Technical standards opportunity identification of Chinese manufacturing industry

Xueyuan Wang[1,*], and *Meixia* Sun[1]

[1]Harbin University of Science and technology, School of economics and Management, 150080, Harbin, Heilongjiang Province

Abstract. Based on Standard Literature Library of CNKI, opportunities for technical standards development are identified. The initial keywords of technical standard are determined based on LDA model, whose similar keywords are further confirmed by Word2Vec, and then combining LDA model with Word2Vec the representative keywords of technical standard of manufacturing industry are determined. According to confirmed representative keywords, the core technical standards of manufacturing industry are identified by using the center degree and "M-core" method. The dimensions of the frequency square matrix are reduced by using the PCA module of Python, and the scatter plot is brawn, the technology gaps are identified by combing the consideration on the core technical standards, and then the future development opportunities are confirmed based on the map of patent technology effect matrix.

1 Introduction

Based on the technical basis of existing standards, identifying the opportunities of future technical standards and scientifically selecting the construction scheme of technical standards are the premise and foundation for the cultivation and development of technical standards in China's manufacturing industry. In the early stage, patent documents were mainly used as data sources for technological innovation and change management research. With the rapid development of science and technology and the rapid progress of manufacturing industry in China, the research of technical standards has more technical value and industrial guidance value. But the research based on technical standard literature is not perfect. First of all, the high frequency words or subject words based on the bibliometric method for technical opportunity recognition, its accuracy needs to be improved. Therefore, this paper uses the method of combining the theme model with the keyword similarity model to carry out bibliometrics to ensure that the extracted topic words are representative. Secondly, the current research uses the network co-occurrence map to determine the technical hot spot and the patent map to excavate the technical blank spot. Based on the combination of centrality and "m- kernel ", this paper removes the weak connection and isolated nodes.

---

*Corresponding author: 79599793@qq.com

2 Theoretical foundations

2.1 The importance of technical standards

Against the background of the in-depth promotion of "standardization ", technical standards have replaced prices, technologies, patents, and become the core of enterprise competition and the key to the direction of technological development in leading industries (Zou Siming, 2017)[1]. Peng Hailing et al. (2019) also pointed out that once a technology is successfully broken through, and the technology is put into application and integrated into the standard, the technology will be rapidly spread and become the mainstream technology in the industry. This technology standardization will help enterprises to grasp the "voice" of the industry and guide the development direction of industry technology [2]. Based on guiding the future technological direction, technological standardization or the bridge between technological innovation achievements and practical applications is the key for enterprises to realize innovation and development. In addition, technical standardization can improve product quality [3, 4].

For the manufacturing industry, technical standards are conducive to the adjustment of the internal economic structure of the industry, standardize the requirements of the internal behavior of the industry, and promote the optimization and upgrading of the industrial structure (Lu Meng 2013) [5]. Tao Zhongyuan et al. (2016) believe that technical standards promote technological innovation, promote advanced technology spillover, promote product research and development, and realize upgrading in value chain by constraining technological diversity, accelerating technology diffusion and transfer [6]. It can be seen that the construction of technical standards is of great strategic significance to the manufacturing industry.

2.2 Technical standard opportunity identification

Sungjoo Lee et al. (2010) proposed a keyword-based patent map creation method, Identify potential technological innovation opportunities by looking for blank areas with sparse but large map pate nt density [7]. MitkovaL team (2015) and Marra A team (2016) further visual patent map construction methods based on semantic structure analysis, text analysis and principal component analysis, And to understand the blank area by combining the characteristics near the blank area[8,9]. However, the current research based on the standard literature information, taking into account the similar words, extracting the subject words to identify the standard blank area, understanding the blank area based on the surrounding core standards and identifying the future development opportunities is still not perfect.

Network co-occurrence diagram. Network co-occurrence is widely used in the future of technology development (Huang Lucheng et al., 2014) [10], technology opportunities and forecasting of evolution trends (Jeong, 2015) [11]. A scholar uses CiteSpace to map a social network of keywords, Research on technological developments, (Wang cuibo et al., 2020) [12].But in the current study, Research on effective access to core technical standards is inadequate, and how to combine the technical efficacy map to determine the more accurate future technology development direction after clarifying the blank area also needs to be deeply studied.

3 Identification of technical standards in manufacturing

3.1 Keyword extraction

(1) Generate text data. In this paper, based on the "Standard Data Base" and using the retrieval strategy of manufacturing industry as the subject word, the obsolete standard documents are manually cleaned out, and 207 relevant technical standard documents which are most in line with the characteristics of the manufacturing industry are selected, and the standards are merged into a complete document. The Jieba word breaker of Python is used for word segmentation. A lot of words with no important text meaning exist in the standard, but they appear frequently. They are called stop words, such as "or ", "yes" and so on. In the python, the edited stop word table is introduced to clean the data. Obtain text data for analysis.

(2) Initial keyword extraction based on LDA model. Using the Sklearn package in the Python software LDA model, the theme word extraction of the cleaned document is carried out by constructing the theme model to form the initial keyword set, as shown in Table 1.

Table. 1. LDA Initial keywords for topic word extraction.

theme	Initial keywords
1	equipment, material science, adhesives, tests, system, process, chemical raw materials, radio frequency, non-metallic minerals, technology
2	hygienic protection, clean, hygiene, technology, energy consumption, chemical raw materials, quality, reliability, systems, methods
3	information, automation, foundation, system, specifications, interfaces, data processing, control system, connect, hygiene
4	experiment, processing, rail transit, control, specification, connect, technical specification, technology, materials, automobiles
5	product, technology, specifications, systems, process, data, machining, quality, reliability, information

(3) Determination of similar words based on Word2Vec keywords. When using the LDA model to extract the theme words, we should set the number of theme K by ourselves. If the number of topics is too large, the theme is not representative, and if the number of topics is too small, some words will be repeated in multiple topics. To solve the problem that there may be too much or too little topic setting, this paper selects a non-repeating keyword as a candidate word under each topic on the basis of extracting the initial keyword by using the LDA model. Word2Vec is used to quantify words and extract words similar to candidate semantics. Determine the similarity of the initial keywords, as shown in Table 2.

Table. 2. Similar words for initial keywords.

Candidate	Similarities	semblance	Similarities	semblance	Similarities	semblance	Similarities	semblance	Similarities	semblance
1.Materials Science	Chemical raw material	0.812	Experiment	0.763	Equipment	0.721	Identification	0.698	Device	0.650
2.hygienic protection	Clean	0.877	Protect	0.757	Reliability	0.623	Energy consumption	0.602	Method	0.559
3.Information	Data	0.738	Data handling	0.693	Process	0.648	Foundation	0.618	Automation	0.579
4.Connect	Checkout	0.678	Integration	0.653	Process	0.594	Cementing	0.586	Control	0.563
……	……	……	……	……	……	……	……	……	……	……
n. Technology	Process	0.591	Consumer	0.585	Product	0.578	Coating	0.566	Quality	0.562

(4) Determination of representative keywords based on LDA models and Word2Vec. Under each theme, four keywords appeared in tables 1 and 2 are selected as the representative keywords of each theme. as shown in Table3.

Table. 3. Representative keywords table.

Theme	Representative Key Words
Materials and equipment	equipment, material science, experiment, chemical raw materials
Environmental Safety	hygienic protection, clean, energy consumption, reliability
Information and access categories	automation, foundation, data processing, information
Operating process classes	machining, control, connect, test
Technical Quality Category	product, technology, processes, quality
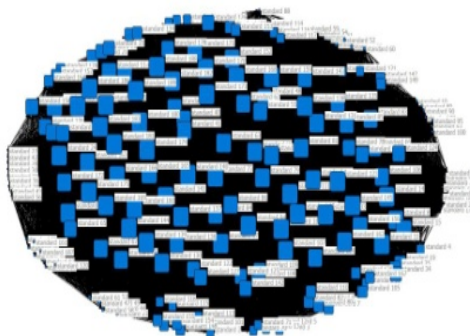
Fig. 2. Common diagram of manufacturing related technical standards network.

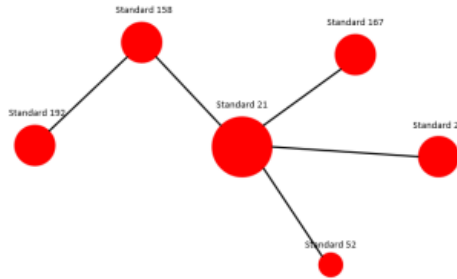## 3.2 Analysis of manufacturing advantage standards based on representative key words

Co-occurrence network centrality analysis. According to the frequency of 20 representative keywords in Table 3 in 207 manufacturing related technical standards, the standard frequency matrix of 189×20 is formed by deleting the standard (8 items) with keyword frequency of 0, Import the matrix into SPSS, Calculate the Euclidean distance between the standards, Form 189×189 standard square array. Import the standard square array into UCINET, Drawing the network co-occurrence diagram using NETDRAW module centrality, Visualizing output, as shown in figure 2.

By using the UCINET software analysis, the manufacturing standard center index is 12.4. From figure 2, we can see that the utilization degree center degree analysis cannot effectively judge the core standard.
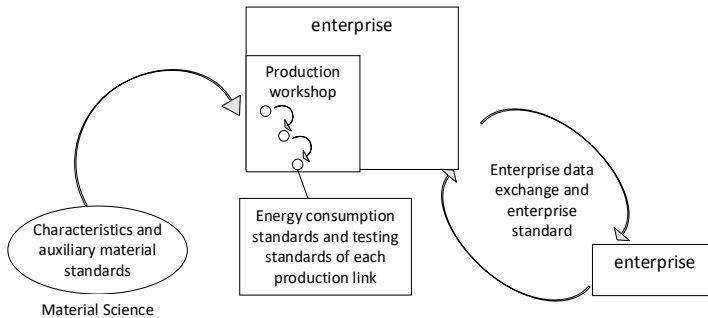
(2) Identification of core manufacturing standards. In Figure2 the number of nodes is large and the core advantage of nodes is not obvious. Therefore, adding "M-core" can eliminate the connections and nodes with low correlation degree, and finally obtain the highest connection strength node and determine it as the core standard. As shown in figure 3, the typical core standard in the diagram is shown in Table 4. The core standards in the manufacturing field mainly focus on the front-end manufacturing process, including the characteristic standards of materials and auxiliary materials required by manufacturing enterprises; the energy consumption standards and quality inspection standards of each link in the production and processing process of enterprises; in addition, it also includes the data exchange standards between enterprises, so as to help enterprises realize efficient and reliable data exchange, improve business efficiency and reduce operating costs, as shown in Figure 4 .

Table. 4. Typical core standards.

Name	Contest
Standard 21- Production Link Loop Monitoring Standard	This standard helps to better understand the contents of loop inspection and provides technical methods for loop inspection, including document inspection, visual inspection, functional inspection (using standard technology to test all components), and gives specifications for technical means to solve errors when errors are found. The following is the first example of this standard
Standard 158- Industrial automation systems; product data display and data exchange	In order to improve the operability of the data in the automation system, the standard specifies the adaptive testing technology of the automation system and the data display, exchange and storage technology
Standard 167- Energy consumption per unit production	Set the energy consumption quota standard of comparable unit output of products and standardize the energy consumption detection technology and energy consumption management technology of enterprises
Standard 192- Material Auxiliary Material Standard	The standard specifies the design and production techniques of materials and auxiliary materials required in the manufacturing process, and the application and inspection techniques of materials

Fig. 3. Manufacturing core technology standards link network.

Fig. 4. Flowchart of manufacturing advantage standards.

3.3 Manufacturing Standard Opportunity Identification and Trend Analysis

(1) The core area is determined based on the standard scatter diagram. The location of the red core standard in the scatter plot is extended to the area with low standard density and sparse distribution, and the surrounding standard of the connecting area forms four polygonal areas surrounded by yellow lines, that is, the technical standard blank area. This paper uses the Python PCA module to reduce the dimension of frequency square array (189×189). By calculation 189 standard scatter diagrams are drawn, as shown in Figure5.
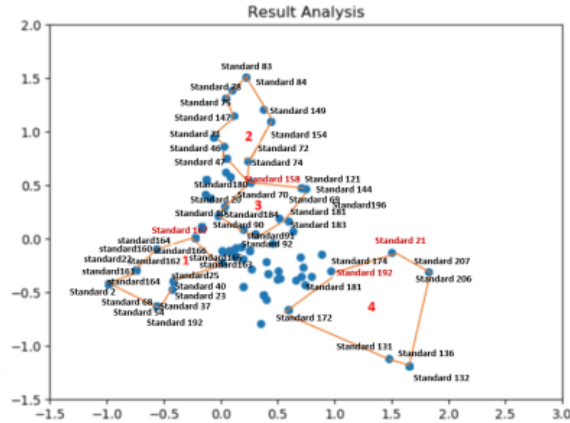
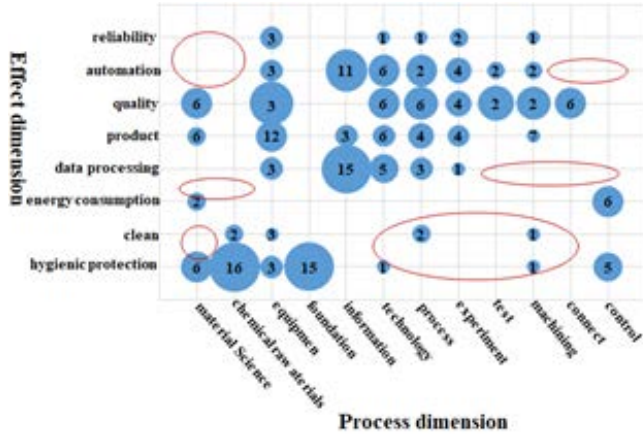**Fig. 5.** Blank area of manufacturing related technical standards.

Fig. 6. Technical efficacy chart.

(2) Technical standard opportunity identification based on technology-effect bubble diagram. Based on the technology-efficacy diagram, this paper determines more detailed and clear technical standard opportunities. First, according to the 20 representative keywords determined in Table 3, the corresponding words of technical process and functional effect dimension are determined, as shown in Table 5.

Table. 5. Dimension division of technical efficacy.

Dimension	Representative Key Words
Technical process dimension(resource + technology + process)	material science, chemical raw materials, equipment, foundation, information, technology, process, experiment, test, machining, connect, control
Functional effect dimension (result + state)	hygienic protection,data processing,product,energy consumption,quality,automation,clean,reliability

Secondly, using the technical process dimension as the horizontal coordinate and the functional effect dimension as the vertical coordinate python drawing the technical-efficacy diagram, as shown in Figure 6, the bubble size represents the frequency of the cross coordinate and the vertical coordinate representative keywords in 189 standards.

Finally, based on the technical efficacy diagram, the existing advantage standards of blank area are expanded, optimized and combined to determine the development opportunities of future manufacturing technical standards, as shown in Table 6.

Table. 6. Key construction contents and development opportunities of manufacturing industry.

Blank space	1—ecological environment	2—Data transfer and integration	3—Automation and interconnection	4—Inspection and Application of New Materials
Core standards	Standard 167	Standard 158	Standard 158	Standard 21 and 192
Identification of blank areas based on core standards	The region includes cleaner production, energy consumption, automated integrated systems environment and energy efficiency standards in different manufacturing industries	This area includes data generation standard, data transfer standard, data processing standard, control system and information system integration model standard, integrated system application implementation specification and dynamic server evaluation specification	In order to improve the operability of the data in the automation system, the standard specifies the adaptive testing technology of the automation system and the data display, exchange and storage technology	The region focuses on carbon substances, carbon materials standards and adhesives for linking materials, adhesive standards and monitoring standards for environmental stability during material processing
Blank area corresponds to technical efficacy dimension	clean, energy consumption (effectiveness dimension)	data processing (effect dimension)	automation (effect dimension)	material (process dimension)
Optimizing dimensions	material science, information, process, test, machining(Process dimension)	material science, machining, connect(Process dimension)	material science, connect, control(Process dimension)	reliability, automation(Effect dimension)
Optimization principles	Combination and Optimization: automation +clean technologies for green production；Process optimization	Integration and expansion: Data integration system and extended application outside the enterprise	local optimization: Forward optimization of material acquisition automation,Ensure the automation of different links	Expansion and integration: New materials for production, testing and other links; New Material Science+ measuring apparatus
Future development opportunities	Share of clean materials, process flow, Optimization of Production Link Operation and Inspection Standards, and full coverage information systems, Optimization of Big Data Analysis Standards for Green Production	Building data integration standards,In particular, systematized data management standards based on data expression, exchange, processing and integration are constructed in processing and inspection,Development and Improvement of Technology, Information and Resource Connectivity Standards for Supply Chain Node Enterprises	Refine front-end automation standards,Construction of automation standards in raw material acquisition, dynamic processing and inspection, improvement of automation standards in connection and control, and improvement of production and quality inspection efficiency	Construction of Technical Standards for Testing and Processing of New Materials and Development and Application of Measuring Instruments,Improve the reliability, accuracy and automation of measuring instruments and instruments

5 Conclusions

Based on the standard literature, this paper designs a systematic method to identify the technical standard opportunities of manufacturing industry. The LDA model is used to determine the initial keywords of manufacturing technical standards, and the similar words Word2Vec to determine the initial keywords are used to determine the five technical standards and 20 representative keywords of manufacturing materials and equipment, environmental protection and safety, information and utilization, operation process and technical quality based on the LDA model and Word2Vec to ensure the comprehensiveness and objectivity of text data mining. Secondly, based on 20 representative keywords, the combination of centrality and "M-core" is used to remove weak connection and isolated nodes, and the core technical standards of manufacturing industry are identified. The

Python PCA module is used to reduce the dimension of the frequency square array, draw the scattered plot, and determine the ecological environment--the standard blank area, data transmission and integration, automation and interconnection, and the inspection and adoption of new materials in combination with the core technical standards. Finally, the technology efficacy chart is used to identify the future technical standard opportunities of manufacturing industry, such as green production standard based on material and process, systematic digital management standard in processing and inspection stage, resource connection standard across enterprises, front-end automation standard, and technical standards for the development and application monitoring of testing measuring instruments.

References

1. Zou Siming, Zeng Deming, Zhang Lifei, et al. Network Relationship, Technology Diversification and Enterprise Technology Standardization Ability [J]. Scientific research management, 2017, 38(9): 12-20.

2. Peng Hailing, Wang Posen, Cheng Tao, et al. An Analysis of the Influence of Enterprise Standard Competition Strategy on Technological Innovation [J]. China Standardization, 2019, (4): 68-72.

3. Yu Ming Steel. Establishment of technological standards and the role of intellectual property in enterprise innovation [J]. Development Mass Standardization, 2017, (9): 63-65.

4. Zheng Ying, Han Shuo. Construction and Empirical Analysis of the Evaluation Model of Transforming Scientific and Technological Achievements into Technical Standards [J]. Science and technology management research, 2018, (23): 44-49.

5. Lu Meng. The role of technical standardization and technological innovation in economic development [J]. Technological Innovators,2013,(15):136.

6. Tao Zhongyuan, Wang Xiaoqing, Xue Chen. The Empirical Study on the Impact of Technology Innovation and Standardization Synergy on China's Manufacturing's Manufacturing Industry Value Chain--Taking Seven Subdivision Industries as Example [J]. Industrial technology economy, 2016, (3): 64-72.

7. SUNGJOO LEE, BYUNGGUN YOON, YONGTAE Park. An approach to discovering new Technology opportunities: keyword-based patent map approach [J]. Technovation, 2010, 6(29): 481-497.

8. Mitkova L, Wang X. Combining SAO semantic analysis and morphology analysis to identify technology opportunities: a case in Dye-Sensitized Solar Cells [J]. Post-Print, 2015, 34(6): 34-45.

9. Marra A R, Monte J C, Prado M, et al . SYSTEM AND A METHOD FORMONITORING HAND SANITIZING OPPORTUNITIES IN HOSPITAL WARDS, US20160125723[P]. 2016.

10. Huang Lucheng, Zhang Jing. A Study on the Identification Method of Industry Common Technology Based on Patent Analysis [J]. Science and Technology Management, 2014, (4): 80-86.

11. Jeong S, KimJC, Choi J Y. Technology convergence: What development stage are we in?[J]. Scientometrics, 2015, 104(3): 841-871.

12. Wang Cuibo, Xiong Kun, Liu Wenjun. Visual Analysis [J]. Technology Prediction Research Based on CiteSpace Technical economy, 2020, 6(39): 147-154.