Summer 7-26-2022

Research on the Evolution of Logistics Identification Marking Technology Based on Mallet-LDA Model and Social Network

Xiaoli Li
School of Management, Wuhan Textile University, Wuhan, 430073, China, linderde@163.com

Yan Kuang
School of Management, Wuhan Textile University, Wuhan, 430073, China

Full Research Paper

Research on the Evolution of Logistics Identification Marking Technology Based on Mallet-LDA Model and Social Network

Xiaoli Li[1*], *Yan Kuang*[1]

(School of Management, Wuhan Textile University, Wuhan, 430073, China)

Abstract: The purpose of this paper is to propose a technology evolution analysis framework based on the Mallet-LDA models combined Social Network Analysis(SNA) method and to study the development trends of critical technologies in logistics identification marking technology. This paper establishes a complete and systematic framework for analyzing the technological evolution based on the logistics identification marking technology selected from the national Key industry patent information service platform in China as the data source database for the period 1993 to 2018 and uses topic evolution and SNA methods for evolution analysis to predict the development trend of the critical technologies in the future time. Results show that Mallet-LDA model combined SNA method is feasible for technology evolution analysis. In addition, the main development trend in logistics identification marking technology is concentrated in the direction of data collection and processing, drone logistics, cold chain logistics, intelligent terminals. And the technical topics have changed from the early days with the primary label, barcode technology, and other individual areas to the development state of multi-technology integration. The findings of this study are expected to enrich technology evolution research based on patent analysis and provide insights into the potential of using patent data for technology evolution development prediction in the context of smart logistics in China.

Keywords: LDA; Patent analysis; Technological evolution; Logistics identification marking technology; SNA

1. INTRODUCTION

Since the outbreak of COVID-19, mature logistics operations have played an essential role in the supply and deployment of emergency supplies. At the same time, because the international trade friction and the global epidemic continued, China's supply chain flexibility and flexibility to enhance the demand has been imminent. In 2015 China's government put forward the "Made in China 2025" development strategy, which sets intelligent manufacturing as the main direction of China's manufacturing industry[1], which requires the transformation of the whole cycle, the entire movement, and the whole chain with the use of digital intelligence technology. Smart logistics is based on intelligent manufacturing and runs through the entire supply chain process. It can be seen that for the logistics industry and even the entire national economy, the strategic decision to build an intelligent logistics supply chain system has a far-reaching impact[2].

Intelligent information technology is a crucial way to realize the transformation from traditional logistics to digital logistics[3], among which logistics identification marking technology is the crucial intellectual technology field for real-time control and management of the logistics supply chain. The technology is mainly composed of radio frequency identification (RFID), automatic identification, logistics bar-code symbology, symbol printing, electronic data interchange and other technologies. It realizes automation of merchandise retailing (POS), stocking, inventory replenishment, sales analysis and other business operations through standardized coding and symbol representation of trade objects operating in logistics. Thus, it achieves the goal of dynamic real-time management of logistics and supply chain.

With the surge of massive objects of logistics operation and the demand for digital logistic transformation,

* Corresponding author. E-mail: linderde@163.com (Xiaoli Li)

logistics identification marking technology has higher requirements, such as mobility and multi-objective identification. The popularization of China's logistics identification marking technology requires logistics enterprises to introduce advanced information technology and sensing technology, making the fundamental realization of intelligence in commodity storage, transportation, packaging, distribution, supply chain, and other links. The technology is in line with the development trend of modern logistics automation, intelligence, real-time and visualization. It also caters to the development trend of Internet of Things technology. However, how to scientifically judge the specific process of logistics identification marking technology evolution? This requires a systematic framework for analyzing the evolution of the technology. Patent literature carries over 90% of the world's technical information[4]. With the gradual popularization of machine learning theory, it means that machine learning algorithms can be introduced to identify and analyze the theme and evolution of logistics identification and marking technology patent literature. This will help to clarify the direction in which existing logistics identification technologies are moving and to strengthen future technological innovation activities in this technology area.

In this paper, we focus on the need for addressing the shortcomings of the current research methods for analyzing technological evolution in the field of logistics identification marking, proposing a combination of the Mallet-LDA models and social network analysis to fill the gap in the work of technological evolution in the field of logistics. Meanwhile, it is of practical significance for logistics enterprises to determine the R&D direction and layout of key technology research early. By analyzing the evolution of logistics identification marking technology in China, this study can qualitatively and quantitatively grasp the latest research hotspots and the latest innovation achievements in this field from a macro perspective. And we can also quickly understand the future development direction of this technology. Therefore, it can guide the optimal R&D direction for Chinese enterprises in the process of transformation from backward to beyond. And it allows related companies to do a good job of patent layout in front of emerging technologies. Therefore, this study has crucial practical significance and application value for enterprise R&D departments and strategy makers.

2. REVIEW OF RELEVANT LITERATURE

At present, domestic and foreign scholars have formed a series of research results on logistics identification marking technology. On the one hand, from the practical research perspective of applied research and development, scholars focus more on performing logistics identification marking subsystem and the application research of related technology under the qualitative perspective, such as Liao et al. proposed a calculation method of port loading and unloading efficiency based on ship automatic identification system, which can monitor port throughput in real-time and enhance the digital management level of port[5]. Recognizing the importance of contactless delivery, Fu et al. proposed a mathematical model to make automated delivery robots more flexible[6]. On the other hand, from the perspective of theoretical research on development mode and path, scholars have already analyzed the current situation of logistics identification marking technology development by quantitative methods, Wang et al. combined patent maps to explore the problems of logistics identification marking technology innovation process in terms of time change trend, technology maturity, innovation subject, and main IPC distribution structure[7].

Patent documents are the most effective carriers of technical information and innovation activities. To explore the technological trends hidden in patent texts, the traditional approach is usually to examine representative patents based on expert experience to get the technical subject content. However, two obvious shortcomings of such methods are that patent representativeness is difficult to measure and that it is difficult to analyze numerous patent documents. Related studies have explored many technical subject discovery methods. Applying the improved PageRank algorithm to the number of citations and application year of patents, this combination of research aims to identify the core patents of domain technologies[8]; Based on the study of patent citation features, a patent citation network is first mapped out, and the construction of this network provides the possibility of using the main path analysis algorithm, and the combination of the two enables the discovery of technical topics[9]. The work applies the LDA topic model to discover technology topics and model the change of topics over time, identifies paths by analyzing the change of technology topics in different segments, and presents a visualization of technology topics trends[10].

From what we have discussed above, the previous research in logistics identification marking technology in China is mostly related to some subfields, and there are few research results in sorting out the overall logistics identification marking systematic technical system; there is a lack of quantitative analysis, especially the analysis of technological topics evolution development trend based on text mining. Based on the above research situation, this paper scientifically divides the technological topics in logistics identification marking by constructing a topic mining analysis model, analyzes the thematic characteristics of logistics identification marking technologies according to the division time slice method, and finally introduces the social network analysis method to analyze the development relationship between the technical clusters, to provide a reference for realizing the high-quality, sustainable development of intelligent logistics. It can also provide relative information for researchers and patent managers in logistics. It also provides relatively rich intelligence and a certain reference basis for researchers and patent managers in logistics.

3. RESEARCH METHODOLOGY AND FRAMEWORK

3.1 Mallet-LDA topic probability model

Topic model is an algorithm to extract useful information from a textual corpus. LSA and pLSA are early topic models that decompose documents into potential vector representations using SVM and probabilistic models respectively[11]. The Latent Dirichlet Allocation (LDA) algorithm, proposed by Blei et al., is a set of algorithms for unsupervised learning to extract useful information from a text corpus algorithm for extracting useful information from a text corpus in an unsupervised learning manner in 2003[12]. The algorithm automatically discovers potential topic structures from a document or text corpus using a Bayesian hierarchical model, where each topic is determined by the probability distribution of words. In simple terms, we derived potential topics from topic probabilities based on document distributions and word probabilities based on topic distributions.

The LDA algorithm can significantly reduce the data dimensionality of large-scale documents, thus facilitating the extraction of potential subject terms that encompass various perspectives. Specifically, the bag-of-words model is used to transform documents into word-frequency vectors and map the complex high-dimensional lexical space of unstructured documents into the low-dimensional space of "document-topic-lexicon " to uncover the hidden topics of patent texts. In this paper, we assume that the potential topics in patent documents obey the hyper-parametric Dirichlet prior distribution.

$$Dir(\theta_d \mid \alpha) = \frac{\Gamma\left(\sum_{k=1}^{K} \alpha_k\right)}{\prod_{k=1}^{K} \Gamma(\alpha_k)} \prod_{k=1}^{K} \theta_{dk}^{\alpha_k - 1} \tag{1}$$

where θ_{dk} denotes the probability distribution of patent literature d in topic k. The larger the parameter α the more obvious the theme will be. The topic-vocabulary distribution is $\phi_k \sim Dir(\beta)$; the document-topic distribution is $\theta_d \sim Dir(\alpha)$. Generate topic $z_{dn} \sim Multinomial(\theta_d)$ and vocabulary $w_{dn} \sim Multinomial(\phi_{z_{dn}})$ for the n-th word item in each patent document. Thus, the LDA likelihood model in this paper is shown in (2) as follows:

$$p(W \mid \alpha, \beta) = \prod_{d=1}^{D} \int p(\theta_d \mid \alpha) \prod_{n=1}^{N_d} \sum_{z_d} p(z_{dn} \mid \theta_d) p(w_{dn} \mid \phi_{z_{dn}}) d\theta_d \tag{2}$$

The Mallet toolkit is dedicated to machine learning task processing. Using the Mallet package, natural language processing, text classification, clustering, topic modeling, and information extraction can be efficiently performed. The Gensim toolkit based on the Python language provides a wrapper for implementing Mallet's LDA algorithm (Mallet-LDA model) inside Gensim. To ensure the effectiveness of LDA topic clustering, this paper chooses to optimize Gensim's built-in LDA algorithm with Mallet.

3.2 Topic strength measures

Since LDA only provides the topic extraction algorithm for patent text, it does not involve the specific evolution mechanism and the explanation of topic evolution, and the greater the topic intensity, the higher attention to the technology[13], so this paper shows the evolution trend of logistics identification hot technology based on the change of topic intensity in the time axis. Referring to the topic intensity calculation method provided by Li et al.[14], we showed the calculation formula as follows:

$$Q(Z_{t,k}) = \frac{\sum_{d=1}^{D_t} \theta_{d,k}}{D_t} \tag{3}$$

In (3): $Q(Z_{t,k})$ denotes the intensity of the subject *K* matter t in the current time slice; $\theta_{d,k}$ represents the probability of the first *K* subject matter in the first d patent document; *Dt* represents the number of all patents on the time slice *t*.

3.3 Research framework

We aim to identify logistics identification marking technology topics and reveal their dynamic evolution process with the help of the Mallet-LDA topic model, and further demonstrate the technology development characteristics and future development direction. The paper mainly includes the pre-logistics identification marking technology patent data collection and pre-processing, logistics identification marking technology topic mining based on the Mallet-LDA model and evolutionary results, trend presentation, etc. Subsequently, the social network analysis method is introduced to visualize the mined hot topics and further explore the relationships within and among the technical topics. The specific process of topic evolution analysis is shown in Figure 1.

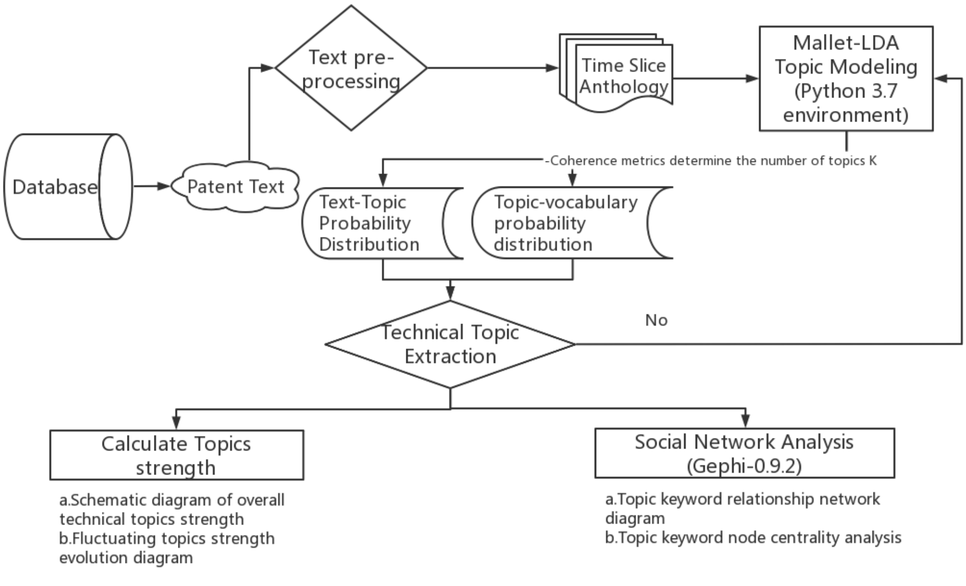


Figure 1. Schematic diagram of the general research framework

4. EMPIRICAL ANALYSIS

4.1 Patent search

In this paper, we use the Patent Information Service Platform to select the "identification technology" under the logistics industry for retrieval. In the first retrieval in November 2020, 235,475 invention patents and utility model patents are preliminarily retrieved. As this paper mainly studies the evolution of identification technology in logistics, we conduct a secondary screening basis on the above retrieval, enter “logistics or supply chain” in the search box of the abstract field. Meanwhile, considering the 18-month lag period between patent application and disclosure, this paper selects the relevant patent application data up to 2018. In this way, we retrieved 4952 invention patents and utility model patents for logistics identification marking technology. To ensure the accuracy of the data source, all the captured data were standardized and sorted out, and we eliminated the repeated patents. Finally, 4857 valid data were got.

4.2 Data pre-processing

Data pre-processing is one of the most important aspects of LDA topic modeling. This paper link the training efficiency of the model and the accuracy of the topic clustering results to this process. First, the jieba package of Python is called to carry out preliminary word separation of text data. However, because of the strong professionalism of the patent text, by analyzing the patent data, this paper constructs a patent dictionary of logistics identification marking technology, a deactivation word list, and manually processes the word separation results to guarantee the correctness of the generated topics and provide good experimental data for subsequent modeling.

4.3 Logistics identification marking technology topic mining

According to previous studies, the hyper-parameter $\alpha = 50/K$, $\beta = 0.1$ are set after considering the topic granularity and the readability of the results in the LDA topic modeling process. The number of topics in the LDA model K has a significant influence on the effect of the final topic extraction. In this paper, the interval value of the number of topics is set to [0, 30] regarding the topic settings in previous literature. By invoking the built-in LDA model of Gensim and running the Mallet-LDA model, the consistency index of both is calculated. In general, the model is better when the consistency value is more significant. Therefore, comparing Figure 2 and Figure 3, we can see that the Mallet-LDA model is better and has the optimal number of topics $K = 9$.

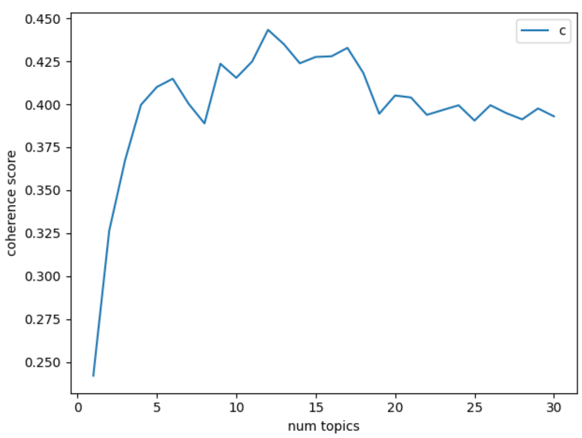


Figure 2. Coherence score based on traditional LDA

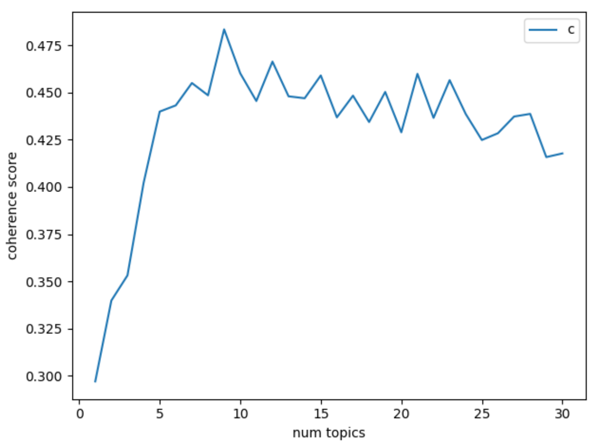


Figure 3. Coherence score based on Mallet-LDA model

Mallet-LDA algorithm is used to carry out topic modeling according to the above optimal number of topics, and the meaning of each topic is summarized according to the high probability feature keywords on each topic. Nine topics mined by Mallet-LDA topic modeling are a collection composed of feature words with a high likelihood of appearing in logistic identification technology and are interrelated. Manual ranking finds that the keywords of Topic2 and Topic7 are in different directions in RFID technology but belong to the same technical field, so the two topics are combined into the same topic. Finally, eight technical topics are identified. The results of topic mining are shown in Table 1.

Table 1. Logistics identification marking technology topic mining results

Topic	Keywords	Technical Topic Name
1	Sorting, Sensors, Automatic identification, Automation, Control System, Shelves, Robotics, bar-code	Logistics Automation Sorting
2	Electronic labels, RFID, Storage, Read, RF, Efficiency, Testing, Logistics box	Radio Frequency Identification (RFID)
3	Commodities, Bar-Code, Computers, Camera, Scanning, Image Recognition, Packages, Inspection	Data acquisition and processing
4	Information Technology, Testing, Logistics box, Automated transportation, Drones, Cost, Load, Self-help	Drone Logistics
5	Data, Monitoring, Real-Time, Internet of Things, Agricultural Products, Temperature and humidity, Cold Chain Logistics, Data Transfer	Cold Chain Logistics
6	Self-adhesive, ic chip, Digital, Logistics Automation, Marking, Records, Technology, Paper	Labeling Technology
7	Enterprise, Commodities, Traceability, Anti-counterfeiting, Block-chain, Supply Chain, Cloud Server, Database	Block-chain technology anti-counterfeit traceability
8	Logistics Marking, Terminal, Servers, Platform, Smart, Clients, Services, Client	Smart Terminal

4.3.1 Topic intensity evolution analysis

The evolution of subject intensity fully reflects the dynamic nature of the subject and its development status. The higher the subject intensity, the more patent applications and the higher the research intensity of the patent technology in the corresponding time window. Based on the document-topic and topic-word probability distributions given by the Mallet-LDA topic model, the above eight technical topic strengths were calculated by Equation (3). The intensity of each technical topic is visualized and analyzed from two dimensions[15]: on the macro level, the intensity of each technical topic is displayed in the overall year as the unit of measurement, as shown in Figure 4; on the micro-level, we calculate the intensity of each topic for each year from 2011 to 2018. On the micro-level, the intensity value of each topic in each time slice from 2011 to 2018 is calculated, and the curve of topic intensity over time is plotted. We show the trends of topic intensity in Figure 5 and Figure 6.

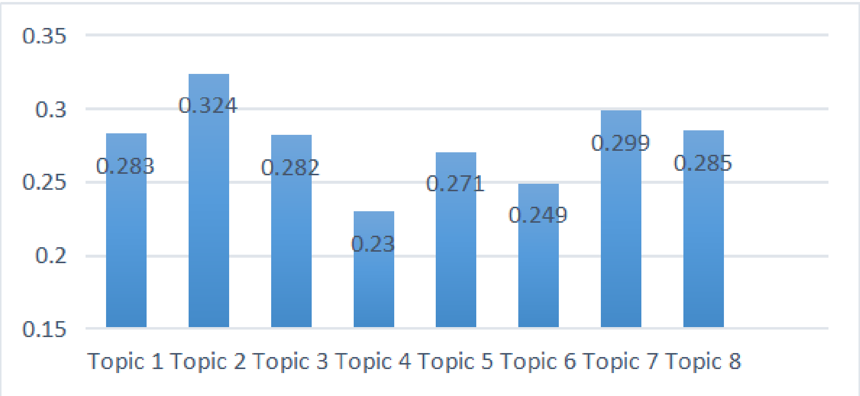


Figure 4. Schematic diagram of the overall technical theme strength

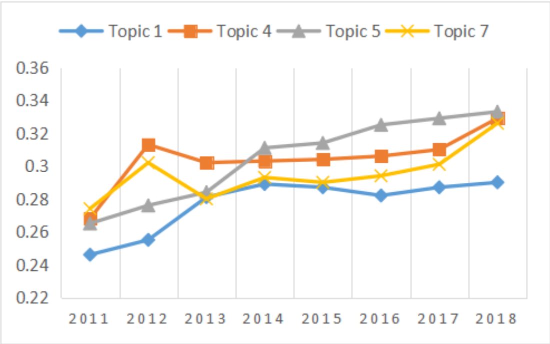


Figure 5. Technical topics uptrend

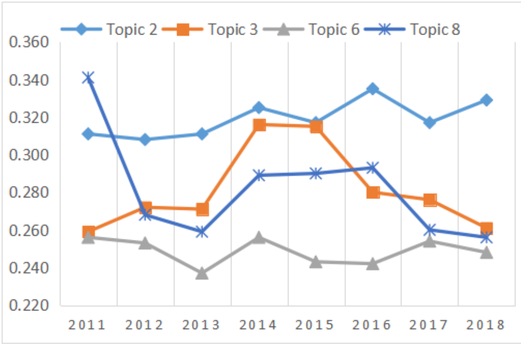


Figure 6. Technical topics wave trend

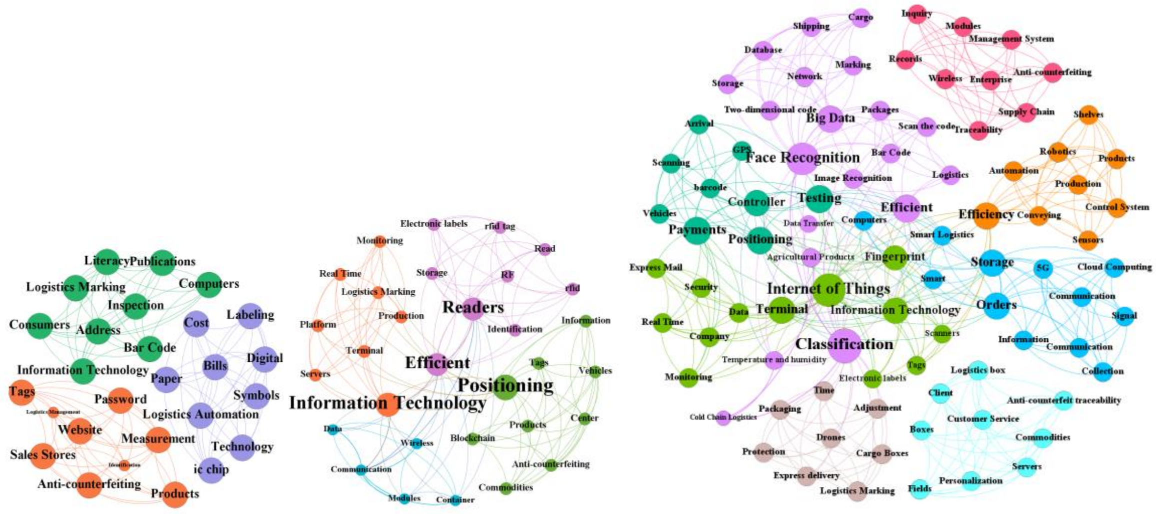
Figure 4 demonstrates the intensity distribution of each technology topic based on the overall year of mining. Topic2 has a higher intensity and is the key focus direction, followed by Topic1, Topic3, Topic5, Topic7, and Topic8, while Topic4 and Topic6 are less intense and weaker than other topics in widespread attention.

Figure 5 shows the intensity change of the four topics mined in the entire text set, all of which offer an upward trend. Among them, the intensity of Topic4 (drone logistics) and Topic7 (block-chain technology anti-counterfeit traceability) offer the same trend of change; Topic1 is one of the technical conditions necessary for advanced logistics parks. With the rise of cloud computing, artificial intelligence, and intelligent logistics concepts, the importance of automated logistics sorting has become more prominent, and the research enthusiasm has accumulated; Topic5 research heat basically shows a straight up trend. Cold chain logistics is a low-temperature logistics process. Because of the improvement of the living standard of the masses, the demand for the purchase of fresh agricultural products has surged. This makes cold chain logistics gain more and more attention.

Figure 6 is the change situation of 4 topics with fluctuating intensity. Topic2 is the core and basic technology of logistics identification marking technology. The application field is vast, so RFID technology has been the research hotspot of logistics identification marking. Its subject intensity also presents fluctuating upward trend on a higher level. The initial technical intensity of Topic6 is not high, and also is not the focus of logistics identification marking technology exploration, swinging down on a lower level; Topic3 is to receive data at the same time and to process the data and output, to simplify data input, tracking commodity logistics, etc. to achieve the bar-code scanning, decoding and data communication with sensors and other terminal devices. Because of the continuous development and improvement of basic technology, the research enthusiasm of both topics shows the trend of fluctuation in the early stage and a decline in the later stage.

4.3.2 Network evolution analysis of technical topic relationships

For the patent data of three-time slices (1993-2001, 2002-2012, and 2013-2018), Mallet-LDA modeling training was conducted to get the optimal topic of the corresponding time slice. According to the results of topic mining, the top ten keywords of each topic in each period are extracted. Then, the matrix formed is imported into the Gephi tool to draw the hot topic keyword relation network in three stages, with topic keywords as nodes and their co-occurrence relations as edges. Finally, the community optimization discovery algorithm based on modularity_class was clustered to form the three-stage thematic keyword relationship network graph, as shown in Figure 7. The larger the node size is, the more paramount it is and the more mature it is.



(a)1993-2000（Logistics 2.0 stage） （b）2001-2011（Logistics 3.0 stage） （c）2012-2018（Logistics 4.0 stage）

Figure 7. Three-stage key technology evolution trend

To measure the evolution of patented technologies in logistics identification marking at each stage, the study uses the Gephi tool to calculate social network indicators. Based on the results of the computed indicators, the centrality value of sorting technology is more significant, showing that sorting technology in logistics identification marking in China has technical advantages and is a key core technology, which combined with Figure 9, can be seen to be closely linked and integrated with data collection, Internet of Things and other technologies. In the early stage of China's logistics industry, the development of label and bar-code technology was slow, and the digitalization of logistics identification marking was a weak link. But from 2012 to 2018, the centrality of technologies such as the Internet of Things, cloud computing, and artificial intelligence increased. This reflects that China has improved the layout of patents, making technology diversified development.

From three stages, the intermediate center value of the leading technology node increases and the near-neutral value decreases. This shows that, with the expansion of the scale of technological development, the overall communication distance of the network is shortened, and the connection and integration between technologies are improved. This shows that the innovation mode of China's logistics identification marking technology has changed from the previous development of single-core technology to the cluster R&D mode of multi-technology integration. From the evolution of the technology node, the logistics recognition identification from the early basic label, bar-code, development to and electronic label, automatic diagnosis for combination, and then to face recognition, cold chain logistics, block-chain and robotics, and other fields between the interpenetration, applied to more different scenarios. Thus, the intelligent and digital development of logistics identification marking technology is the general trend.

5. CONCLUSIONS

Because of the current demand for technological innovation from the development of intelligent logistics, the research is based on 4,857 patent application data of logistics identification technology in the logistics industry collected by the National Key Industry patent Service platform from 1993 to 2018. This paper proposes a method to combine the mallet-LDA topic model and social network index, which is superior to traditional LDA topic clustering. The optimal number of topics was determined at the macro overall level, and the calculation formula of the topic intensity measured the variation rule of each technical topic intensity. From the micro perspective, we carried theme clustering out for three-time slices (1993-2001, 2002-2012, and 2013-2018). Gephi software is used to draw the evolution trend chart of hot technologies in three stages, and the overall development trend of logistics identification marking technology in China and the relationship between key technologies are explored. In addition, the following research conclusions are drawn:

1)In this paper, the optimal number of topics divided by using the Mallet-LDA topic mining analysis model yields more reasonable results By comparing the Mallet-LDA topic mining analysis model with the traditional LDA model, Mallet-LDA shows a large consistency index, which indicates a more reasonable number of topics. Logistics identification marking technology topic mining results are specifically: logistics automated sorting, radio frequency identification(RFID), data acquisition and processing, drone logistics, cold chain logistics, labeling technology, blockchain technology anti-counterfeit traceability, and smart terminal.

2)The overall technology topic intensity shows a stable upward trend and two development patterns of up and down fluctuations. According to the analysis of the evolution of technology topic intensity, it is concluded that: logistics automatic sorting, drone logistics, cold chain logistics, and blockchain technology anti-counterfeit traceability show a stable rising trend and are the technology research hotspots for a while in the future; radio frequency identification technology (RFID), data collection and processing, label technology and intelligent terminal are in an up-and-down fluctuating state, especially the topic intensity of label technology always stays at a relatively low position, showing that the technology is more basic; while radio frequency identification technology, although also in a state of up and down fluctuation, but the topic of the technology has always remained in a relatively high position, indicating that although the technology is not a hot research direction, in the next period the research enthusiasm will not be significantly reduced.

3)According to the measurement index of the subject nodes in the three-stage technology evolution network, it is concluded that the innovation mode of China's logistics identification marking technology has changed from the previous development of single-core technology to the cluster R&D mode of multi-technology integration. As the intermediary center value of the nodes increases and the near-neutral value decreases, the technology topic changes from the earlier single field with label and bar-code as the core technology to the development state of multi-technology integration, and the overall network becomes dense, and it also shows that the communication efficiency between technologies increases. This suggests that R&D personnel should not only focus on the core technologies of the industry but also actively explore crucial industries and key technologies that promote the development of industrial correlation. Thus, new technologies can be created by combining different industries to further promote industrial development.

However, there are still some limitations in this study: the processing of patent text in this paper is demanding, and how to optimize the word separation steps and model configuration parameters is a problem that needs attention in subsequent research; in addition, relying solely on patent text data for decision making is also imperfect, and other data, such as scientific and technical literature and industrial reports, should be combined, so that it is possible to grasp the future direction of industry and technology development more accurately.

ACKNOWLEDGEMENT

This research was supported by the National Social Science Foundation of China under Grant 19BGL040.

REFERENCES

[1] Zhou Ji. (2015).Intelligent manufacturing - the main direction of "Made in China 2025"[J]. China Mechanical Engineering, 26(17): 54-55(in Chinese)

[2] Liu, W., Shanthikumar, J. G., Lee, P. T. W., Li, X., and Zhou, L. (2021). " Special issue editorial: Smart supply chains

and intelligent logistics services”, Transportation Research Part E: Logistics and Transportation Review, Vol. 147, pp. 102256.

[3] Mikl, J., Herold, D. M., Ćwiklicki, M., and Kummer, S. (2020). “ The impact of digital logistics start-ups on incumbent firms: a business model perspective”, The International Journal of Logistics Management, Vol. 32 No. 4, pp. 1461-1480.

[4] Krestel, R., Chikkamath, R., Hewel, C. and Risch, J. (2021), “A survey on deep learning for patent analysis”, World Patent Information, Vol 65, p. 102035.

[5] Liao, L.F., Le, F.G. and Zhu, Y.L. (2017), “ Application of LDA model in patent text classification”, Modern Intelligence, Vol 37 No 3, pp. 35-39.

[6] Fu, C., Zhu, N., Ma, S., and Liu, R. (2021), “ A two-stage robust approach to integrated station location and rebalancing vehicle service design in bike-sharing systems”, European Journal of Operational Research, Vol 298, pp. 915-938.

[7] Wang, C.L., Wang, Y. and Li, C.C. (2014), “ Research on technological innovation of logistics identification marking based on patent map”, China Science and Technology Forum, Vol 10, pp. 28-34(in Chinese)

[8] De Paulo, A. F., Ribeiro, E. M. S. and Porto, G. S. (2018), “ Mapping countries cooperation networks in photovoltaic technology development based on patent analysis”, Scientometrics, Vol. 117 No. 02, pp. 667-686.

[9] Ji, J., Barnett, G. A., and Chu, J. (2019), “ Global networks of genetically modified crops technology: A patent citation network analysis”, Scientometrics, Vol. 118 No. 3, pp. 737-762.

[10] Hofmann, P., Keller, R., and Urbach, N. (2019), “ Inter-technology relationship networks: Arranging technologies through text mining”, Technological Forecasting and Social Change, Vol. 143, pp. 202-213.

[11] Np A, Mvbts A, Lrb B, et al. Women's e-commerce clothing sentiment analysis by probabilistic model LDA using R-SPARK[J]. Materials Today: Proceedings, 2021.

[12] Blei, D.M., Ng, A. and Jordan, M. (2003), “ Latent dirichlet allocation”, Journal of Machine Learning Research, Vol 3, pp. 993-1022.

[13] Li, Z.B., Zhang, L. (2020), “ Study on the evolution of patent technology topics in steel materials based on LDA model”, Science and Technology Management Research, Vol. 40 No. 24, pp. 175-183.

[14] Li, X.D., Zhang, J. and Yuan, M. (2014), “Research on topic evolution of science and technology journals based on LDA model”. Journal of Intelligence, Vol 33 No 7, pp. 115-121

[15] Keller, T.R., Hase, V. and Thaker, J. (2020), “ News media coverage of climate change in India 1997–2016: using automated content analysis to assess themes and topics”, Environmental Communication, Vol. 14 No. 2, pp. 219-235.