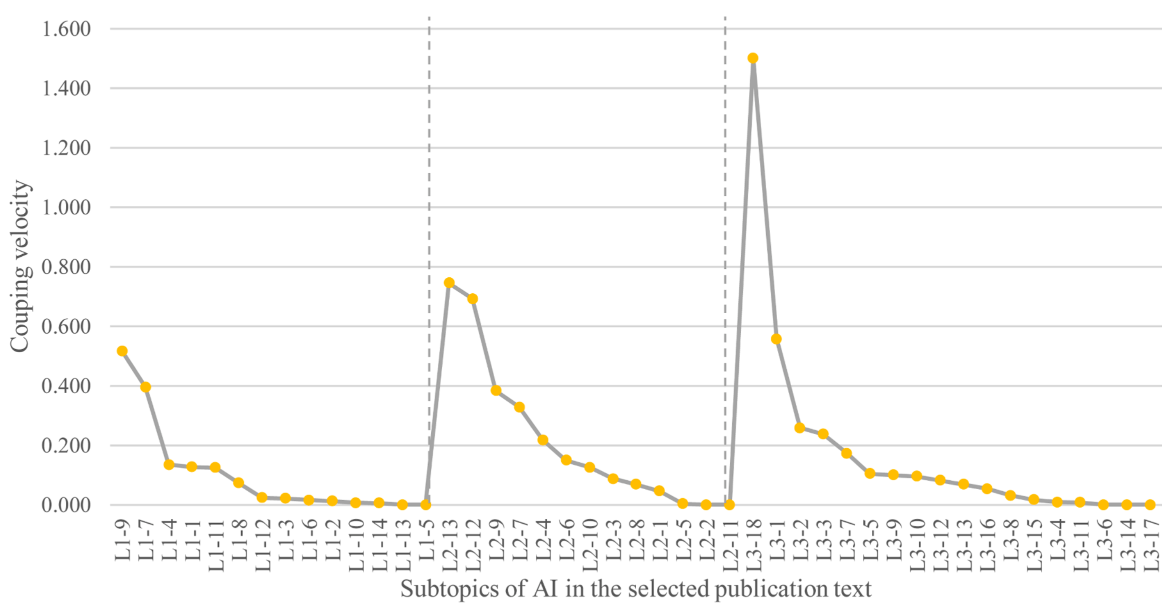

Fig. 9. The coupling velocity of those AI subtopics in the selected publication texts.

the model, and the LDA model is built with the Gensim algorithm package[2] The default parameters in the Gensim package are retained, and the maximum number of iterations is 500.

The experiment includes two processes: optimal topic selection and LDA topic modeling. The first step is to determine the optimal number of topics for each period before running the LDA model. According to the topic number settings chosen in previous studies (Hu et al., 2014), this paper first sets the interval value of the topic classification number to [5, 30] and selects the minimum number of topics in each period that can significantly increase the coherence of the model as the optimal number. The second step is to run the LDA model with the optimal number of topics as the parameter. The results of LDA mainly focus on two aspects: one is the topic distribution $\theta^{(d)}$ of the documents; the other is the distribution $\phi^{(z)}$ of feature words in the chosen topic, which represents the probabilities of feature words in the topic.

According to the output of the LDA model, the topic distribution $\theta^{(d)}$ and feature word distribution $\phi^{(z)}$ of the publication data in each of the three periods can also be obtained, as shown in Table 2, and the relevant results regarding the patent data are shown in Table 3.

In Table 2, the frequencies of the corresponding documents can partly reflect the research fronts in different periods. Each topic is divided into the documents with the highest probabilities, and the top five topics with the highest document frequencies are selected; these can be considered the hot topics in each AI development period.

In addition to the results presented in Table 2 and Table 3, the distances between different subtopics can be calculated based on Eqs. (5) and (6); furthermore, the thickness of the line between every two subtopics is calculated to indicate the connection strength, which is the reciprocal of distance. Then, the results of the connection strength between each pair of subtopics, especially the subtopics in different periods, are shown in Figs. 6 and 7. For information on the topic tags, see

[2] https://pypi.org/project/gensim/

Table A1. Topic distribution in the selected AI publications.

1990–1999			
L_{1-1} Parallel learning training	L_{1-2} Movement neuron cell	L_{1-3} Machine learning and parameter prediction	L_{1-4} Expert system decision
L_{1-5} Optimization design of genetic algorithm	L_{1-6} Classification and classifier	L_{1-7} Associative memory and pattern recognition	L_{1-8} Protein structure sequence
L_{1-9} Adaptive fuzzy control	L_{1-10} Nonlinear dynamic simulation	L_{1-11} Signal sensor	L_{1-12} Robot motion simulator
L_{1-13} Heuristic path planning		L_{1-14} Clinical medicine	
2000–2009			
L_{2-1} Physical properties of materials	L_{2-2} Protein structure sequence prediction	L_{2-3} Spectral analysis of cancer	L_{2-4} Support vector machine classifier
L_{2-5} Cell morphology and spatial distribution	L_{2-6} Quality control system model	L_{2-7} Fuzzy control system	L_{2-8} Optimization design of genetic algorithm
L_{2-9} Sensor fault diagnosis based on neural network	L_{2-10} Cortical neural network	L_{2-11} Integrated speech learning	L_{2-12} Dynamic neural network stability
L_{2-13} Artificial neural network prediction model			
2010–2017			
L_{3-1} Adaptive fuzzy control system	L_{3-2} Text-based cancer detection	L_{3-3} Power supply system design	L_{3-4} Human behavior learning
L_{3-5} Parametric prediction model	L_{3-6} Spectral resolution	L_{3-7} Fuzzy model and fuzzy system	L_{3-8} Wind speed prediction model
L_{3-9} System stability	L_{3-10} Support vector machine classification	L_{3-11} Protein sequence prediction	L_{3-12} Sensor structure design
L_{3-13} Pathological diagnostic analysis	L_{3-14} Molecular descriptor	L_{3-15} Optimization design of genetic algorithm	L_{3-16} Mechanics of materials
L_{3-17} Cortical neuron activity		L_{3-18} Neural network model	

Table A2. Topic distribution in the selected AI patents.

1990–1999			
P_{1-1} Neural network image processing	P_{1-2} Neural network system and parameters	P_{1-3} Expert system	P_{1-4} Neural network learning
P_{1-5} User computer system	P_{1-6} Signal sensor based on neural network		
2000–2009			
P_{2-1} Traffic sensor	P_{2-2} Material sample testing	P_{2-3} Expert system	P_{2-4} Image description
P2–5 Neural network mapping method	P2–6 Neural network pattern recognition	P2–7 Expression of cellular proteins in patients	
2010–2017			
P_{3-1} Vehicle human-computer interaction	P_{3-2} Sensor equipment	P_{3-3} Cell disease test	P_{3-4} User computer system equipment
P_{3-5} Vector drawing method	P_{3-6} Water quality monitoring model	P_{3-7} Modular control device system	P_{3-8} Neural network data transmission
P_{3-9} Signal system and drawing method		P_{3-10} Mapping algorithm	

the Appendix.

Regarding the results presented in Fig. 6, branched topics in the field of AI, such as "m*achine learning and parameter prediction* (L_{1-3})", "*decision-making expert system* (L_{1-4})", "*genetic optimization design* (L_{1-5})", "*protein structure sequence* (L_{1-8})", "*nonlinear dynamic simulation* (L_{1-10})", "*signal sensor* (L_{1-11})" and "*clinical medicine* (L_{1-12})", have relatively stronger connections with the second-period studies than other topics, i.e., these AI subtopics in the first period exhibit improved continuity in the second period. The subtopic "*associative memory and pattern recognition* (L_{1-7})" is far from all the topics in the second period, indicating that this topic may be weakened or fade away to some extent during the second period.

Regarding the connection strengths of the AI subtopics between the second and third periods, the inheritance and dependency relationships of the subdivided themes are clearer than those between the first and second periods. For example, these AI subtopics, such as "*protein sequence prediction* (L_{2-2} & L_{3-11})", "*support vector machine classifier* (L_{2-4} & L_{3-10})", and "*genetic algorithm optimization design* (L_{2-8} & L_{3-15})", all exhibit significant continuity, indicating that these AI subtopics are still popular in artificial intelligence studies.

In addition, in the second period, the subtopic "*cancer spectroscopic analysis* (L_{2-3})" exhibits above-average connection strength with AI subtopics in the third period, including "*spectral resolution* (L_{3-7})", " *sensor structure design* (L_{3-12})" and "*pathological diagnosis analysis* (L_{3-13})". Similarly, the subtopic "*cell morphology and spatial distribution* (L_{2-5})" gives rise to subtopics in the third period, such as "*human behavioral learning* (L_{3-4})", "*fuzzy models and fuzzy systems* (L_{3-7})" and "*pathological diagnostic analysis* (L_{3-13})". Furthermore, the subtopic "*cortical neuron network* (L_{2-10}) " in the second period can branch to issues such as "*human behavioral learning* (L_{3-4})", "*sensor structure design* (L_{3-12}) " and "*cortical neuron activity* (L_{3-17})". In addition, the subtopic "*adaptive fuzzy control* (L_{1-9})" appears to encounter high-low-high volatility or oscillating processes. Therefore, the presented information in Figs. 6 and 7 not only shows the similarities or connections but can also partly depict the evolutionary connections between subtopics and even the branching processes among different AI subtopics in the different periods.

Based on the computational results of the distances between different subtopics in the AI patent texts, some subtopics also exhibit above-average connection strength. For example, some AI subtopics in the first period, such as "*neural network system and parameters* (P_{1-2})", "*expert system* (P_{1-3})", "*user computer system* (P_{1-5})", and "*signal sensor based on neural network* (P_{1-6})", have improved continuity in the second period. In particular, the relationship between subtopic "*neural network learning* (P_{1-4})" and subtopic "*neural network mapping method* (P_{2-5})" is much closer than average.

Regarding the relationships between the second period and the third period, the subtopic "*expert system* (P_{2-3})" seems to play an important role in binding the different subtopics between periods. For example, the connection between "*expert system* (P_{2-3})" and "*user computer system equipment* (P_{3-4})" exhibits the highest strength, and the subtopic "*modular control device system* (P3–7)" also has above-average connection strength with the subtopic "*expert system* (P_{2-3})". In addition, the subtopics "*vehicle human-computer interaction* (P_{3-1})" and "*neural network transmission* (P_{3-8})" are not closely related to all of the patent topics in the second period, so the relevant issues with respect to P_{3-1} and P_{3-8} may be emerging technologies in the field of AI.

The continuity or inheritance of knowledge production related to AI studies and engineering applications can be partially revealed in Fig. 5 and Fig. 6; however, the differences between AI studies and AI applications can also be partly found. Furthermore, coupling analyses between AI studies and AI engineering applications can provide additional technological intelligence, so the corresponding work is conducted in the following section.

4.4. Coupling analyses between publications and patents regarding AI

To further explore the possible or underlying coupling relationships between the selected AI publications and AI patents, two indicators, coupling strength and coupling velocity, were introduced above. Based on formulas (8) and (9), the computational results are shown in Figs. 8

Table A3. Distances of two subtopics between period #1 and period #2 (AI publications)

	L_{2-1}	L_{2-2}	L_{2-3}	L_{2-4}	L_{2-5}	L_{2-6}	L_{2-7}	L_{2-8}	L_{2-9}	L_{2-10}	L_{2-11}	L_{2-12}	L_{2-13}
L_{1-1}	0.475	0.497	0.436	0.505	0.417	0.465	0.486	0.575	0.471	0.436	0.533	0.532	0.584
L_{1-2}	0.364	0.386	0.325	0.394	0.282	0.354	0.375	0.464	0.330	0.229	0.422	0.323	0.323
L_{1-3}	0.219	0.205	0.160	0.249	0.161	0.209	0.230	0.319	0.215	0.180	0.277	0.276	0.328
L_{1-4}	0.282	0.304	0.243	0.312	0.224	0.272	0.253	0.382	0.278	0.243	0.340	0.339	0.391
L_{1-5}	0.224	0.246	0.185	0.254	0.166	0.214	0.235	0.225	0.220	0.185	0.282	0.281	0.333
L_{1-6}	0.336	0.358	0.279	0.176	0.278	0.326	0.347	0.436	0.332	0.297	0.394	0.393	0.445
L_{1-7}	0.659	0.715	0.654	0.723	0.635	0.683	0.704	0.793	0.689	0.654	0.751	0.750	0.802
L_{1-8}	0.245	0.149	0.206	0.227	0.187	0.215	0.256	0.345	0.241	0.206	0.303	0.302	0.354
L_{1-9}	0.594	0.616	0.555	0.624	0.536	0.584	0.457	0.694	0.59	0.555	0.652	0.651	0.703
L_{1-10}	0.281	0.303	0.242	0.311	0.223	0.271	0.292	0.381	0.241	0.242	0.339	0.308	0.390
L_{1-11}	0.164	0.242	0.181	0.250	0.162	0.21	0.231	0.320	0.174	0.181	0.278	0.277	0.329
L_{1-12}	0.315	0.277	0.276	0.345	0.257	0.305	0.326	0.415	0.311	0.276	0.373	0.372	0.424
L_{1-13}	0.420	0.442	0.381	0.450	0.362	0.382	0.431	0.520	0.416	0.381	0.478	0.477	0.529
L_{1-14}	0.247	0.269	0.184	0.277	0.189	0.237	0.258	0.347	0.243	0.208	0.305	0.304	0.356

Table A4. Distances of two subtopics between period #2 and period #3 (AI publications)

	L_{2-1}	L_{2-2}	L_{2-3}	L_{2-4}	L_{2-5}	L_{2-6}	L_{2-7}	L_{2-8}	L_{2-9}	L_{2-10}	L_{2-11}	L_{2-12}	L_{2-13}
L_{3-1}	0.369	0.391	0.330	0.399	0.311	0.315	0.180	0.469	0.309	0.330	0.427	0.402	0.478
L_{3-2}	0.370	0.392	0.331	0.352	0.298	0.360	0.381	0.470	0.366	0.331	0.428	0.427	0.479
L_{3-3}	0.282	0.304	0.243	0.312	0.224	0.206	0.193	0.382	0.186	0.243	0.340	0.315	0.391
L_{3-4}	0.204	0.226	0.165	0.234	0.146	0.194	0.215	0.304	0.200	0.143	0.262	0.261	0.313
L_{3-5}	0.240	0.230	0.173	0.236	0.182	0.198	0.251	0.340	0.236	0.201	0.298	0.297	0.159
L_{3-6}	0.266	0.288	0.227	0.296	0.160	0.256	0.277	0.366	0.262	0.227	0.324	0.323	0.375
L_{3-7}	0.212	0.234	0.145	0.210	0.154	0.146	0.161	0.312	0.170	0.173	0.27	0.245	0.251
L_{3-8}	0.307	0.277	0.268	0.337	0.249	0.265	0.318	0.407	0.303	0.268	0.365	0.364	0.276
L_{3-9}	0.317	0.339	0.278	0.347	0.259	0.265	0.286	0.417	0.271	0.278	0.375	0.324	0.426
L_{3-10}	0.291	0.313	0.252	0.117	0.233	0.281	0.302	0.351	0.287	0.252	0.349	0.348	0.400
L_{3-11}	0.278	0.146	0.239	0.308	0.22	0.268	0.289	0.378	0.274	0.239	0.336	0.335	0.357
L_{3-12}	0.195	0.195	0.136	0.185	0.137	0.185	0.206	0.233	0.191	0.156	0.253	0.252	0.304
L_{3-13}	0.198	0.220	0.111	0.228	0.140	0.162	0.183	0.298	0.168	0.159	0.256	0.231	0.307
L_{3-14}	0.305	0.207	0.266	0.335	0.247	0.295	0.316	0.405	0.301	0.266	0.363	0.362	0.414
L_{3-15}	0.328	0.350	0.289	0.358	0.270	0.318	0.339	0.072	0.324	0.289	0.386	0.385	0.437
L_{3-16}	0.194	0.298	0.237	0.306	0.218	0.226	0.287	0.376	0.272	0.237	0.334	0.333	0.385
L_{3-17}	0.220	0.242	0.181	0.250	0.162	0.210	0.231	0.320	0.186	0.111	0.278	0.205	0.289
L_{3-18}	0.458	0.48	0.419	0.488	0.400	0.448	0.469	0.558	0.382	0.359	0.516	0.241	0.263

Table A5. Distances of two subtopics between period #1 and period #2 (AI patents)

	P_{2-1}	P_{2-2}	P_{2-3}	P_{2-4}	P_{2-5}	P_{2-6}	P_{2-7}
P_{1-1}	0.260	0.254	0.326	0.250	0.216	0.250	0.292
P_{1-2}	0.195	0.205	0.213	0.267	0.193	0.205	0.243
P_{1-3}	0.212	0.190	0.146	0.216	0.264	0.206	0.228
P_{1-4}	0.295	0.233	0.261	0.259	0.107	0.145	0.271
P_{1-5}	0.178	0.156	0.080	0.198	0.230	0.226	0.194
P_{1-6}	0.193	0.163	0.235	0.225	0.165	0.161	0.201

Table A6. Distances of two subtopics between period #2 and period #3 (AI patents)

	P_{2-1}	P_{2-2}	P_{2-3}	P_{2-4}	P_{2-5}	P_{2-6}	P_{2-7}
P_{3-1}	0.366	0.304	0.376	0.366	0.378	0.374	0.342
P_{3-1}	0.193	0.163	0.255	0.245	0.257	0.253	0.221
P_{3-3}	0.269	0.207	0.279	0.269	0.281	0.277	0.171
P_{3-4}	0.177	0.195	0.115	0.195	0.231	0.227	0.233
P_{3-5}	0.318	0.256	0.328	0.168	0.258	0.232	0.294
P_{3-6}	0.331	0.247	0.297	0.295	0.343	0.285	0.307
P_{3-7}	0.269	0.279	0.247	0.341	0.353	0.349	0.317
P_{3-8}	0.457	0.395	0.427	0.421	0.269	0.351	0.433
P_{3-9}	0.182	0.200	0.248	0.188	0.202	0.228	0.238
P_{3-10}	0.283	0.221	0.293	0.209	0.223	0.249	0.259

and 9.

In general, the coupling strengths of the subtopics of AI publications with the relevant patent subtopics exhibit dynamic and fluctuating trends, even those publication subtopics that maintain weak connections with patent subtopics. Since 2000, the R&D of AI technologies and applications has become more rapid than before; in particular, in the past five years, an increasing number of patents on AI technologies have emerged.

Furthermore, a time lag with respect to the technology transformation can also be observed; for example, the publication subtopic L_{1-5} can hardly be observed in the patents developed during the same period, and the coupling strength of L_{1-5} is almost zero in the first period; however, the coupling strength of the similar subtopic L_{2-8} or L_{3-15} with L_{1-5} is significantly larger than zero in the second and third periods, respectively.

In the third period, some publication subtopics exhibit significant coupling relationships with patent subtopics; for example, "*neural network model* (L_{3-18})", "*adaptive fuzzy control system* (L_{3-1})", "*cancer detection based on texture* (L_{3-2})" and "*power system design* (L_{3-3})" have high coupling strengths with patent texts, which means that these AI subtopics achieve relatively successful technological transformations. Additionally, some AI subtopics, such as "*sensor structure design* (L_{3-12})", "*pathological diagnosis analysis* (L_{3-13})", "*wind speed prediction model* (L_{3-8})" and "*material mechanics* (L_{3-16})", can be observed, and their coupling relationships with patent subtopics should be further mined to a degree for technological opportunities. Some of the possible emerging topics, including "*cortical neuronal activity* (L_{3-17})", "*molecular descriptors* (L_{3-14})", "*spectral resolution* (L_{3-6})", and "*human behavior learning* (L_{3-4})", have low degrees of patent transformation or are even blank; therefore, in the upcoming years, these AI subtopics may require additional attention regarding technology transformation and opportunity discovery.

While considering the factor of time, the indicator of coupling strength can be integrated into the indicator of coupling velocity, and the coupling velocities of all AI subtopics in the selected publication

Table 1
Top 50 most frequent terms in the selected publication text.

No	Term	Publications	Patents	No	Term	Publications	Patents	No	Term	Publications	Patents
1	neural network	140,065	16,073	19	pattern recognition	9929	802	37	exponential stability	4337	1
2	genetic algorithm	67,378	3738	20	sensor	9483	2823	38	wavelet transform	4243	127
3	artificial neural network	44,553	2182	21	image processing	8662	3534	30	backpropagation	4114	37
4	support vector machine	32,136	3820	22	protein	8602	366	40	multilayer perceptron	3976	47
5	identification	21,097	2515	23	particle swarm optimization	8126	169	41	fault diagnosis	3619	436
6	signal	20,731	5265	24	adaptive control	7701	385	42	cellular neural network	3196	213
7	classifier	18,198	1674	25	disease	7406	643	43	differential evolution	2917	26
8	machine learning	17,595	1345	26	global optimization	7301	154	44	climate	2664	36
9	nonlinear system	13,479	254	27	feature extraction	6954	617	45	visual cortex	2619	5
10	temperature	13,066	1333	28	fuzzy logic	6210	296	46	deep learning	2531	565
11	neuron	12,817	1454	29	principal component analysis	6010	214	47	associative memory	2338	93
12	power system	12,579	1625	30	multiobjective optimization	5847	44	48	random forest	2207	48
13	artificial intelligence	12,341	3642	31	cancer	5672	366	49	convolutional neural network	1924	1476
14	expert system	12,093	2384	32	recurrent neural network	5523	351	50	mobile robot	1362	113
15	database	11,602	2266	33	descriptor	4967	154				
16	evolutionary algorithm	11,584	100	34	spectroscopy	4757	106				
17	image classification	10,900	1048	35	data mining	4679	306				
18	brain	10,045	408	36	time varying delay	4542	12				

Table 2
The most popular subtopics in AI publications.

1990–1999			2000–2009			2010–2017		
Topic code	Publications	Topic tag	Topic code	Publications	Topic tag	Topic code	Publications	Topic tag
L_{1-6}	20,957	Classification and classifier	L_{2-13}	57,945	Artificial neural network prediction model	L_{3-7}	79,750	Fuzzy model and fuzzy system
L_{1-7}	19,345	Associative memory and pattern recognition	L_{2-12}	51,028	Dynamic neural network stability	L_{3-18}	67,900	Neural network model
L_{1-10}	13,698	Nonlinear dynamic simulation	L_{2-8}	48,661	Optimization design of the genetic algorithm	L_{3-5}	64,369	Parametric prediction model
L_{1-14}	13,200	Clinical medicine	L_{2-2}	48,181	Protein structure sequence prediction	L_{3-10}	63,070	Support vector machine classification
L_{1-9}	12,593	Adaptive fuzzy control	L_{2-3}	45,952	Spectral analysis of cancer	L_{3-15}	60.883	Optimization design of the genetic algorithm

Table 3
The most popular subtopics in the AI patents.

1990–1999			2000–2009			2010–2017		
Topic code	Publications	Topic tag	Topic code	Publications	Topic tag	Topic code	Publications	Topic tag
P_{1-4}	2001	Neural network learning	P_{2-5}	5494	Neural network mapping method	P_{3-8}	12,276	Data transmission of neural network
P_{1-1}	1731	Neural network image processing	P_{2-3}	4363	Expert system	P_{3-10}	11,979	Mapping algorithm
P_{1-6}	1686	Signal sensor based on neural network	P_{2-6}	3787	Neural network pattern recognition	P_{3-5}	11,503	Vector drawing method
P_{1-5}	1369	User computer system	P_{2-4}	3695	Image processing	P_{3-7}	10,980	Modular control device system
P_{1-2}	1318	Neural network system and parameters	P_{2-1}	3589	Traffic sensor	P_{3-9}	10,724	Signal system and drawing method

texts are computed; the results are shown in Fig. 9.

Comparing Fig. 9 with Fig. 8, although these two curves look very similar, some differences are still significant because the local ranking orders change. For example, some subtopics of AI publications, such as L_{1-11}, L_{2-8} and L_{3-10}, have lower coupling velocity ranks than their ranks of coupling strength. In contrast, other subtopics of AI publications, such as L_{1-4}, L_{2-10} and L_{3-5}, exhibit faster coupling velocities than their ranks of coupling strength.

Regarding the two indicators of CS and CV, some interesting phenomena are uncovered. Some subtopics of AI publications exhibit high CSs and CVs, e.g., L_{1-9}, L_{1-7}, L_{2-13}, L_{3-18} and L_{3-1}, which can mean that these relevant issues on AI have broad applications, so the corresponding technology transformation also becomes very popular during the same period and in neighboring periods. Conversely, other subtopics in the selected AI publications exhibit lower CSs and CVs, e.g., L_{1-13}, L_{2-5} and L_{3-17}, which can imply that these issues in AI studies have difficulty achieving technological transformations or that their applications are not clear or have low value.

Based on the coupling analyses of AI publications and patents, many relevant subtopics, such as "*neural network*", "*fuzzy control*", "*support vector machine*", "*genetic algorithm*" and "*adaptive control system*", are not only popular in scientific research but are also very popular with respect to technological development. Furthermore, such relevant issues of AI studies, including "*protein sequence prediction*", "*pathological diagnosis*", "*spectrum analysis*", and "*human brain cortex neurons*", can be popular in scientific studies, but the corresponding patents are below average in terms of coupling; therefore, these subtopics either have much room for technological transformation or could face more challenges than other subtopics in the process of application development.

To verify the above results based on text mining and topic modeling, we interviewed several domain experts on AI, and positive feedback was the main factor. However, some experts noted that the knowledge boundary and the connotation of AI remain controversial, and the search queries related to AI publications and patents also have room for refinement. For example, one expert argued that if the term or the topic of 'fuzzy control' can be attributed to AI, additional terms or topics can also be included; therefore, the results of coupling analyses between AI publications and AI patents should be further discussed with more domain experts. In general, the coupling analysis technique for publications and patents can have the potential to identify the TE of a specific topic.

5. Conclusions and discussions

In terms of the mainstream studies on TE, TOA and TF (technological forecasting), the relevant publications and patents related to a specific topic are still the main data sources used. In relative terms, patent analysis and relevant text mining with patent data were more popular in prior studies (Wang et al., 2015; Zhang et al., 2016; Suominen et al., 2017).

Actually, the interactions between scientific research and technological development can also bring valuable intelligence for S&T policymaking and the strategic planning of firms (Daim et al., 2006; Suominen et al., 2019). Therefore, finding the latent connections between publications and patents within a specific topic is still valuable. To enrich the relevant methods for performing TE or TOA, an integrated framework based on several new indicators and computing methods on the coupling analysis between publications and patents is proposed to explore the TE or TOA in the AI field.

Essentially, accurately identifying TE or TO is still one of the challenges in the study of technology management and policymaking in STI (science, technology and innovation). The latest studies have begun to describe a promising approach that focuses on the underlying connections or coupling relationships between publications and patents within a specific topic. Although the proposed indicators and computing techniques could be controversial and have room for improvement, they could still be helpful for future work on TE and TOA.

Regarding the TE and TOA related to AI, corresponding studies are still very rare for several reasons: (1) intuitively, AI seems to be a scientific topic rather than a technical topic; (2) the relevant knowledge on AI has diffused in many different areas of S&T; and (3) the knowledge and application boundaries of AI have become blurred and dynamic. However, with the rapid growth of publications and patents related to AI, probing or capturing the relevant TE or TO on AI could also be valuable for knowledge communication and policymaking.

In terms of the empirical study, some popular subtopics of AI publications and patents were partly revealed in this paper. In particular, the dynamic trends of subtopics in the different periods were also partly observed based on the indicators of the coupling strength and coupling velocity. Therefore, in addition to the proposed indicators and computing methods for coupling analyses between publications and patents within a specific topic, a case study on AI can also be helpful for related technology forecasting or policymaking.

In summary, to further enrich the relevant research methodology on TE and TOA, several new indicators and computing techniques for coupling analysis between publications and patents are proposed and applied to the AI field. Therefore, the marginal contributions of this paper can be attributed to the following aspects: (1) further enriching the analytical methods for TE, TOA, or TF; (2) introducing a new perspective, several new indicators and computing formulas for coupling analyses between publications and patents within a specific topic; and (3) presenting the topic evolution in the AI field, in particular, the dynamic change from academic studies to technology development during the past decades; and then the underlying implications of policy for AI development could be meaningful.

Certainly, this paper also has some limitations. Considering the search formula (Boolean operation formula) related to AI, although a hybrid query was introduced, there is still much room for optimization. The connotation and knowledge boundary of AI still need to be further explored and clarified. Furthermore, the proposed indicators of the coupling strength and coupling velocity could be controversial, and the analytical framework proposed in this paper also needs additional cases to verify its generalization capability in future studies on TE, TOA or TF.

The work has not been published previously, and it is not under consideration for publication elsewhere.

CRediT authorship contribution statement

Munan Li: Conceptualization, Methodology. **Wenshu Wang:** Data curation, Visualization, Software. **Keyu Zhou:** Data curation, Software, Visualization.

Declaration of Competing Interest

There is no conflict of interest.

Acknowledgement

A part of the idea for this manuscript was reported at the 9th Global TechMining Conference, Atlanta, USA. We appreciate the suggestions from the audience, which were very helpful for us to complete and improve this paper. The work is partly supported by the National Natural Science Foundation of China (72074081 & 71673088), the Natural Science Foundation of Guangdong Province (2020A1515010438), the Foundation of Guangdong Soft Science (2019B101001002), and the Fundamental Research Funds for the Central Universities (ZDPY201903).

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.techfore.2021.121064.

Appendix

Table A1, A2, A3, A4, A5, A6

References

Abbas, N.N., Ahmed, T., Shah, S.H.U., et al., 2019. Investigating the applications of artificial intelligence in cyber security. Scientometrics 121 (2), 1189–1211.
Agrawal, A., Gans, J.S., Goldfarb, A., 2019. Exploring the impact of artificial Intelligence: prediction versus judgment. Inf. Econ. Policy 47, 1–6.
Aletras, N., Baldwin, T., Lau, J.H., et al., 2017. Evaluating topic representations for exploring document collections. J. Assoc. Inf. Sci. Technol. 68 (1), 154–167.
Aristodemou, L., Tietze, F., 2018. The State-of-the-Art on Intellectual Property Analytics (IPA): a Literature Review on Artificial Intelligence. Machine Learning and Deep Learning Methods for Analysing Intellectual Property (IP) Data. World Patent Information 55 (12), 37–51, 10.1016/j.wpi.2018.07.002.

Arora, S., Rong, G., Halpern, Y., et al., 2018. Learning Topic Models Provably and Efficiently. Commun. ACM 61 (4), 85–93.
Avila-Robinson, A., Miyazaki, K., 2013. Dynamics of scientific knowledge bases as proxies for discerning technological emergence - The case of MEMS/NEMS technologies. Technolog. Forecast.. Soc. Change 80 (6), 1071–1084.
Bastani, K., Namavari, H., Shaffer, J., 2019. Latent Dirichlet allocation (LDA) for topic modeling of the CFPB consumer complaints. Expert Syst. Appl. 127, 256–271.
Baumer, E.P.S., Mimno, D., Guha, S., et al., 2017. Comparing grounded theory and topic modeling: extreme divergence or unlikely convergence? Journal of the Association for Information Science & Technology 68 (6), 1397–1410.
Bench-Capon, T.J.M., Dunne, P.E., 2007. Argumentation in artificial intelligence. Artif. Intell. 171 (10), 619–641.
Borges, A.F.S., Laurindo, F.J.B., Spinola, M.M., et al., 2021. The strategic use of artificial intelligence in the digital era: systematic literature review and future research directions. Int. J. Inf. Manag. 57 (4), 102225.
Bruck, P., Rethy, I., Szente, J., et al., 2016. Recognition of emerging technology trends: class-selective study of citations in the US Patent Citation Network. Scientometrics 107 (3), 1465–1475.
Brunette, E.S., Flemmer, R.C., Flemmer, C.L., 2009. A Review of Artificial Intelligence. ICARA 2009 - Proceedings of the 4th International Conference on Autonomous Robots and Agents 385–392, 10.1109/ICARA.2000.4804025.
Burmaoglu, S., Sartenaer, O., Porter, A., 2019. Conceptual definition of technology emergence: a long journey from philosophy of science to science policy. Technol. Soc. 59, 101126.
Blei, D.M., 2012. Probabilistic Topic Models. Commun. ACM 55 (4), 77–84.
Blei, D.M., Ng, A., Jordan, M., 2003. Latent Dirichlet Allocation. J. Mach. Learn. Res. 3, 993–1022.
Carley, S.F., Newman, N.C., Porter, A.L., et al., 2018. An indicator of technical emergence. Scientometrics 115 (1), 35–49.
Daim, T.U., Rueda, G., Martin, H., Gerdsri, P., 2006. Forecasting emerging technologies: use of bibliometrics and patent analysis. Technol. Forecas. Soc. Change 73 (8), 981–1012.
Daugherty, P.R., Wilson, H.J., Chowdhury, R., 2019. Using Artificial Intelligence to Promote Diversity. MIT Sloan Management Review 60 (2), 10–12.
Gosavi, A., 2009. Reinforcement Learning: a Tutorial Survey and Recent Advances. INFORMS J. Comput. 21 (2), 178–192.
Guo, Y., Ma, T., Porter, A.L., et al., 2012. Text mining of information resources to inform Forecasting Innovation Pathways. Technol. Analy. Strat. Manag. 24 (8), 843–861.
Guo, Y., Barnes, S.J., Jia, Q., 2017. Mining meaning from online ratings and reviews: tourist satisfaction analysis using Latent Dirichlet Allocation. Tourism Manag. 59, 467–483.
Hu, Y., Boyd-Graber, J., Satinoff, B., et al., 2014. Interactive topic modeling. Mach. Learning 95 (3), 423–469.
Huang, Y., Schuehle, J., Porter, Alan, L., et al., 2015. A systematic method to create search strategies for emerging technologies based on the Web of Science: illustrated for 'Big Data. Scientometrics 105 (3), 2005–2022.
Huang, Y., Zhu, D.H., Qian, Y., et al., 2017. A hybrid method to trace technology evolution pathways: a case study of 3D printing. Scientometrics 111 (1), 185–204.
Klinger, J., Mateos-Garcia, J., Stathoulopoulos, K., 2021. Deep learning, deep change? Mapping the evolution and geography of a general purpose technology. Scientometrics 126 (7), 5589–5621.
Kwon, S., Liu, X., Porter, A.L., et al., 2019. Research addressing emerging technological ideas has greater scientific impact. Res. Policy 48 (9), 103834.
LeCun, Y., Bengio, Y., Hinton, G., 2015. Deep Learning. Nature 521 (7553), 436–444.
Li, M.N., 2015. A novel three-dimension perspective to explore technology evolution. Scientometrics 105 (3), 1679–1697.
Li, M.N., Porter, A.L., Suominen, A., 2018. Insights into relationships between disruptive technology/innovation and emerging technology: a bibliometric perspective. Technol. Forecast. Soc. Change 129, 285–296.
Li, M.N., Porter, A.L., Suominen, A., Burmaoglu, S., et al., 2021. An exploratory perspective to measure the emergence degree for a specific technology based on the philosophy of swarm intelligence. Technolog. Forecast. Soc. Change 166, 120621.
Li, X., Xie, Q., Jiang, J., et al., 2019. Identifying and monitoring the development trends of emerging technologies using patent analysis and Twitter data mining: the case of perovskite solar cell technology. Technol. Forecast. Soc. Change 146, 687–705.
Love-Koh, James, et al., Peel, Alison, Rejon-Parrilla, Juan Carlos, 2018. The Future of Precision Medicine: Potential Impacts for Health Technology Assessment. Pharmacoeconomics 36 (12), 1439–1451.
Ma, T., Porter, A.L., Guo, Y., et al., 2014. A technology opportunities analysis model: applied to dye-sensitised solar cells for China. Technol. Analy. Strat. Manag. 26 (1), 87–104.
Mimno, D.M., Wallach, H.M., Talley, E.M., et al., 2011. Optimizing Semantic Coherence in Topic Models. Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing 262–272.
Porter, A.L., 1999. Tech forecasting an empirical perspective. Technol. Forecast. Soc. Change 62 (1), 19–28.
Porter, A.L., Garner, J., Carley, S.F., et al., 2019. Emergence scoring to identify frontier R&D topics and key players. Technol. Forecast. Soc. Change 146, 628–643.
Porter, A.L., Newman, N.C., 2011. Mining external R&D. Technovation 31 (4), 171–176.
Rotolo, D., Hicks, D., Martin, B.R., 2015. What is an emerging technology? Res. Policy 44 (10), 1827–1843.
Shin, H.C., Lu, L., Kim, L., et al., 2017. Interleaved text/image deep mining on a large-scale radiology database for automated image interpretation. J. Mach. Learn. Res. 17 (1), 3729–3759.
Silver, D., Huang, A., Maddison, C.J., et al., 2016. Mastering the game of Go with deep neural networks and tree search. Nature 529 (7587), 484.
Simon, H.A., Munankata, T., 1997. AI Lessons. Commun. ACM 40 (8), 23–25.
Stevens, K., Kegelmeyer, P., Andrzejewski, D., et al., 2012. Exploring Topic Coherence over many models and many topics. Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning 952–961.
Suominen, A., Ranaei, S., Dedehayir, O., 2019. Exploration of Science and Technology Interaction: a Case Study on Taxol. IEEE Transac. Eng. Manag., Early Access 1–16. https://doi.org/10.1109/TEM.2019.2923634.
Suominen, A., Toivanen, H., 2016. Map of Science with Topic Modeling: comparison of Unsupervised Learning and Human-Assigned Subject Classification. J. Assoc. Inform. Sci. Technol. 67 (10), 2464–2476.
Suominen, A., Toivanen, H., Seppanen, M., 2017. Firms' knowledge profiles: mapping patent data with unsupervised learning. Technol. Forecast. Soc. Change 115, 131–142.
Venugopalan, S., Rai, V., 2015. Topic based classification and pattern identification in patents. Technol. Forecast. Soc. Change 94, 236–250.
Wang, X.F., Ma, P., Huang, Y., et al., 2017. Combining SAO semantic analysis and morphology analysis to identify technology opportunities. Scientometrics 111 (1), 3–24.
Wang, X.F., Qiu, P.J., Zhu, D.H., et al., 2015. Identification of technology development trends based on subject-action-object analysis: the case of dye-sensitized solar cells. Technol. Forecast. Soc. Change 98, 24–46.
Wang, X.F., Yang, X.M., Wang, X.L., et al., 2020. Evaluating the competitiveness of enterprise's technology based on LDA topic model. Technol. Analys. Strat. Manag. 32 (2), 208–222.
Yoon, B., Park, I., Coh, B.Y., 2014. Exploring technological opportunities by linking technology and products: application of morphology analysis and text mining. Technol. Forecast. Soc. Change 86, 287–303.
Yoon, J., Kim, K., 2012. Detecting signals of new technological opportunities using semantic patent analysis and outlier detection. Scientometrics 90 (2), 445–461.
Zhang, Y., Robinson, D.K.R., Porter, A.L., et al., 2016. Technology roadmapping for competitive technical intelligence. Technol. Forecast. Soc. Change 110, 175–186.
Zhou, Y., Dong, F., Kong, D., et al., 2019. Unfolding the convergence process of scientific knowledge for the early identification of emerging technologies. Technol. Forecast. and Soc. Change 144, 205–220.

Munan Li is Professor of Business School at South China University of Technology in China. He obtained his Ph.D. in computer science and two-year postdoctoral experience in management science from South China University of Technology. His-research area of interest includes the management of emerging technology, disruptive innovation, bibliometrics analysis and text mining. His-work has been published by such journals as the Technological Forecasting & Social Change, Journal of Information Science, Technology Analysis & Strategic Management and the Scientometrics etc. Email: limn@scut.edu.cn.

Wenshu Wang is Candidate of Master degree at School of Business Administration, South China University of Technology in China. Her research interest includes the technology emergence, emerging technology, bibliometric analysis. Email: 464,894,229@qq.com.

Keyu Zhou is Candidate of Master degree at School of Business Administration, South China University of Technology in China. His-research interest includes the emerging technology, bibliometric analysis and text mining. Email: zkeyu2018@163.com.