

Evolution of Research Hotspots in China's Logistics Industry Based on Cluster Analysis

Dan Liu^{1,*} Yanfeng Fang^{2,3}

¹School of Economics and Management, Fuzhou University of International Studies and Trade, Fuzhou, Fujian 350202, China

²Fujian Institute of Scientific and Technological Information, Fuzhou, Fujian 350001, China

³Fujian Key Laboratory of Information Network, Fuzhou, Fujian 350001, China

*Corresponding author. Email:liudan@fzfu.edu.cn

ABSTRACT

Topic is a highly condensed content of the subject research. Word frequency analysis, time series and cluster analysis can reveal the core content, research hotspots and directions of the subject research. In order to detect the research hotspots and direction trends of China's logistics industry, this paper obtains 2,342 academic topics from 2011-2019 from the official website of China Society of Logistics. Through Chinese words segmentation and text mining, 91 high-frequency words are obtained. The paper conducts time series and cluster analysis of hot words and finds that the society subject shows the characteristics of not obvious differences in themes, strong interdisciplinarity, attention to national strategy and policy guidance, and attention to public events and so on. It also points out that future research topics should focus on differences, continue to play to the interdisciplinary nature and pay attention to the application of new technologies.

Keywords: *China Society of Logistics, research topics, hotspots, time series analysis, cluster*

I. INTRODUCTION

China Society of Logistics is a national logistics academic research organization approved and registered by the Ministry of Civil Affairs. It has the status of juridical association and is headquartered in Beijing. There are more than 10,000 members, covering all sectors of politics, industry, academia and research in the logistics field, with strong resource integration capabilities and extensive social influence. The research topics of the China Society of Logistics play an active and important role in integrating social research forces, coordinating and guiding research directions, advancing academic theoretical innovation, and promoting the integration of production, education and research [1]. Its research topics and hotspots not only represent the research direction of various fields of politics, industry, academia and research in the logistics field, but are more likely to become the vane guiding logistics research and practice in the future.

The topic selection is a highly condensed content of the subject research and a concentrated reflection of the subject research. By constructing the weight matrix of topic, hot spots can be found, and time series and cluster analysis can be performed on them, which can reveal the core content of the subject research to represent the research hotspot and direction [2].

Looking back at the literature, it is found that the research of hotspot analysis starting from keywords is common in China and foreign countries, but this method is rarely reported in the research of societies and associations, and there is currently a lack of research on hot trends. This paper uses the weight analysis method to reveal the current status and hotspots of the China Society of Logistics, and analyzes the major policy adjustments in the logistics industry to reveal the relationship between the hotspots and the national strategic orientation.

II. MATERIALS AND METHODS

A. Research materials

This research is based on the project that was published on the official website of the China Society of Logistics from 2011 to 2019.

B. Research methods

1) *Search method:* This research is based on the notification document issued by the China Society of Logistics over the years on issuing the "Research project plan of China Society of Logistics and China Federation of Logistics and Purchasing", and a total of 2,342 projects have been approved by the society.

2) *Data processing process:* The data processing process is divided into the following steps:

CLC Number:F253.9

- to extract data, merge entities, build a stop vocabulary, and complete data cleaning;
- to use Python software to compile a program to analyze the year, application unit, and name of the 2,342 academic projects, and get the school distribution and the word frequency of the project.
- for the extraction of high-frequency keywords, high-frequency keywords should be obtained according to Zipf's second law [3], and then the corresponding matrix will be generated, exported and saved to a document in xlsx format.
- to use PivotTable and function tools to process the keyword data and get corresponding charts.

III. DATA PROCESSING RESULTS

A. General situation

This research has obtained a total of 2,342 topics from the China Society of Logistics in the past five years. Since 2014, it has distinguished major topics, key topics and general topics; Among them: there are 13 major projects, 38 key projects, and 2,291 general projects. The application units cover all sectors of industry, education and research including universities, enterprises, military, and research institutes.

B. Years distribution

The number of topics and year distribution of the China Society of Logistics from 2011 to 2019 are shown in "Table I" and "Fig. 1".

TABLE I. NUMBER OF TOPICS OF CHINA SOCIETY OF LOGISTICS FROM 2011 TO 2019

Year	2011	2012	2013	2014	2015	2016	2017	2018	2019
Quantity	208	227	267	279	277	289	239	280	276

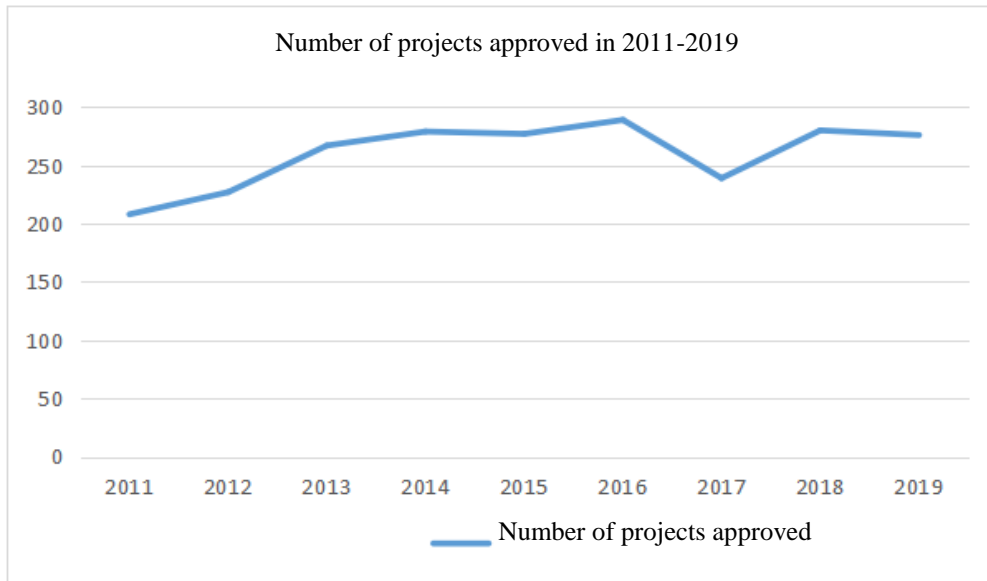


Fig. 1. Number of projects approved by China Society of Logistics in 2011-2019.

C. Distribution of approved project units

Judging from the categories of approved project units, from 2011 to 2019, 978 project units have been approved by China Society of Logistics, many of which apply as secondary units and enterprises in the name of branches. After combining the above data, 576 application units are obtained. It includes 296 universities, 120 junior colleges, 11 secondary technical schools, 10 military academies, 27 scientific research institutes (studios), 96 enterprise groups or companies,

and 16 industry associations or societies. Colleges and universities accounted for 72.45%.

From the perspective of the number of approved projects, the units that have applied for the largest number of academic projects are mainly finance and economics colleges, military academies, and professional research institutions. Among them, 10 military academies have a total of 239 approved projects. Among the top 10 units, there are 4 military academies, as shown in "Fig. 2".

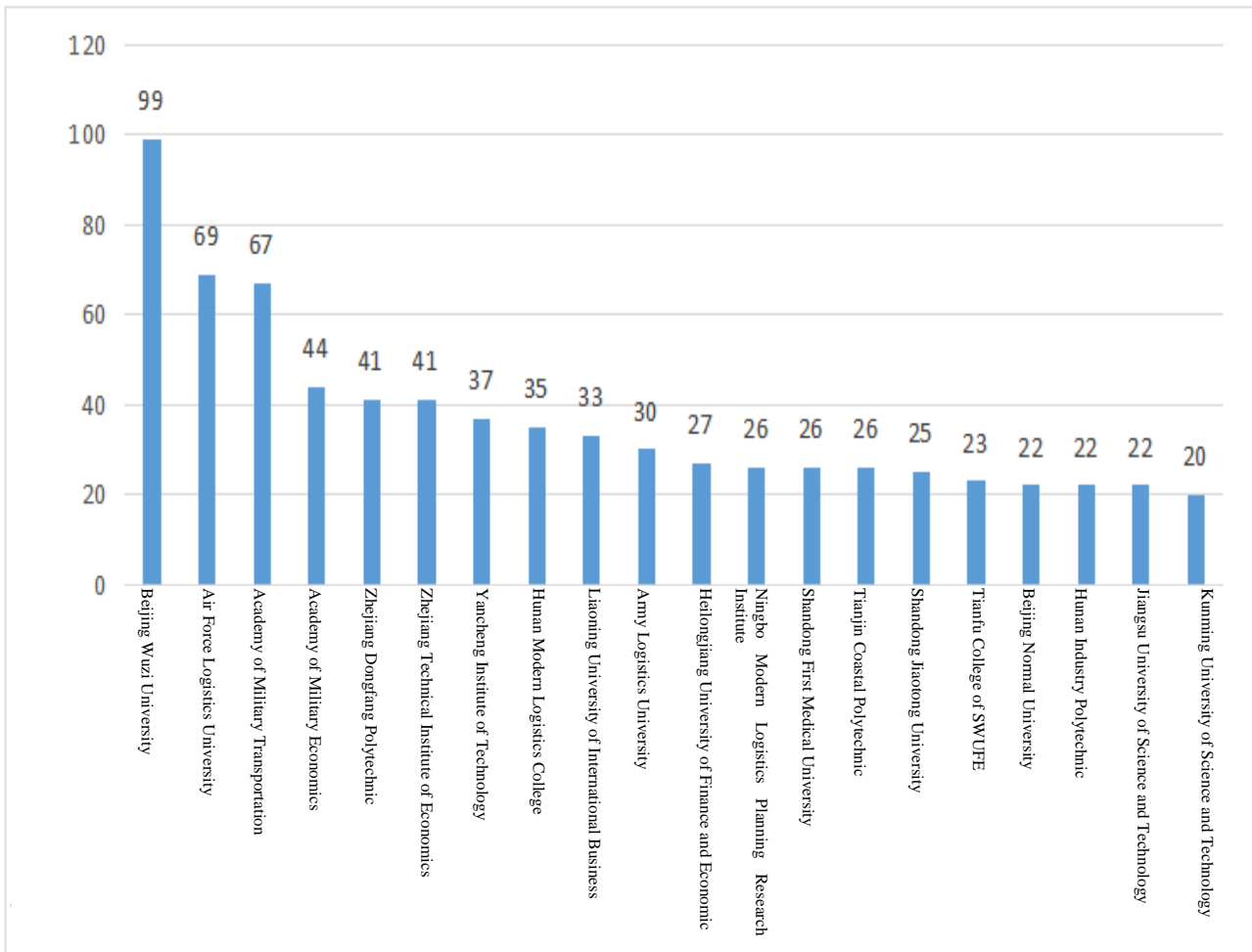


Fig. 2. Distribution of the top 20 universities in the number of topics of the China Society of Logistics.

In terms of the geographical distribution of the above-mentioned colleges and universities, Beijing, Chongqing, Tianjin, Zhejiang, Hunan, Liaoning and other places have the largest number of projects.

D. Statistical results of high-frequency topics

According to Zipf's second law, 2401 word segmentation of the subject of the society is obtained; there are 99 high-frequency keywords with frequency ≥ 14 times, and the cumulative percentage is 4.1%; there are 101 words with frequency ≥ 30 times, and the cumulative percentage reaches 4.2%, and there are 20 words with frequency greater than 100, and the cumulative percentage reaches 0.83%.

After cleaning results and removing residual tail, according to the Pareto Principle, there are 91 high-frequency words, and the cumulative word frequency accounts for 80.11%, as shown in "Table II".

TABLE II. STATISTICAL RANKING OF HIGH-FREQUENCY TOPIC WORDS (CUMULATIVE TOP 80%)

Serial number	Keywords	Word frequency	Grand total Percentage	Serial number	Keywords	Word frequency	Grand total Percentage	Serial number	Keywords	Word frequency	Grand total Percentage
1	Logistics	1483	13.18%	31	E-commerce	86	50.92%	61	Higher vocational education	54	68.69%
2	Development	364	16.41%	32	Strategy	85	51.68%	62	New	53	69.16%
3	Model	330	19.35%	33	Region	83	52.41%	63	Environment	52	69.63%
4	Supply chain	305	22.06%	34	City	82	53.14%	64	International	51	70.08%
5	Enterprise	289	24.62%	35	Evaluation	80	53.85%	65	Finance	51	70.53%
6	Optimization	186	26.28%	36	Internet	75	54.52%	66	Fresh	49	70.97%
7	Construction	171	27.80%	37	Emergency	74	55.18%	67	Practice	48	71.39%
8	System	157	29.19%	38	Internet	73	55.83%	68	Information	48	71.82%
9	Innovation	144	30.47%	39	Military	73	56.47%	69	Control	47	72.24%
10	Management	135	31.67%	40	Logistics	72	57.11%	70	Ability	47	72.66%
11	Agricultural products	135	32.87%	41	Logistics park	72	57.75%	71	Planning	45	73.06%
12	Background	132	34.04%	42	Influence	72	58.39%	72	Method	45	73.46%
13	Application	131	35.21%	43	Mechanism	68	59.00%	73	Data	45	73.86%
14	Construction	130	36.36%	44	Networking	68	59.60%	74	Operation	44	74.25%
15	Our country	117	37.40%	45	Modern	67	60.20%	75	Transportation	43	74.63%
16	Delivery	105	38.34%	46	Path	67	60.79%	76	Demonstration	42	75.00%
17	System	104	39.26%	47	Talent cultivation	66	61.38%	77	Big	42	75.38%
18	Countermeasure	104	40.18%	48	Express delivery	65	61.96%	78	Guarantee	41	75.74%
19	Problem	102	41.09%	49	Green	65	62.53%	79	Cross-border	41	76.10%
20	Economy	102	42.00%	50	Object	62	63.09%	80	China	40	76.46%
21	Industry	98	42.87%	51	Logistics and delivery	62	63.64%	81	Third party	40	76.82%
22	Perspective	98	43.74%	52	Military and civilian	61	64.18%	82	Integration	39	77.16%
23	Service	96	44.59%	53	E-commerce	60	64.71%	83	Intelligence	39	77.51%
24	Analysis	93	45.42%	54	Wisdom	58	65.23%	84	Car	39	77.85%
25	Fusion	93	46.25%	55	Model	57	65.73%	85	Warehouse	37	78.18%
26	Technology	89	47.04%	56	The Belt	57	66.24%	86	Supplies	37	78.51%
27	Cold chain	89	47.83%	57	Risk	56	66.74%	87	Competitiveness	36	78.83%
28	Collaboration	88	48.61%	58	Port	56	67.24%	88	Design	36	79.15%
29	Strategy	88	49.39%	59	Profession	55	67.72%	89	Center	36	79.47%
30	Platform	86	50.16%	60	Road	55	68.21%	90	Facing	36	79.79%
								91	Manufacturing	36	80.11%

E. Time series and cluster analysis

1) Changes in research hotspots (2011-2019):
 Analysis method: In information retrieval, TF-IDF (term frequency-inverse document frequency) is a statistical method used to evaluate the importance of a

word in a document collection or corpus. It is often used as a weighting factor for information retrieval, text mining, and user models. The value of TF-IDF will increase as the number of times the word appears in the document increases, and will decrease as the number of

times the word appears in the corpus increases. TF-IDF is one of the most popular word frequency weighting schemes today [4], and formula (1) is a commonly used idf calculation formula.

$$idf_i = \log \frac{|D|}{|\{j : t_i \in d_j\}|} \quad (\text{Formula 1})$$

It writes word segmentation programs, exports the results and ranks the high-frequency words of themes from 2011 to 2019. It then selects the top 20 words of TF-IDF (term frequency—inverse document frequency) annually, and removes stop words with little meaning, such as analysis, countermeasures, perspective,

application, model, system, optimization, etc. It takes 15 high-frequency words for analysis and finds that three high-frequency words of logistics, supply chain, and delivery appears every year, and the cold chain appears for 8 consecutive years. Starting from 2013, the research hotspot has been agricultural products, which continues for 7 years; E-commerce has appeared in hotspot since 2015 and continues for 5 years; emergency logistics management is a hotspot in 2011-2013; the research on fresh logistics has been popular since 2015; research on the Belt and Road Initiative is hotspot in 2016 and 2017, and then begins to decline, as shown in "Table III".

TABLE III. TF-IDF HIGH-FREQUENCY WORDS BY YEAR

<i>High-frequency words</i>	Logistics	Supply chain	Delivery	Cold chain	Enterprise	Agricultural products	E-commerce	Innovation
<i>Number of rankings</i>	9	9	9	8	7	7	5	5
<i>Year</i>	2011-2019	2011-2019	2011-2019	2011-2018	2011-2017	2013-2019	2015-2019	2015-2019
<i>High-frequency words</i>	Collaboration	Talent cultivation	Industry	Emergency	Low carbon	Fresh	The Belt and Road Initiative	
<i>Number of rankings</i>	3	3	3	3	2	2	2	
<i>Year</i>	Random	Random	2011-2013	2011-2013	2011-2012	2015-2017	2016-2017	

2) *Topic hotspot cluster*: Through the cluster analysis of subject high-frequency keywords, it is found that the topic subject hotspots can be clustered into 4 subject categories, showing the following distribution:

- industry category: (Cluster 0), including keywords: logistics, e-commerce, fresh, finance, agricultural products, transportation, distribution, cross-border, car, military.
- research entry point: (Cluster 1) including keywords: strategy, perspective, application, model, system, optimization, our country.
- logistics technology: (Cluster 2) including keywords: network, information, Internet, wisdom, data.
- management functions: (Cluster 2) including keywords: supply chain, logistics, strategy, control, planning, operation, transportation, mechanism, and delivery.

IV. TOPIC FEATURE ANALYSIS OF THE SOCIETY OF LOGISTICS

A. Obvious characteristics of interdisciplinarity

Logistics science is an applied science, which is comprehensive science, engineering, management, economics and other multi-disciplinary knowledge, so the selection of scientific research methods also presents diverse characteristics. Among more than

2,000 topics, there are 245 topics mentioned in methodology, and among them, 12 technical methods such as analytical network process (ANP), data envelopment analysis (DEA), customer relationship management (CRM), hazard analysis and critical control point (HACCP), SWOT analysis method, game analysis, ant colony algorithm, genetic algorithm, Petri net, Flexsim virtual simulation, SNA network architecture, and SVM support vector machine are more common. To classify the above methods, the main sources of management disciplines are ANP, DEA, CRM, SPD, SWOT; the main source of economics is game theory; the main sources of engineering are algorithms and simulations in the computer field, including ant colony algorithm, genetic algorithm, Petri, Flexsim, SNA, SVM; and there is also DEA derived from the intersection of operations research, management science and mathematical economics.

It can be seen that the interdisciplinary characteristics of logistics research are obvious, so there are no hot words about research methods. This phenomenon matches the nature of logistics management and engineering disciplines.

B. Focusing on national strategy and industry development

In 2013, in order to give full play to the comparative advantages of resources, accelerate the cultivation of regional characteristic industries, and realize agricultural efficiency and farmers' income, the

Ministry of Agriculture revised the "Special Agricultural Products Regional Distribution Plan (2006-2015)" issued in 2007, and formed the "Special Agricultural Products Regional Distribution Plan (2013-2020)"; the State Council issued "Several Opinions on Accelerating the Development of Modern Agriculture and Further Enhancing the Vitality of Rural Development", proposed to coordinate the construction of agricultural product distribution centers, sales areas, and wholesale markets for agricultural products, and build an integrated circulation chain of agricultural product production and marketing [5]. In addition, local governments have also continuously issued policies and plans in the field of agricultural product circulation. Among them, Shanghai has successively issued the "Shanghai Edible Agricultural Products Wholesale and Retail Market Development Plan (2013-2020)" and "Guidelines on Accelerating the Development of Cold Chain Logistics of Municipal Agricultural Products", etc.

In that year, "agricultural products" became a hot word in research, and it continued until 2019.

In September and October 2013, Chinese President Xi Jinping put forward the cooperation proposal of building the "New Silk Road Economic Belt" and the "21st Century Maritime Silk Road" respectively. On March 28, 2015, the National Development and Reform Commission, the Ministry of Foreign Affairs, and the Ministry of Commerce jointly issued the "Vision and Actions for Promoting the Joint Construction of the Silk Road Economic Belt and the 21st Century Maritime Silk Road". From 2016 to 2017, "the Belt and Road Initiative" become a hot topic.

2015 was the year when BAT giants joined forces and emerging e-commerce companies "savagely grew", and also the first year of the development of B2B e-commerce 2.0 [6]. The State Council issued the "Opinions on Vigorously Developing E-commerce and Accelerating the Cultivation of New Economic Impetus"; the General Office of the Ministry of Commerce issued the "Notice on Issuing the "Internet + Circulation" Action Plan", and the purpose of "Notice" is to accelerate the in-depth integration of the Internet and the circulation industry, promote the transformation and upgrading of the circulation industry, innovate ways to serve the people's livelihood, and release consumption potential. Therefore, in 2015, e-commerce logistics entered the top ten research hot words, and it continues to this day.

C. Paying attention to public events and reflecting the function of logistical care

From 2011 to 2013, emergency logistics was a research hotspot, but after 2013, emergency logistics gradually faded out of the line of research hot words. This is because the 2008 Wenchuan earthquake urged

the public to pay attention to emergency logistics. And after a major public event, emergency logistics was relatively indifferently treated.

In terms of quantity, from 2011 to 2019, the overall ranking of emergency logistics keywords is 37, and the word frequency is 74. A total of 51 projects including emergency logistics topics have been approved by military academies, accounting for 68.9%. The term "logistics" is derived from military logistics, and emergency material support is an important task of military logistics. The research on emergency logistics management in military academies has always been very professional and in-depth, and its advantages and internal drive in the field of emergency research are irreplaceable.

D. The subject difference is not obvious enough

From the perspective of subject hot words, cold chain logistics, delivery, agricultural product logistics, and e-commerce logistics have received more attention. The themes meet the needs of society and are basically coordinated with social development. However, the data from 2011 to 2019 is highly repetitive, and the hot words of the subject do not change much each year. Logistics, supply chain, delivery, cold chain, and agricultural products have always been hot words. Among the 9-year subject hot words, there are 20 high-frequency words each year. After removing the stop words, they are concentrated in 18, as shown in "Table IV". It can be seen from this phenomenon that the concentration of the subject words of the China Society of Logistics is very high.

TABLE IV. TF-IDF HIGH-FREQUENCY WORDS BY YEAR

Year	Hot word								
	1	2	3	4	5	6	7	8	9
2011	Logistics	Supply chain	Delivery	Cold chain	Enterprise				
2012	Logistics	Supply chain	Delivery	Cold chain	Enterprise				
2013	Logistics	Supply chain	Delivery	Cold chain	Enterprise	Agricultural products			Collaboration
2014	Logistics	Supply chain	Delivery	Cold chain	Enterprise	Agricultural products			
2015	Logistics	Supply chain	Delivery	Cold chain	Enterprise	Agricultural products	E-commerce	Innovation	Collaboration
2016	Logistics	Supply chain	Delivery	Cold chain	Enterprise	Agricultural products	E-commerce	Innovation	
2017	Logistics	Supply chain	Delivery	Cold chain	Enterprise	Agricultural products	E-commerce	Innovation	Collaboration
2018	Logistics	Supply chain	Delivery	Cold chain		Agricultural products	E-commerce	Innovation	
2019	Logistics	Supply chain	Delivery			Agricultural products	E-commerce	Innovation	Collaboration
Year	Hot word								
	10	11	12	13	14	15	16	17	18
2011		Industry	Emergency		Low carbon	Logistics			
2012		Industry	Emergency		Low carbon	Logistics			
2013		Industry	Emergency						
2014	Logistics park					Logistics		Talent cultivation	
2015	Logistics park			Internet					Fresh
2016	Logistics park			Internet			The Belt and Road Initiative	Talent cultivation	
2017				Internet		Logistics	The Belt and Road Initiative		Fresh
2018									
2019						Logistics			

V. CONCLUSION

A. Enlightenment from the study of the topic selection of the society

1) *The differences in research topics need to be improved:* The selection of scientific research topics has both uniformity and difference. The unity not only reflects the trend of scholars' topic selection, but also is a powerful interpretation of the development direction of logistics discipline. And the difference is the embodiment of the differentiation and specialization of research institutions [2].

On the whole, the differences in the topic of the subject are not obvious enough. In addition to the better depth of military academies and professional research institutions, there is a higher convergence of research topics in higher education institutions. In the future selection of research topics, researchers can develop ideas and improve the innovation, difference and industry characteristics of the topic. The depth of research still needs to be further explored, and the

diversification of sources of research topics needs to be further expanded. Regional characteristics and urban logistics still need further exploration.

2) *Continuing to play to the advantages of interdisciplinary:* Disciplinary intersections are often new growth points and new frontiers of science, and they are most likely to produce major scientific breakthroughs and revolutionize science. At the same time, interdisciplinary science is a comprehensive and systematic product, which is conducive to solving major and complex scientific problems, social problems, and global problems faced by mankind. To accelerate the development of scientific research, it's necessary to vigorously promote interdisciplinary and pay attention to the development of interdisciplinary science. Therefore, it is of great significance to propose and solve interdisciplinary science problems [7].

The logistics industry, academia, and research circles can actively use their own resources, platforms, and the differentiated advantages of the preliminary

research foundation to maintain their unique advantages in the cross-disciplinary field of logistics research. It can integrate research ideas, research frameworks, and research methods in management, economics, operations research, and engineering to improve the quality of research topics and innovation.

3) *The application of new technology needs attention:* Among the 2,342 topics, there are only 225 themes containing terms such as "application", "technology", "platform", "wisdom", "smart", and "blockchain", accounting for less than 10%. Among them, it also includes research on concepts and theoretical frameworks. However, there are not many real research projects on the application of new technologies in the field of logistics. This is not matched with the smart and intelligent development direction of modern logistics. The reason is that although the overall level of logistics in China is on the rise, the overall level of development is not high, and there is still a certain gap between it and developed countries. However, compared with practical work, scientific research is forward-looking and leading, which requires scientific research workers in the logistics industry and academia to improve the level of science and technology, broaden their horizons, aim at the new momentum of economic and social development, and promote the new development of logistics technology.

B. Limitations of this paper

It is feasible to use text cluster method to study topic subject hotspots analysis.

Since the China Society of Logistics has not published the summary information text of each item, this paper mainly analyzes the short text of the title. Because natural language processing technology still has problems such as insufficient precision and ambiguity for short texts, the conclusions of the analysis are relatively rough, which are left for further improvement.

References

- [1] Introduction to research projects of China Society of Logistics [EB/OL]. [2011-10-04] (quoted date) <http://csl.chinawuliu.com.cn/items/1000047.html>. (in Chinese)
- [2] Zhang Li, Wang Honghong. Based on cluster analysis to explore the hot topics of graduate students' thesis of nursing science [J]. Nursing Journal of People's Liberation Army, 2019, (1): 6-10,59. (in Chinese)
- [3] Zheng Yanning, Xu Xiaoyang, Liu Zhihui. Research on the research front identification method based on keyword co-occurrence [J]. Library and Information Service, 2016, (4): 85-92. (in Chinese)
- [4] Wang Rifen. Comprehensive research on bibliometrics and content analysis [D]. Nanjing: Nanjing University of Science and Technology, 2007. (in Chinese)
- [5] Yu Xin. The impact of macro-industrial policies on the development of modern commerce and circulation industry. Business Economics Research [J]. 2020,(6):21-23. (in Chinese)
- [6] Sun Liang, Zhang Yan. Evolution and innovation of B2B e-commerce model: Empirical analysis based on major B2B platforms [J]. E-commerce. 2015, (1): 21-24. (in Chinese)
- [7] Lu Yongxiang. The significance of interdisciplinary and interdisciplinary science. Bulletin of the Chinese Academy of Sciences [J]. 2005,(1):58-60. (in Chinese)