Research Progress on Emerging Technology Topic Identification Methods: Postprint

Authors: Liu Xiaoling, Tan Zongying

Date: 2023-04-01T16:15:55+00:00

Abstract

[Purpose/Significance] Emerging technology topic identification not only helps to timely track technology development dynamics, but also enables early capture of future development opportunities and potential change trends in technology fields. Reviewing quantitative research methods for emerging technology topic identification and comparing their advantages and disadvantages can provide references for the improvement and perfection of emerging technology topic identification methods. [Method/Process] First, analyze and differentiate the connotations of concepts such as "emerging technology" and "emerging technology topic identification"; then investigate and systematically review qualitative and quantitative research methods for emerging technology topic identification both domestically and internationally, focusing on quantitative research methods primarily based on bibliometrics and data mining, and classify them into three categories: keyword or literature statistical methods, citation network clustering methods, and text mining analysis methods; finally, comprehensively analyze the similarities, differences, and existing defects of various research methods in aspects such as technology topic extraction, construction of emerging technology topic identification indicator systems, and method effectiveness validation, as well as preliminary thoughts on method improvement. [Results/Conclusions] The three categories of methods each have their own characteristics, advantages, and disadvantages in the main steps of emerging technology topic identification, all having room for further improvement. Future research can explore using deep learning and other technologies for accurate extraction of technology topics, construct more comprehensive and systematic emerging technology topic identification indicator systems, and conduct more rigorous method effectiveness validation based on machine learning.

Full Text

Abstract

Identification of emerging technology topics not only facilitates timely tracking of technological development dynamics but also enables early capture of future development opportunities and potential trends. This paper reviews quantitative research methods for emerging technology topic identification and compares their advantages and disadvantages to provide references for methodological improvement and refinement. First, the connotations of concepts such as "emerging technology" and "emerging technology topic identification" are analyzed and distinguished. Then, qualitative and quantitative research methods for emerging technology topic identification at home and abroad are investigated and systematically reviewed, with particular focus on quantitative methods based on bibliometrics and data mining. These methods are categorized into three types: keyword/document statistical methods, citation network clustering methods, and text mining analysis methods. Finally, the paper comprehensively analyzes the similarities, differences, and deficiencies of these methods in key steps including technology topic extraction, construction of emerging technology topic identification indicator systems, and method effectiveness validation, along with preliminary thoughts on methodological improvements. The three categories of methods each have distinct characteristics, advantages, and disadvantages in the main steps of emerging technology topic identification, and all have room for further improvement. Future research could explore using deep learning and other techniques for accurate technology topic extraction, construct more comprehensive and systematic indicator systems for emerging technology topic identification, and conduct more rigorous validation based on machine learning.

Keywords: emerging technology; topic identification; methodology research; bibliometrics; text mining

2 Related Concepts

2.1 "Emerging Technology" and Related Concepts

In the mid-1990s, the "Emerging Technology Management Research Program" at the Wharton School of the University of Pennsylvania first proposed the concept of "Emerging Technology." Wharton researchers defined emerging technology as "innovation based on science that may create a new industry or transform an existing one," encompassing not only technologies arising from radical innovation—such as biopharmaceuticals, digital imaging, high-temperature superconductors, microrobotics, and laptop computers—but also technologies that become more innovative through the integration of previously independent research outcomes, such as MRI, fax machines, electronic finance, and internet technologies [1]. Following this concept's introduction, domestic researchers including Lu Ruoyu [2], Li Shiming [3], and Xu Jianguo [4], as well as international scholars such as S. Cozzens [5] and A. Breitzman et al. [6], provided various definitions (see Table). For instance, Lu Ruoyu et al. [2] viewed emerging technology as "a new concept, method, or invention based on science that can create a new industry or transform an existing one, thereby exerting significant influence on economic structure." S. Cozzens et al. [5], based on analysis of nearly 2,000 articles on emerging technologies, defined emerging technology as "technology that is rapidly growing, emerging, possesses unexploited market potential, and is based on high technology. Such technology has enormous potential but has not yet demonstrated value or achieved industry consensus." They identified four characteristics: rapid growth; transformation into something new during the process; unexploited market or economic potential; and increasingly close connection with scientific research. In summary, emerging technology is a newly emerged, rapidly developing technology typically based on high technology that may open new technological and scientific fields, possesses enormous market potential, and may create a new industry or transform an existing one, while still carrying uncertainty at its current stage.

"Emerging technology" shares similarities with but differs from concepts such as "emerging research field," "disruptive technology," and "frontier technology." An "emerging research field" generally refers to new scientific research focused on exploring new scientific questions and theoretical studies. Compared to emerging technology, it has lower requirements for market application and may have smaller socioeconomic impact. However, with the rapid development of science and technology, the boundary between science and technology has become increasingly blurred, leading some studies to not strictly distinguish between "emerging research field" and "emerging technology," with empirical analyses typically covering basic research, applied research, and technological research. C. M. Christensen [9] noted that when new technological innovation overthrows existing dominant technologies in the market, it is called "disruptive technology." Frontier technology refers to major technologies in high-tech fields that are forward-looking, pioneering, and exploratory—important foundations for future high-tech updates and emerging industry development. Compared with disruptive and frontier technologies, emerging technology carries higher uncertainty, and its commercial value remains only potential and not yet fully realized.

2.2 "Emerging Technology Topic Identification" and Related Concepts

Technology topics lack clear definition, with different studies interpreting them variously according to their research purposes and questions, typically referring to branch technology fields, technology directions, or technology problems within a technical domain, with varying levels of granularity. In existing quantitative research, technology topics are generally revealed through a set of keywords/phrases or a collection of papers or patents [10]. Patent literature serves as an important information source for technology analysis, integrating technical, legal, and economic information with characteristics of novelty, accessibility, standardization, retrievability, and long time series. As the boundary between science and technology becomes increasingly blurred, the connection between technology topics and research topics grows closer, leading many scholars to adopt both paper and patent literature simultaneously.

Concepts related to "emerging technology topic identification" include "emerging topic detection," "emerging research front identification," "emerging trends detection," "burst word detection," and "new event detection." These studies share the common feature of identifying or detecting emerging topics or themes that have appeared in recent scientific research activities but have not yet gained widespread recognition. "Emerging technology topic identification" generally refers to the identification of newly emerged branch technology fields, directions, or topics within technology domains, differing from other related research in terms of identification objects.

3 Methods for Emerging Technology Topic Identification

The evolution, monitoring, and identification of emerging topics in science and technology have long been of interest to governments, enterprises, and scientists, with continuous government funding for such research. In the late 1990s, the U.S. Defense Advanced Research Projects Agency (DARPA) implemented the "Topic Detection and Tracking (TDT) Program," which ran for several years [11]. In 2010, the America COMPETES Act [12] explicitly identified the recognition of emerging and innovative fields as a work objective. In 2011, the Intelligence Advanced Research Projects Activity (IARPA) under the U.S. Director of National Intelligence funded the "Foresight and Understanding from Scientific Exposition (FUSE)" program [13], aiming to develop automated methods to systematically, continuously, and comprehensively assess emerging technologies using information from scientific and technical literature and patents. The EU's PromTech project [14] identified emerging technologies through analysis of paper literature.

Methods for identifying emerging technologies or topics fall into two major categories: qualitative research methods based primarily on expert subjective judgment, and quantitative research methods using bibliometrics and data mining on paper and patent literature. Qualitative methods include Delphi method, expert brainstorming, technology roadmapping, scenario analysis, and TRIZ methods. For example, the Institute for Prospective Technological Studies (IPTS) at the European Commission's Joint Research Centre (JRC) developed a method (IPTS-TIM) [15] that identifies existing and future technologies by evaluating their commercialization potential to support technology transfer processes. F. M. Tseng et al. [16] proposed combining scenario planning, Delphi method, and technology substitution models to identify emerging technologies. Tan Yi et al. [17] integrated technology roadmapping with real options methods to identify and select emerging technologies. Wei Guoping [18] used expert scoring to identify emerging technologies. With the explosion of information and development of computer technology, increasing numbers of scholars have begun exploring quantitative analysis based on paper and patent literature data to supplement expert judgment and compensate for its strong subjectivity. This paper focuses on quantitative research methods for emerging technology topic identification, which are categorized into three types according to their focus on different literature features and attributes: 1) keyword or document statistical methods; 2) citation network clustering methods; and 3) text mining analysis methods. The first type focuses primarily on quantitative features of literature keywords/subject terms and documents themselves; the second emphasizes citation relationships between documents; and the third delves into text content to reveal semantic meaning and relationships. These three method types exhibit numerous differences in technology topic extraction, identification indicator system construction, and method effectiveness validation.

3.1 Keyword or Document Statistical Methods

These methods typically obtain technology fields or topics based on existing paper or patent classification systems or through keyword/phrase retrieval, then represent them using papers, patents, or their subject terms/clusters, and finally identify emerging technology topics according to features such as changes in keyword or document quantities over time. A representative method is J. Kleinberg's burst detection algorithm [19], which uses an infinite-state automaton to model time-series data, with state transitions in the time series marking the emergence of burst events. Initially applied to analyzing news article data streams, this method has been widely used in emerging technology identification research [20] and has been incorporated into tools such as CiteSpaceII [21], SCI2, and NetworkWorkbench [22]. M. Bengisu [23] retrieved papers and patents in major subfields of materials science and engineering through keyword/phrase searches, compared the temporal growth of papers and patents across subfields, and extracted rapidly developing emerging technology fields. E. Schiebel [24] and I. Roche [14] classified keywords into unusual words, established words, and cross-disciplinary words based on their occurrence frequency, TF-IDF values, and Gini coefficients to reflect their diffusion across other technology fields, thereby identifying emerging technologies. T. U. Daim et al. [25] combined patent analysis with scenario analysis and growth curve analysis to forecast emerging technologies.

3.2 Citation Network Clustering Methods

Citation relationships between papers or patents can reflect the relevance of their content and topics to some extent. These methods cluster direct citation networks, bibliographic coupling networks, or co-citation networks of papers or patents to group those with similar content or topics into technology topics, while using citation relationships to measure topic evolution paths and trends and identify emerging technology topics through a series of indicators. Y. Kajikawa et al. [26] clustered direct citation networks of paper literature to track emerging technology changes in the energy field, using the average publication year of papers in each cluster as an indicator for identifying emerging technology topics. H. Small et al. [27] combined co-citation network clustering and direct citation network clustering methods to identify scientific and technological topics with novelty and rapid growth. J. Hopcroft et al. [28] used bibliographic coupling analysis to identify several emerging topics in the computer field. P. Órdó et al. [29] conducted cluster analysis on patent sets in target fields based on how patents in those fields were cited by patents from other fields, extracted sub-technology sets, and captured the emergence and development trajectories of emerging technologies by analyzing temporal changes in these sub-technology sets. A. Breitzman et al. [6] clustered "hot patents" and "next-generation patents" that cited them based on co-citation relationships among hot patents, and evaluated clustering results using indicators such as the proportion of public sector patentees, science index, originality index, and reference index to extract emerging technology sets. S. Zhang et al. [30] identified emerging technology topics in the solar photovoltaic field by combining patent direct citation network clustering with network analysis algorithms. Li Bei et al. [31] established an emerging technology identification model and related indicator system based on patent bibliographic coupling clustering according to the core characteristics of emerging technologies and patents, conducting empirical analysis on the nanotechnology field using the US Patent and Trademark Office's authorized patent database.

3.3 Text Mining Analysis Methods

With the development of data mining and text analysis technologies, increasing numbers of scholars have attempted to use such methods for technology trend analysis and emerging technology topic identification. Commonly employed methods include Subject-Action-Object (SAO) structure extraction, vector space models, LDA topic models, and machine learning. J. Kim et al. [32] used text mining and decision tree methods for technology forecasting, extracting features representing technology topic domains from fields such as paper authors, journals, and subject areas, as well as patent assignees and subject areas. S. Choi [33], J. Yoon [34], Li Xin [35], and Z. Xiao [36] used SAO structure-based semantic analysis to identify emerging technologies from patents. A patent's SAO structure reflects the functional characteristics of the patented technology. S. Choi et al. constructed co-occurrence networks of nouns and verbs in SAO structures and analyzed technology development trends and identified emerging technology topics based on social network analysis indicators such as node degree and centrality. J. Yoon and Li Xin et al. calculated similarity between SAO structures to obtain patent similarity, built patent networks, and identified emerging technology topics through outlier analysis or clustering. Z. Xiao et al. used text mining methods such as topic word clusters and SAO structure analysis, combined with technology roadmapping and expert judgment, to identify potential innovations and commercial applications in the solid lipid nanoparticle field. Ren Zhijun [37], Zhou Yuan [38], and Dong Fang [39] used LDA models to construct patent technology topics, representing a patent as a probability distribution across several topics and a topic as a probability distribution across several words, then combined a series of indicators with expert judgment for emerging technology identification. K. I. Filippovich et al. [40] identified emerging technologies in agriculture and food sectors through machine learning, ontology mining, and entity association techniques. P. Yu et al. [41] used self-organizing maps to identify emerging technology topics. Domestic scholar Wang Lingyan et al. [42] constructed a co-word network of high-frequency subject terms (title keywords) in the industrial biotechnology patent field, conducted cluster analysis to obtain nine technology topics, and then used a series of indicators to determine emerging technology topics.

4 Review of Emerging Technology Topic Identification Methods

All three categories of emerging technology topic identification methods involve steps including target domain determination, dataset construction, technology topic extraction, identification indicator system construction, and method effectiveness validation. They differ primarily in technology topic extraction, identification indicator system construction, and method effectiveness validation. This paper compares the three method types across these main steps using a two-dimensional coordinate diagram, with the vertical axis representing the three method categories and the horizontal axis representing the three main steps of emerging technology topic identification (see Figure [Figure 1: see original paper]).

The differences among the three methods are most pronounced in technology topic extraction, while they share some common indicators in identification indicator system construction but with different emphases. Method effectiveness validation shows the greatest commonality across methods. Each method category has its own advantages and disadvantages at each identification step, which are analyzed and discussed below.

4.1 Technology Topic Extraction

Keyword or document statistical methods primarily obtain paper or patent data topic divisions based on existing classification systems or keyword/phrase retrieval. Existing classification systems include Web of Science (WoS) or Scopus journal classifications and international patent classifications. While this approach uses universal classification methods that are easily recognized, it struggles to reflect dynamic changes in scientific research and technological development. Citation network clustering methods construct technology topics based on citation relationships between papers or patents, including direct citation relationships, bibliographic coupling relationships, and co-citation relationships [43-44]. Compared with methods based on existing classification systems, this approach reveals relationships between individual documents and can reflect dynamic technological changes, but has limitations such as diverse citation motivations, thematic dissimilarity between cited and citing documents, and time lags since citations occur after publication. Text mining analysis methods typically construct technology topics based on text content or word co-occurrence relationships, such as the co-word network clustering, SAO structures, vector space models, and LDA topic models mentioned above. These methods, based on rapidly developing data mining and deep learning technologies, provide deep revelation of text content and can extract technology topics more accurately, though they still have room for improvement. Vector space models calculate based on term frequency, but term frequency struggles to accurately reflect word semantics and inter-word relationships, making the constructed vectors inadequate for accurately measuring text topic content. Some studies have improved vector space models by using external lexicons such as WordNet to measure semantic similarity between words and combining TF-IDF algorithms for text representation and classification, but these methods still struggle to accurately measure word meanings based on contextual information [45-46]. The LDA topic model [47] is a statistical model used to discover abstract topics in a document collection, representing each document as a probability distribution across topics and each topic as a probability distribution across words. However, probability distributions only describe co-occurrence statistical relationships in the corpus and are not the optimal choice for text feature representation, often making it difficult to determine exact topic meanings from a set of words.

4.2 Construction of Emerging Technology Topic Identification Indicator Systems

The three method categories each emphasize different indicators in constructing identification systems. Keyword or document statistical methods typically use quantity change indicators for papers or patents and subject terms, such as the paper and patent quantity change indicators used by M. Bengisu et al. [23], and keyword occurrence frequency, TF-IDF values, and Gini coefficients used by P. Órdó [29] and E. Schiebel et al. [24], which essentially reflect changes in document and keyword quantities. Citation network clustering methods employ more diverse indicators, such as average publication year of papers or patent grant year in clusters, and changes in the number of papers or patents in clusters. Y. Kajikawa et al. [26] used the average publication year of papers in topic clusters as an indicator for identifying emerging technologies in the energy field, with H. Small using similar indicators. Text mining analysis methods use more indicators that reveal text content, such as relationships and changes of words in SAO structures and similarity comparisons with known emerging technology topics. Since citation network clustering and text mining analysis methods can construct technology topic relationship networks based on citation or content relationships, they also employ social network analysis indicators such as degree centrality, betweenness centrality, and structural holes, as used by S. Zhang [30] and Wang Lingyan et al. [42]. A few scholars have used multi-indicator approaches for emerging technology topic identification, such as J. Kim et al. [32] using numbers of paper authors, journals, patent assignees, and subject fields;

A. Breitzman et al. [6] using assignee types, technology-science links, technology originality index, and references to prior technology; and Li Bei et al. [31] using median patent grant year and number of patent claims. The above analysis shows that most existing indicator systems inadequately reflect the characteristics of emerging technology topics and have room for improvement and refinement.

Based on the connotations and characteristics of emerging technology topics and existing research, the author has constructed an indicator system comprising seven characteristic dimensions: novelty, scale, growth rate, influence, science linkage, market potential, and uncertainty (see Table), which will be evaluated through experiments in future research.

4.3 Method Effectiveness Validation

All three method categories use expert consultation, policy documents or roadmaps for corroboration, or comparison with other methods to validate the effectiveness of identification methods and indicator systems. Text mining analysis methods have begun exploring more rigorous validation through construction of training and test sets using relevant evaluation indicators, though such studies remain limited in number. Expert consultation has the disadvantage of expert subjectivity and is also influenced by experts' knowledge scope. Some scholars use policy documents or roadmaps for corroboration, such as Y. Kajikawa et al. [26] comparing identified emerging energy technologies with expert roadmaps drawn by Japanese government agencies, though this method faces issues of mismatched granularity between identified topics and roadmaps, as well as reliance on manual interpretation for evaluation. Comparison with existing research methods is also commonly used for validation, such as Q. Wang [48] comparing identified emerging topics with those mentioned in existing literature, though this may encounter problems of different definitions of emerging topics and varying granularity between studies.

4.4 Future Prospects for Emerging Technology Topic Identification Methods

The accuracy of technology topic extraction based on papers or patent literature depends on accurate understanding and analysis of document content, which can be enhanced by rapidly developing data mining and deep learning technologies. In recent years, some deep learning-based natural language processing models have achieved good results in text semantic analysis, such as neural network language models and Google's 2013 Word2Vec model, which learns distributed word vectors for text representation using word context information. This can solve problems such as data sparsity and lack of semantic expression capability, and to some extent address the inability of co-word network clustering, vector space models, and LDA topic models to accurately reflect word meanings. Therefore, the application of such methods to technology topic extraction can be explored.

Emerging technologies are characterized by being newly emerged, rapidly developing, based on high technology, and having enormous market potential. Existing research on emerging technology topic identification indicators typically considers only one or a few of these features. The author argues that more comprehensive and systematic indicators for emerging technology topic identification should be considered, and then selected through appropriate methods to optimize identification effectiveness. Emerging technology topic identification is essentially a classification problem—dividing a set of technology topics into emerging and non-emerging categories. Therefore, constructing training sets for emerging technology topic identification could be considered to select identification indicators through supervised machine learning, followed by more rigorous validation of indicator system and method effectiveness using test sets. Although a few studies have begun using such methods for validation, they remain exploratory with considerable room for future development.

Moreover, quantitative research methods for emerging technology topic identification are often based on small datasets from specific domains. However, with increasing interdisciplinary integration, emerging technologies are likely to appear at the intersection of multiple disciplines. Therefore, discovering emerging technologies from large-scale datasets across multiple disciplines is significant. Few studies have analyzed large datasets, with H. Small et al. [27] and the ERA-CEP project [55] identifying emerging research topics based on full-domain paper data over a period, though not specifically emerging technology topics. Most current emerging technology topic identification involves retrospective analysis of predetermined domains rather than focusing on methodological research for identification, typically using topics with burst potential as empirical research data and then verifying that these topics have burst characteristics or identifying subtopics with the most prominent burst features. Strictly speaking, such studies are not true emerging technology topic identification.

Through systematic investigation and comprehensive analysis of domestic and international research on emerging technology topic identification, this paper categorizes quantitative research methods into three types: keyword/document statistical methods, citation network clustering methods, and text mining analysis methods. These methods differ in technology topic extraction, identification indicator system construction, and method effectiveness validation, each with advantages but also deficiencies. With the development of deep learning and other technologies, problems in accurately parsing paper and patent text content and extracting technology topics can be better addressed. Future research could explore deep learning-based natural language processing models for technology topic extraction, construct more comprehensive identification indicator systems based on deep understanding of "emerging technology topic" connotations, and build training and test sets for emerging technology topics. Supervised machine learning could be used to learn from emerging technology topic training sets, select truly relevant indicators, and then more rigorously validate indicator system and method effectiveness through test sets.

References

[1] George Day, Paul Shoemaker, Shi Ying. Wharton on Managing Emerging Technologies [M]. Beijing: Huaxia Publishing House, 2002. [2] Lu Ruoyu, Zhang Hongqi. Emerging Technology Product Renewal Strategy Based on Fast-Changing Markets [J]. Chinese Journal of Management, 2005(3): 67-70. [3] Li Shiming, Xiao Lei, Xiao Yangao. Review of Emerging Technology Management Research [J]. Journal of Management Sciences in China, 2007, 10(6): 12. [4] Xu Jianguo, Li Mengjun, You Hanlin. Research Progress on Emerging Technology Identification [J]. Journal of Intelligence, 2018, 37(12): 11-16. [5] Cozzens S, Gatchair S, Kang J, et al. Emerging technologies: quantitative identification and measurement [J]. Technology Analysis & Strategic Management, 2010, 22(3): 361-376. [6] Breitzman A, Thomas P. The Emerging Clusters Model: a tool for identifying emerging technologies across multiple patent systems [J]. Research Policy, 2015, 44(1): 195-205. [7] Cheng Yue, Yin Lu. Research on Emerging Technology Evolution Model Based on Corporate Dynamic Capabilities and Case Studies [J]. Chinese Journal of Management, 2010, 7(1): 43-49. [8] Rotolo D, Hicks D, Martin B R. What is an emerging technology? [J]. Research Policy, 2015, 44(10): 1827-1843. [9] Christensen C M. The Innovator's Dilemma [M]. Boston: Harvard Business Review Press, 1997. [10] Hu Apei, Zhang Jing, Lei Xiaoping, et al. Review of Patent Technology Topic Analysis Based on Text Mining [J]. Journal of Intelligence, 2013, 32(12): 88-92. [11] Allan J, Carbonell J G, Doddington G, et al. Topic detection and tracking pilot study final report [R]. Virginia: Defense Advanced Research Projects Agency, 2000. [12] United States Congress. America COMPETES Act [EB/OL]. [2019-09-13]. https://www.congress.gov/111/plaws/publ358/PLAW-111publ358.pdf. [13] Intelligence Advanced Research Projects Activity. Foresight and Understanding from Scientific Exposition (FUSE) [EB/OL]. [2019-09-13]. https://www.iarpa.gov/index.php/research-programs/fuse. [14] Roche I, Besagni D, François C, et al. Identification and characterisation of technological topics in the field of Molecular Biology [J]. Scientometrics, 2010, 82(3): 663-676. [15] Moncada-Paternó-Castello P, Rojo J, Bellido F, et al. Early identification and marketing of innovative technologies: a case study of RTD result valorisation at the European Commission's Joint Research Centre [J]. Technovation, 2003, 23(8): 655-663. [16] Tseng F M, Cheng A C, Peng Y N. Assessing market penetration of emerging technologies using a combination of scenario analysis, Delphi, and the technological substitution model: the case of the OLED TV market [J]. Technological Forecasting and Social Change, 2009, 76(7): 897-909. [17] Tan Yi, Huang Yanli. Research on Process-Based Emerging Technology Planning and Selection Model [J]. Science and Technology Management Research, 2007, 27(8): 5-8. [18] Wei Guoping. Research on Emerging Technology Management Strategy [D]. Hangzhou: Zhejiang University, 2006. [19] Kleinberg J. Bursty and hierarchical structure in streams [J]. Data Mining and Knowledge Discovery, 2003, 7(4): 373-397. [20] Mane K K, Börner K. Mapping topics and topic bursts in PNAS [J]. Proceedings of the National Academy of Sciences, 2004, 101(S1): 5287-5290.

[21] Chen C. CiteSpace II: Detecting and visualizing emerging trends and transient patterns in scientific literature [J]. Journal of the American Society for Information Science and Technology, 2006, 57(3): 359-377. [22] Börner K, Huang W, Linnemeier M, et al. Rete-netzwerk: analyzing and visualizing scholarly networks using the Network Workbench Tool [J]. Scientometrics, 2010, 83(3): 863-876. [23] Bengisu M. Critical and emerging technologies in materials, manufacturing, and industrial engineering: a study for priority setting [J]. Scientometrics, 2003, 58(3): 473-487. [24] Schiebel E, Hörlesberger M, Roche I, et al. An advanced diffusion model to identify emergent research issues: the case of optoelectronic devices [J]. Scientometrics, 2010, 83(3): 765-781. [25] Daim T U, Rueda G, Martin H, et al. Forecasting emerging technologies: Use of bibliometrics and patent analysis [J]. Technological Forecasting and Social Change, 2006, 73(8): 981-1012. [26] Kajikawa Y, Yoshikawa J, Takeda Y, et al. Tracking emerging technologies in energy research: toward a roadmap for sustainable energy [J]. Technological Forecasting and Social Change, 2008, 75(6): 771-782. [27] Small H, Boyack K W, Klavans R. Identifying emerging topics in science and technology [J]. Research Policy, 2014, 43(8): 1450-1467. [28] Hopcroft J, Khan O, Kulis B, et al. Tracking evolving communities in large linked networks [J]. Proceedings of the National Academy of Sciences, 2004, 101(S1): 5249-5253. [29] Órdó P, Makovi K, Somogyvári Z, et al. Prediction of emerging technologies based on analysis of the US patent citation network [J]. Scientometrics, 2013, 95(1): 225-242. [30] Zhang S, Han F. Identifying emerging topics in a technological domain [J]. Journal of Intelligent & Fuzzy Systems, 2016, 31(4): 2147-2157. [31] Li Bei, Chen Xiangdong. Identification of Emerging Technologies in the Nano Field Based on Patent Citation Coupling Clustering [J]. Journal of Intelligence, 2015(5): 35-40. [32] Kim J, Hwang M, Jeong D H, et al. Technology trends analysis and forecasting application based on decision tree and statistical feature analysis [J]. Expert Systems with Applications, 2012, 39(16): 12618-12625. [33] Choi S, Yoon J, Kim K, et al. SAO network analysis of patents for technology trends identification: a case study of polymer electrolyte membrane technology in proton exchange membrane fuel cells [J]. Scientometrics, 2011, 88(3): 863-883. [34] Yoon J, Kim K. Detecting signals of new technological opportunities using semantic patent analysis and outlier detection [J]. Scientometrics, 2012, 90(2): 445-461. [35] Li Xin, Wang Jingjing, Yang Zi, et al. Research on Emerging Technology Identification Based on SAO Structure Semantic Analysis [J]. Journal of Intelligence, 2016, 35(3): 80-84. [36] Xiao Z, Lu H, Alan P, et al. Tracing the system transformations and innovation pathways of an emerging technology: solid lipid nanoparticles [J]. Technological Forecasting and Social Change, 2018, 146(2): 785-794. [37] Ren Zhijun, Qiao Xiaodong, Zhang Jiangtao. Research on Emerging Technology Discovery Model [J]. New Technology of Library and Information Service, 2016, 32(7): 60-69. [38] Zhou Yuan, Liu Yufei, Xue Lan. A Machine Learning-Based Emerging Technology Identification Method: A Case Study of Robotics Technology [J]. Journal of the China Society for Scientific and Technical Information, 2018, 37(9): 83-99. [39] Dong Fang, Liu Yufei, Zhou Yuan. Emerging Technology Prediction Based on LDA-SVM

Multi-Classification of Paper Abstracts [J]. Journal of Intelligence, 2017(7): 44-49, 137. [40] Filippovich K I, Denisovich B P, Stanislavovna L A. Big-data-augmented approach to emerging technologies identification: case of agriculture and food sector [J]. Social Science Research Network, 2017(1): 130-134. [41] Yu P, Lee J H. A hybrid approach using two-level SOM and combined AHP rating and AHP/DEA-AR method for selecting optimal promising emerging technology [J]. Expert Systems with Applications, 2013, 40(1): 300-314. [42] Wang Lingyan, Fang Shu, Ji Peipei. Research on Technical Framework for Identifying Emerging Technology Topics Using Patent Literature [J]. Library and Information Service, 2011, 55(18): 74-23. [43] Furukawa T, Mori K, Arinobu K, et al. Identifying the evolutionary process of emerging technologies: a chronological network analysis of World Wide Web conference sessions [J]. Technological Forecasting and Social Change, 2015, 91: 280-294. [44] Zhang Y, Porter A L, Hu Z, et al. "Term clumping" for technical intelligence: a case study on dye-sensitized solar cells [J]. Technological Forecasting and Social Change, 2014, 85(2): 26-39. [45] Patil L H, Atique M. A novel approach for feature selection method TF-IDF in document clustering [C]//Advance Computing Conference (IACC), 2013 IEEE 3rd International. Ghaziabad: IEEE, 2013: 858-862. [46] Huang Chenghui, Yin Jian, Hou Yu. A Text Similarity Measurement Method Combining Term Semantic Information and TF-IDF [J]. Chinese Journal of Computers, 2011, 34(5): 856-864. [47] Yan E. Research dynamics: measuring the continuity and popularity of research topics [J]. Journal of Informetrics, 2014, 8(1): 98-110. [48] Wang Q. A bibliometric model for identifying emerging research topics [J]. Scientometrics, 2017, 69(2): 290-304. [49] Goh H, Weingart S, Börner K. Mixed-indicators model for identifying emerging research areas [J]. Scientometrics, 2011, 89(1): 421-435. [50] Christopher E, Silberglitt R. Identification and analysis of emerging technologies [J]. Technological Forecasting and Social Change, 2017, 69(2): 290-304. [51] Guo Ruilan. Application of Scenario Planning in Emerging Technology Strategy Formulation [D]. Chengdu: University of Electronic Science and Technology of China, 2007. [52] Noh H, Song Y, Lee S. Identifying emerging core technologies: case study of patents published by leading telecommunication organizations [J]. Telecommunications Policy, 2016, 40(10/11): 956-970. [53] Zhang Wei, Chen Shaogang. Research on Technology Risk at Each Stage of Emerging Technology Adoption Life Cycle [J]. Science and Technology Management Research, 2007, 27(5): 161-163. [54] Huang Lucheng. Research on Emerging Technology Identification Based on Attribute Comprehensive Evaluation System [J]. Science Research Management, 2009, 30(4): 190-194. [55] European Research Council. ERACEP-emerging research areas and their coverage by ERC-supported projects final report [EB/OL]. [2019-09-13]. https://erc.europa.eu/sites/default/files/document/file/Eracep_{{{final}}_{{report}}}.pdf.

Author Contributions: Liu Xiaoling: Participated in paper framework design, wrote initial draft, revised paper. Tan Zongying: Proposed research ideas and paper framework, reviewed and revised paper.

Abstract: [Purpose/significance] Identification of emerging technology topics not only contributes to tracking technological development dynamics but also enables early capture of future development opportunities and trends. Reviewing quantitative methods for emerging technology topic identification and comparing their advantages and disadvantages can provide references for methodological improvement. [Method/process] Concepts such as "emerging technology" and "emerging technology topic identification" were analyzed; qualitative and quantitative research methods for emerging technology topic identification were investigated, focusing on bibliometrics and data mining; quantitative methods were divided into three categories: keyword/document statistical methods, citation network clustering methods, and text mining analysis methods; similarities, differences, and shortcomings were analyzed regarding technology topic extraction, indicator system construction, and method validation; preliminary improvement thoughts were provided. [Result/conclusion] The three method types have distinct characteristics, advantages, and disadvantages in the three steps of emerging technology topic identification, with room for further improvement. Future research could explore deep learning for accurate topic extraction, build more comprehensive indicator systems, and conduct rigorous validation based on machine learning.

Keywords: emerging technology; topic identification; methodology research; bibliometrics; text mining

Note: Figure translations are in progress. See original paper for figures.