Research on Key Technology Identification Methods Integrating Literature Knowledge Clustering and Complex Networks: Postprint

Authors: Wang Yanpeng, Han Tao, Chen Fang

Date: 2023-04-01T00:00:00+00:00

Abstract

[目的/意义] Grounded in the perspective of intelligence research, this study proposes a scientific, effective, and reusable/generalizable methodology for key technology identification, aiming to provide intelligence-based support for nations, regions, enterprises, and innovation institutions in discovering, deploying, and promoting forward-looking planning for key technology R&D. [方法/过程] Based on the definition of key technology types and concepts, literature knowledge clustering is employed to identify hotspot technologies. A complex network is constructed with each hotspot technology as a node, and the technology structure network is displayed through secondary node clustering and visualization methods. Structural holes theory is adopted to analyze network and node characteristics, thereby selecting generic technologies. Link prediction methods are utilized to predict the likelihood of missing edges forming connections in the technology structure network, and to analyze the phenomenon of hotspot technology cross-fusion promoting the formation of innovative technologies, thereby identifying potential emerging technologies. [结果/结论] An empirical study on key technology identification is conducted using the intelligent manufacturing domain as an example. Through comparison with national authoritative planning documents and literature investigation, the operability and effectiveness of the methodology are preliminarily verified.

Full Text

A Method for Identifying Key Technologies Based on Literature Knowledge Clustering and Complex Networks

Wang Yanpeng[1,2], **Han Tao**[1,2], **Chen Fang**[1,2] [1] National Science Library, Chinese Academy of Sciences, Beijing 100190 [2] Department of Library, Information and Archives Management, School of Economics and Management, University of Chinese Academy of Sciences, Beijing 100190

Abstract: [Purpose/Significance] From the perspective of intelligence research, this paper proposes a scientific, effective, and reusable method for identifying key technologies, aiming to provide intelligence support for nations, regions, enterprises, and innovation institutions to discover, deploy, and promote prospective R&D layouts of key technologies. [Method/Process] Based on the definition and classification of key technology types, we utilize literature knowledge clustering to identify hotspot technologies, construct a complex network with these hotspot technologies as nodes, visualize the technology structure network through secondary node clustering, employ structural holes theory to analyze network and node characteristics for selecting generic technologies, and apply link prediction methods to forecast the likelihood of missing edges forming connections in the technology structure network. By analyzing how cross-fertilization of hotspot technologies promotes the formation of innovative technologies, we identify potentially emerging technologies. [Result/Conclusion] Taking the intelligent manufacturing domain as an example, we conduct an empirical study on key technology identification and preliminarily validate the operability and effectiveness of our method through comparison with national authoritative planning documents and literature research.

Keywords: Key technology identification; Complex network; Structural holes; Link prediction; Technical structure **Classification Number:** G250 **DOI:** 10.13266/j.issn.0252-3116.2020.16.011

Key technologies refer to critical and indispensable components or technologies within a system or technical field. These may represent current technological hotspots and challenges, future breakthrough points, or crucial knowledge for a particular domain. At the 2018 Academicians' Conference, President Xi Jinping emphasized the need to "take key generic technologies, frontier leading technologies, modern engineering technologies, and disruptive technological innovations as breakthrough points, dare to forge new paths where predecessors have not ventured, and strive to achieve independent control of core technologies."Against the backdrop of China-US trade tensions, a series of key technology threats arising from unbalanced technological development have become increasingly prominent, such as high-end photolithography machines, intelligent core algorithms, lithium battery separators, and photovoltaic inverters, which have permeated and impacted China' s social development, economic progress, and public welfare. In the current wave of technological revolution and industrial transformation, clearly understanding key technologies in strategic frontier domains holds significant practical importance for nations, regions, and enterprises to develop prospective R&D layouts.

1 Related Research

The goal of achieving technology identification and foresight, selecting technologies that may generate maximum economic and social benefits, largely depends on a scientifically sound theoretical and methodological system. With in-depth research and practice by governments, academia, and industry, methods for key technology identification and foresight continue to enrich and improve. Current research methods can be summarized into the following categories:

1.1 Expert Wisdom-Based Methods

These methods rely on domain experts' years of research and practical experience to identify and forecast the evolution direction, development roadmap, and implementation timeline of key technologies within current social, economic, and political contexts. Specific implementation methods include the Delphi method and technology roadmapping. Japan' s 10th Science and Technology Foresight activity employed the Delphi method for multidisciplinary, multi-topic surveys to identify and foresee key technologies. Similarly, China' s Ministry of Science and Technology conducted technology foresight surveys and key technology identification for the 13th Five-Year Science and Technology Plan, while the Chinese Academy of Sciences has continuously conducted 20-year technology foresight research. Beijing and Shanghai have also deployed relevant work on key technology identification and foresight. N. Uchihira proposed a method for developing concurrent system technology roadmaps to identify key R&D characteristics.

Expert wisdom-based methods consolidate experts' extensive research experience and produce scientifically robust and authoritative results. However, they are significantly influenced by experts' subjective perceptions, making it difficult to reach consensus during the early stages of new technologies and ideas, potentially overlooking some potentially emerging key technologies.

1.2 Multi-Indicator Assessment Methods

These methods establish multi-indicator assessment frameworks to identify and forecast key technologies, often validated through technical case studies. Current multi-indicator assessment methods can be divided into three categories: First, quantitative research based on objective indicators, such as C.Y. Lee et al.' s construction of a multi-patent indicator system using machine learning for early identification of emerging technologies. Second, qualitative research based on subjective indicators, such as D. Nagy et al.' s three-step method for assessing disruptive innovations, including technological innovation characteristics, value chain aspects, and comparison with incumbent technologies. Third, mixed qualitative-quantitative methods based on subjective scoring, where experts score subjective evaluation indicators that are then statistically weighted, as exemplified by the Chinese Academy of Engineering' s "China Engineering Science and Technology 2035 Development Strategy Research."

Multi-indicator assessment methods provide comprehensive, systematic evaluation through multi-dimensional indicator systems, offering strong scientific rigor and objectivity. However, some assessment indicators still require expert scoring, which affects objectivity.

1.3 Model-Based Methods

These methods construct models based on theoretical frameworks and criteria through statistical approaches to identify and forecast key technologies. Common models include TRIZ theory, data envelopment analysis, catastrophe theory, trend extrapolation, and out-of-sample forecasting methods. J.G. Sun et al. employed TRIZ theory to determine whether a technology is mainstream or lagging and forecast its future prospects. Li Zheng et al. used mathematical modeling (cusp catastrophe equations) to construct a key technology foresight model for S&T evaluation and forecasting.

Model-based methods draw on fundamental theories from economics, operations research, and mathematics to form relatively complete analytical frameworks and criteria, enabling more systematic and objective technology identification. However, these methods require researchers to possess comprehensive domain expertise and thorough understanding of technology market status and trends, demanding high professional competence and practical skills.

1.4 Text Content and Relationship-Based Methods

These methods primarily utilize multi-source textual data such as scientific literature and social media, employing bibliometrics, text mining, and relationship analysis for technology identification and foresight. Current approaches include traditional bibliometric methods and techniques integrating machine learning, big data, and complex network theory. For instance, F. Dotsika et al. extracted domain keywords to construct co-occurrence networks for emerging technology identification. Bai Guangzu et al. conducted disruptive technology foresight from the perspective of literature knowledge associations. T.C. Jin et al. built large citation networks from patent data to identify radical technological innovations through main path evolution analysis. Zhou Yuan et al. employed LDA and SVM models with time series methods to forecast emerging technology trends.

Text content and relationship-based methods adopt a quantitative research perspective with strong objectivity, mitigating expert cognitive bias impacts on foresight results. They feature concise processes and objective outcomes, serving as powerful tools for key technology identification. However, these methods remain in continuous development without a unified, widely recognized methodological framework.

In summary, expert wisdom, multi-indicator assessment, and model-based methods require deep domain expert involvement, emphasizing qualitative analysis with high costs, strong subjectivity, and weak result interpretability. In contrast, text content and relationship-based methods emphasize quantitative analysis, achieving key technology identification through multi-source textual data and explicit analytical processes, offering advantages of low operational cost, high reproducibility, and strong interpretability. Therefore, this paper integrates machine learning methods and complex network theory from the perspective of text content and relationships to construct a scientific, effective, and reusable methodological system, providing a new solution for key technology identification.

2 A Method for Identifying Key Technologies Based on Literature Knowledge Clustering and Complex Networks

Based on the definition and classification of key technology types, this study employs literature knowledge clustering to identify hotspot technologies, constructs a complex network with these hotspot technologies as nodes, visualizes the technology structure network through secondary clustering and visualization methods, applies structural holes theory to analyze network and node characteristics for selecting generic technologies, and utilizes complex network link prediction algorithms to forecast the likelihood of missing edges forming connections in the technology structure network. By analyzing how cross-fertilization of hotspot technologies promotes scientific and technological innovation, we identify potentially emerging technologies. The overall research methodology is illustrated in Figure 1 [Figure 1: see original paper].

2.1 Definition and Classification of Key Technology Types

Based on literature research, this paper defines three types of key technologies: Hotspot technologies: Main research directions and technical topics that have received widespread attention from researchers and produced corresponding research outcomes in recent years within a discipline, reflecting the current R&D status and overall technology structure; Generic technologies: Technologies that exert broad influence on other research directions and technical topics within a discipline, whose research outcomes can be referenced, learned from, and shared, reflecting important R&D foundations and serving as "idea sources" within the field; Potentially emerging technologies: Future research directions and technical topics that may emerge due to scientific cross-fertilization within a discipline, reflecting important R&D prospects and worthwhile technology intersection points.

2.2 Hotspot Technology Identification Based on Literature Knowledge Clustering

Hotspot technologies represent mainstream research directions in current disciplinary fields, essentially covering major research focuses and efforts. Text clustering, as an unsupervised machine learning method, groups documents with similar content into clusters based on the assumption that similar texts have greater similarity while dissimilar texts have less similarity. Combined with domain expertise for document interpretation, this approach summarizes core scientific problems and solutions for each cluster, thereby identifying current hotspot technologies and transforming mathematical spatial relationships into domain knowledge discovery. This study first selects core literature data from a specific discipline, performs text preprocessing and vectorization, and employs unsupervised clustering methods to aggregate literature knowledge into multiple clusters. Through manual interpretation and intelligence research, we decipher the meaning of each cluster and identify those with clear boundaries and explicit meanings as domain hotspot technologies.

2.2.1 Text Preprocessing and Representation Text preprocessing and representation include tokenization, lemmatization, feature word extraction, and vectorization. First, we use Python' s NLTK package to tokenize each document by spaces. Second, we lemmatize each token by removing affixes and extracting stems (e.g., "cars→car" , "ate→eat"). Third, for each document, we select nouns and verbs after stop-word filtering as feature words to represent main content. Finally, we use the TF-IDF method for document vectorization.

2.2.2 Document Clustering Document clustering employs the K-means++ algorithm. When clustering, we must predetermine the number of clusters (K value). We calculate and compare silhouette coefficients for multiple K values while manually assessing actual clustering effects to comprehensively evaluate clustering quality and select the optimal number of clusters.

2.3 Generic Technology Identification Based on Structural Holes Theory

Generic technologies reflect the R&D foundation of current disciplinary fields and play important promoting roles. Structural holes theory describes non-redundant connections, and H.J. Raider' s empirical research demonstrates that positions occupying structural holes are crucial for information control, identification, and transactions. Thus, structural holes reflect nodes' control over network resources, with nodes occupying structural holes able to obtain non-repetitive information from multiple sources. In disciplinary research, research directions occupying structural holes are more likely to be "idea sources"—generic technologies that broadly influence other research directions. This study uses cluster centroid vectors as nodes and vector similarity as connection criteria to construct a complex network, visualizing the technology structure network and performing secondary clustering of centroid vectors. Based on structural holes theory, we calculate constraint indices for each node to reflect their ability to utilize structural holes. Smaller constraint values indicate more structural holes and stronger utilization capabilities, enabling identification of domain generic technologies.

2.3.1 Network Construction and Visualization Cluster centroid vectors, calculated as the mean of all document vectors within a cluster, reflect cluster centers spatially and core content logically. We use centroid vectors to represent hotspot technologies as network nodes, employing cosine similarity to calculate pairwise similarities as connection weights. We visualize the network using Gephi software and apply its integrated modularity algorithm for secondary clustering of centroid vectors, dividing hotspot technologies into several major technology categories.

2.3.2 Structural Holes Calculation We construct a symmetric matrix from pairwise cosine similarities between centroid vectors and input it into Ucinet software via "Network→Ego Networks→Structural Holes" to calculate constraint indices for each node.

2.4 Potential Emerging Technology Identification Based on Complex Network Link Prediction

Cross-disciplinary integration and innovation have become distinctive features of contemporary S&T innovation. Forecasting the likelihood of cross-fertilization among multiple research directions and technical topics helps identify potentially emerging technologies. This study uses the technology structure network constructed above and applies complex network link prediction methods to detect node pairs that are not yet connected but may form strong connections in the future. By analyzing how hotspot technology cross-fertilization promotes scientific innovation, we identify potentially emerging technologies. Specifically, we employ network structure similarity-based link prediction methods, which offer low computational complexity and high accuracy, with network topological properties helping select appropriate similarity metrics. Since the technology structure network is a weighted undirected network, we use the weighted Adamic-Adar (AA) index to calculate connection likelihood, comprehensively considering common neighbor numbers and node degrees for more accurate predictions.

3 Empirical Analysis of Key Technology Identification

To verify the operability and effectiveness of our method, we conduct an empirical study in the intelligent manufacturing domain.

3.1 Core Literature Data Selection

This study analyzes scientific papers as research objects. Based on the Journal Citation Reports, we select Q1 and Q2 journals in "Engineering, Manufacturing" and Q1 journals in "Engineering, Electrical & Electronic", "Engineering, Mechanical", "Engineering, Multidisciplinary", "Computer Science, Artificial Intelligence", "Computer Science, Cybernetics", and "Computer Science, Interdisciplinary Applications". Through evaluation of journal titles, descriptions, and representative papers, we identify 26 journals related to intelligent manufacturing (see Table 1), including 7 closely related journals and 19 related journals. We develop search terms based on the intelligent manufacturing knowledge system and retrieve 15,345 papers from the Web of Science Core Collection (2015-2018) using strategies of searching all articles from closely related journals and limiting searches with keywords for related journals.

3.2 Hotspot Technology Identification and Technology Structure Network Construction

Using titles, abstracts, and keywords from the core literature as analytical data, we perform text preprocessing, vectorization, and K-means++ clustering. Based on silhouette coefficient calculations, we determine that K=100 yields optimal clustering results. Through manual interpretation, we identify 98 valid clusters and 2 mixed clusters, using the 98 valid clusters as intelligent manufacturing hotspot technologies. Using these 98 hotspot technologies as nodes, we construct and visualize the technology structure network, performing secondary clustering to form 10 major technology categories (see Figure 2 [Figure 2: see original paper] and Table 2).

In intelligent manufacturing technologies, three major categories emerge: robotics and industrial control technology, artificial intelligence technology, and IoT and sensor technology. Hotspot technologies mainly include industrial robots, robot tracking and adaptation, motor design and control, actuators and drivers, adaptive control, vehicle dynamics and control, model predictive control, sliding mode control, delay control, consensus control, machine learning, pattern mining, and computer vision applications in manufacturing processes. Additionally, sensor technology, IoT computing, IoT security, network manufacturing, and other core IoT technologies are included.

In manufacturing processes, four major categories emerge: forming processes, milling processes, welding processes, and special processing technologies. Hotspot technologies include material properties and 3D printing, selective laser melting and laser processing, injection molding, extrusion forming, forging, hydraulic forming, deep drawing, tool path planning, tool wear, error estimation, chatter stability, cutting force modeling, chip formation, heat treatment, stress analysis, processing deformation, friction stir welding, electrical discharge machining, and powder metallurgy.

In advanced manufacturing modes, hotspot technologies include design for manufacturability theory, cellular manufacturing, additive manufacturing, remanufacturing, product design optimization, inventory management, facility layout optimization, supply chain optimization, intelligent assembly, production planning and scheduling, project management, manufacturing system reliability assessment, distributed state estimation, control chart optimization, predictive maintenance, and fault monitoring and isolation.

3.3 Generic Technology Identification Analysis

Using the intelligent manufacturing hotspot technology structure network (Figure 2), we calculate constraint indices for each node and select the Top 20 hotspot technologies as generic technologies in intelligent manufacturing (see Table 3).

In intelligent manufacturing processes, 11 generic technologies are identified, including manufacturing process monitoring and control theories, design for manufacturability theory, advanced manufacturing mode exploration, and cellular manufacturing. In manufacturing processing, six generic technologies are identified, including precision machine tools, material properties and 3D printing, selective laser melting and laser processing, injection molding, and tool path planning.

To validate these results, we compare them with the "Intelligent Manufacturing Development Plan (2016-2020)" issued by the Ministry of Industry and Information Technology and the Ministry of Finance in December 2016. The plan mentions 12 key generic technology innovation directions. Our identified 20 generic technologies basically cover these 12 directions (except high-reliability real-time communication), with more detailed granularity and clearer technical scope (see Table 4).

3.4 Potential Emerging Technology Identification Analysis

Based on the intelligent manufacturing hotspot technology structure network, we predict the likelihood of missing edges forming connections and select the Top 10 hotspot technology node pairs as potentially emerging technologies (see Table 5).

In intelligent manufacturing hotspot technologies, the combination of design for manufacturability theory with material properties and 3D printing, as well as the integration of material properties and 3D printing with hydraulic forming, may produce deep coupling and cross-fertilization to become future emerging technologies. Additionally, combinations such as network manufacturing and AI technology applications, advanced manufacturing mode exploration, robot tracking adaptive predictive control, and manufacturing system scheduling and control technology also show potential.

To verify the rationality of these results, we examine literature on "design for manufacturability theory applied to material properties and 3D printing." Professor Liu Shutian from the State Key Laboratory of Structural Analysis for Industrial Equipment at Dalian University of Technology noted that most existing additive manufacturing structures still adopt design configurations for traditional manufacturing processes, failing to fully utilize the novel design space offered by additive manufacturing. This limits performance improvements and, due to immature additive manufacturing technologies, may even underperform traditional manufacturing. Therefore, developing design methods for additive manufacturing to create complete theoretical systems and break traditional design limits represents a crucial research direction. This literature case demonstrates the rationality of identifying this combination as a potentially emerging technology.

Conclusion

With the deepening "data-driven" concept and the gradual popularization of machine learning, big data technology, and complex network theory, new problem-solving approaches have emerged for key technology identification. From an intelligence research perspective, this paper proposes a method for key technology identification based on the fundamental laws of S&T innovation, using literature knowledge clustering and technology structure complex network construction as the main analytical threads, and employing unsupervised clustering, structural holes theory, and complex network link prediction to identify hotspot, generic, and potentially emerging technologies. The empirical study in intelligent manufacturing, validated through comparison with national authoritative documents and literature research, preliminarily confirms the operability and effectiveness of our method.

Limitations and future improvements include: enhancing lemmatization accuracy during text preprocessing, combining structural holes theory with other network and objective metrics for multi-dimensional generic technology revelation, and reducing subjectivity in cluster interpretation that currently requires domain experts or intelligence analysts.

References

[1] Wang Xuefeng, Lai Yuangen, Zhu Donghua. Research on technology threat theory[J]. Studies in Science of Science, 2009, 27(2): 166-169.

[2] Shichijo Naohiro, Ogasawara Atsushi. The 10th Science and Technology Foresight Survey (Vision): Scenario Planning from an International Perspective[J]. Annual Conference Proceedings, 2015, 30(1): 882-885.

[3] Chinese Academy of Sciences Technology Foresight Research Group. China' s 20-Year Technology Foresight[M]. Beijing: Science Press, 2006.

[4] Li Hong, Sun Shaorong, Liu Jiyun. Empirical research on technology foresight for Shanghai's key S&T development areas[J]. Studies in Science of Science, 2005, 23(s1): 101-105.

[5] Mu Rongping. Beijing technology foresight: Practice and reflection[J]. World Science, 2003, 15(4): 41-43.

[6] Uchihira N. Future direction and roadmap of concurrent system technologies[C]//Portland International Conference on Management of Engineering and Technology. Honolulu, HI: IEEE, 2017: 649-656.

[7] Lee C, Kwon O, Kim M, et al. Early identification of emerging technologies: A machine learning approach using multiple patent indicators[J]. Technological Forecasting and Social Change, 2018, 50(127): 291-303.

[8] Nagy D, Schuessler J, Dubinsky A. Defining and identifying disruptive innovations[J]. Industrial Marketing Management, 2016, 46(57): 119-126.

[9] Wang Kunsheng, Zhou Xiaoji, Gong Xu, et al. Technology foresight research for China Engineering Science and Technology 2035[J]. Strategic Study of CAE, 2017, 19(1): 34-42.

[10] Sun J, Gao J, Yang B, et al. Achieving disruptive innovation-Identification of potentially disruptive technologies based upon technical system evolution by TRIZ[C]//IEEE International Conference on Management of Innovation and Technology. Bangkok: IEEE, 2008: 18-22.

[11] Li Zheng, Luo Hui, Li Zhengfeng, et al. Preliminary exploration of S&T evaluation methods based on catastrophe theory[J]. Science Research Management, 2017, 38(s1): 193-200.

[12] Dotsika F, Watkins A. Identifying potentially disruptive technologies by means of keyword network analysis[J]. Technological Forecasting and Social Change, 2017, 49(119): 114-127.

[13] Bai Guangzu, Zheng Yurong, Wu Xinnian, et al. Disruptive technology foresight based on literature knowledge association[J]. Journal of Intelligence, 2017, 36(9): 38-44.

[14] Jin T, Miyazaki K, Kajikawa Y. Identification of evolutionary characteristics of emerging technologies: The case of smart grid technology[J]. Ieice Transactions on Fundamentals, 2007, 90(11): 2443-2448.

[15] Dong Fang, Liu Yufei, et al. Emerging technology prediction based on LDA-SVM multi-classification of paper abstracts[J]. Journal of Intelligence, 2017, 36(7): 40-45.

[16] Liu Jun. Lecture Notes on Whole Network Analysis: A Practical Guide to UCINET Software[M]. Shanghai: Truth & Wisdom Press/Shanghai People' s Publishing House, 2009.

[17] Raider HJ. Market structure and innovation[J]. Social Science Research, 1998, 27(1): 1-21.

[18] Blondel VD, Guillaume JL, Lambiotte R, et al. Fast unfolding of communities in large networks[J]. Journal of Statistical Mechanics: Theory and Experiment, 2008: P10008.

[19] Liu Shutian, Li Quhao, Chen Wenjiong, et al. Integration of topology optimization and additive manufacturing: A design-manufacturing integration method[J]. Aeronautical Manufacturing Technology, 2017, 60(10): 26-31.

Author Contributions

Wang Yanpeng: Research design, experimental design and result analysis, paper writing; **Han Tao:** Research design revision, paper revision; **Chen Fang:** Research design discussion, experimental result analysis.

Abstract: [Purpose/Significance] This paper tries to propose a scientific, effective and reusable method to identify key technologies based on the perspective of intelligence research. It aims to provide information support for nations, regions, enterprises and innovative institutions to discover, deploy and promote prospective R&D of key technologies. [Method/Process] Based on the definition of key technology and its types, this paper used K-means++ algorithm to cluster scientific papers to identify hotspot technologies. Then it used the hotspot technologies as nodes to construct and visualize complex network through secondary clustering and Gephi. Structural holes theory was adopted to analysis the network and attributes of nodes, and thereby selected generic technologies. Link prediction algorithm was used to predict the missing edges in the network according to the structure, and we can identify the potential emerging technologies based on the phenomenon of cross-fusion of hot technologies to promote the formation of innovative technologies. [Result/Conclusion] Taking the Intelligent Manufacturing as an example to carry out empirical research on the method, and validated the operability and effectiveness of the method through national authoritative documents and literature research.

Keywords: key technology identification; complex network; structural holes; link prediction; technical structure