

Analysis of Hot Spots and Trends of Chinese Medical Maintenance Combination based on Sent-LDA Model

Ting Gu

School of Economics and Management, Jiangsu University of Science and Technology, Zhenjiang
212003, China

Abstract. To analyze the hot spots and trends of periodical texts on the topic of medical and maintenance combination can provide the basis to grasp the hot spots of research in this field and promote its further development. With the method of machine learning, 1,459 fund literatures with the theme of "combination of medical and nursing care" in CNKI database from 2013 to 2021 were taken as analysis samples. python was used to crawl the text content and preprocess it. Subject words were selected based on Sent-LDA model, and data were substituted to form a visualization map. Analyze the research trend of Chinese medical combination. Study found that at present our country medical have combined with the research hotspots focus on the allocation of resources, wisdom, provide for the aged, personnel management and rural development, then can pay attention to the field of policy implementation, cross-sectoral coordination mechanism, the old man nursing assessment system and grading system of diagnosis and treatment research.

Keywords: Combination of Medical and Nursing Care; Sent-LDA Model; Text Mining; Research Trend.

1. Introduction

With the increasing aging in our country, the chronic disease of the elderly has gradually become the focus of pension industry research. Since 2013, a series of policies have been released to advocate "integrated care", "diversified health care for the elderly", "medical care and elderly care services" and other key tasks and safeguard measures. It can be seen that "the combination of medical care and elderly care" has become an inevitable trend of the development of China's elderly care industry. Scientific and effective integration of medical and elderly care resources to a certain extent to solve the elderly daily examination, health care, rehabilitation and other diversified needs. Therefore, the integration of medical resources and pension service resources makes medical services and pension services deeply integrated development, and provides the elderly with medical treatment, care, life and other aspects in one service mode, increasingly becoming a research hotspot in the academic circle.

In order to further understand the research status of the combination of medical care and maintenance, this paper analyzes the journal text with the theme of the combination of medical care and maintenance, aiming to discuss the research hotspot and development trend of the combination of medical care and maintenance in China, analyzes the shortcomings of the existing research and points out the future research direction, in order to provide theoretical basis for the subsequent academic research and promote the development of the combination of medical care and maintenance in China. Among them, the age distribution of literature can directly reflect the change relationship between time and quantity of literature in a certain research field, thus revealing the research status and development speed of this field, and taking this as the basis for predicting its development trend. As can be seen from Figure 1, literatures on the combination of medical and nursing care in China show an increasing trend year by year, especially after 2015, with a great breakthrough in the number of literatures, among which the largest number is in 2019. Due to the more rigorous academic policy in 2019, the research literature has slightly decreased since that year, but the number of literature still remains above 800. It can be seen that there is a large development space in the field of combining medical and nursing care in China. The existing research literature is more and more refined and specific while paying attention to high quality.

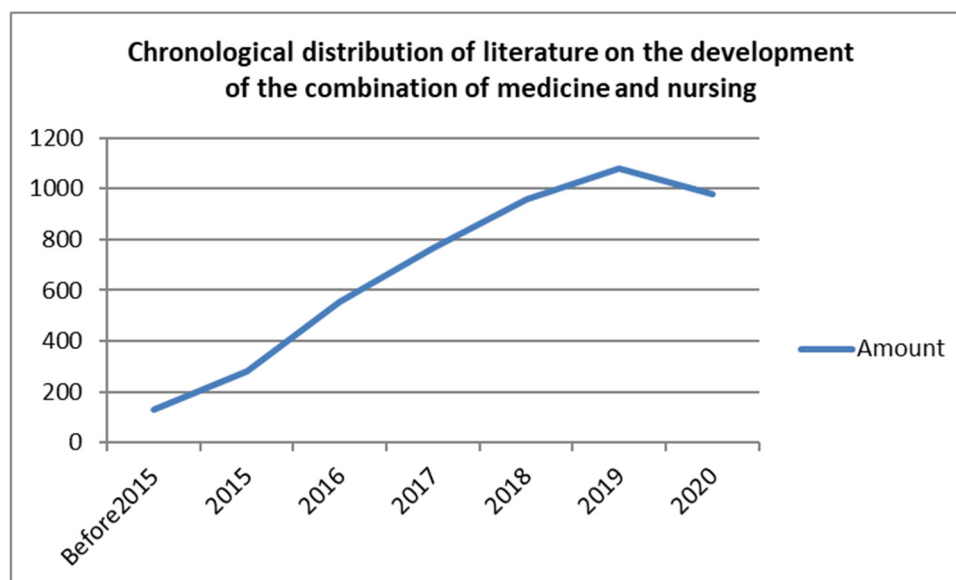


Fig 1. Chronological distribution of literatures on the combination of medical and nursing care

Most of the existing studies on text trends adopt the bibliometric method [1-5]. Citespace software is used to sort out the authors, research institutions and keywords of articles, so as to obtain the visualization analysis of knowledge graph, while few analyze the trends from the perspective of text mining. The use of text mining can directly extract useful and potential information from the text content, in the study of article trend analysis can more accurately reflect the core topic of the article and the relationship between the topic and keywords. Therefore, from the perspective of text mining [6-9], this paper adopts the method of machine learning to analyze and preprocess the contents of 1,482 literatures on the theme of "combination of medical and nursing care" downloaded from the database of China National Knowledge Network (CNKI), and introduces the Sentence-LDA model to the processed text contents. That is, a sentence layer is added on the basis of the original LDA model, which mainly considers the words with the highest proportion of weight in a sentence, so as to be more refined when dealing with a large amount of text data. The model was used to identify and cluster key keywords, and the distribution of key words under each theme was obtained, so as to analyze the hot spots and trends of Chinese medical maintenance combination research.

2. Research Methods

2.1 Data Collection

This paper uses the database of China National Knowledge Network (CNKI) as the data source, and the research object is articles containing the terms "combination of medical and nursing care" or "integration of medical and nursing care" in their titles, keywords and abstracts. Among the published articles by 2021, the documents of the foundation were selected from the advanced search bar and downloaded into PDF format. 23 documents that were inappropriate or undownloadable were deleted, and 1459 articles were retained as data samples.

2.2 Data Preprocessing

The text content of PDF is parsed and preprocessed using python. It mainly includes: 1) Cleaning the data contained in the original text data, such as symbols and English characters, with low value content and chaotic data structure, which can easily lead to the deviation of text mining results; 2) Use the Jieba package in Python to divide a sentence into multiple words to make the content clearer; 3) In text processing, eliminate words such as interjections or modal particles with high frequency but little practical meaning. FIG. 2 shows a comparative analysis diagram before and after data preprocessing.

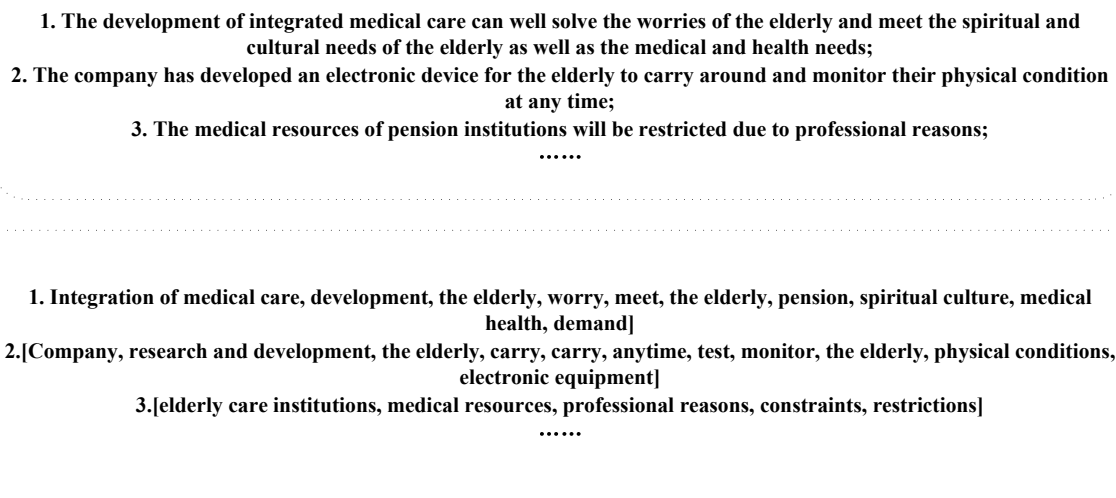


Fig 2. Comparison before and after data preprocessing

2.3 Identification of Key Keywords based on the Sent-LDA Model

In order to dig out the hot spots in the field of combining medical and nursing care from a large number of texts, this paper introduces the topic model. In the field of natural language processing, the topic model is a statistical model to find abstract topics in a series of documents in an unsupervised way, among which the Latent Dirichlet Allocation (LDA) model proposed by Blei [10] in 2003 is the most commonly used one. LDA model uses a three-layer Bayesian probability model to correspond to the three-layer structure of "document - topic - word", and believes that each word is generated by "selecting a certain topic with a certain probability, and then selecting a certain word from this topic with a certain probability", that is, each word in the document is extracted from a potential topic. Based on a large amount of text information, LDA model can realize automatic clustering of these text data, so as to dig out potential topics and calculate the proportion of each topic [11].

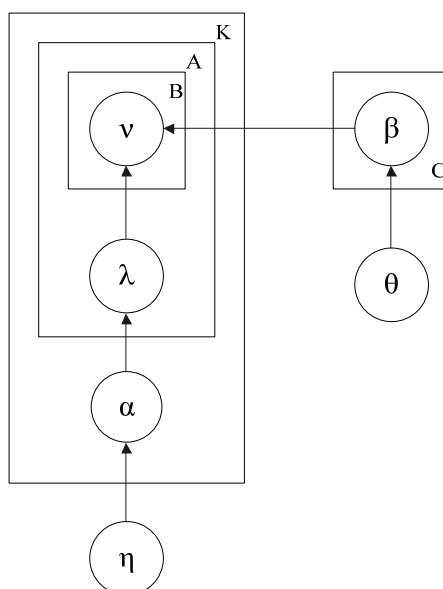


Fig 3. Structure representation diagram of the Sent-LDA model

In theory, the LDA model can be directly used to extract high-frequency subject words in the field of medical and nursing combination from a large number of text contents, but a title often only expresses a topic, and the LDA model believes that multiple topics can be extracted from a sentence. Therefore, in order to better conform to the characteristics of research hotspots in the text content,

this paper introduces the Sent-LDA model proposed by Bao et al. [12] aiming at the characteristics of short texts. Based on the LDA model, the Sent-LDA model adds the "one topic only sentence" restriction, which assumes that all words in a sentence are drawn from an underlying topic. Specifically, based on the three-tier structure of "document -- topic -- word" of the LDA model, the Sent-LDA model adds a "sentence" layer between "document" and "topic", forming a four-tier structure of "document -- sentence -- topic -- word". In this way, it mainly considers the words with the highest proportion of weight in a sentence, so as to be more refined when dealing with a large amount of text data. The structure of the Sent-LDA model is shown in Figure 2, where the meanings of each symbol are summarized in Table 1.

Table 1. Conformance and implication in the Sent-LDA model

Symbol	Meaning	Symbol	Meaning
K	\Number of documents	λ	Theme
A	The number of sentences in the literature	α	The polynomial distribution of Text - topic
B	The number of words in the literature	β	Polynomial distribution of "topic - word"
C	The number of topics in the literature	η	Prior distribution of α
v	word	θ	The prior distribution of β

The generation process of the Sent-LDA model is as follows:

- 1) For each topic $\lambda \in \{1, \dots, C\}$, the distribution of words in the topic is $\beta_\lambda \sim \text{Dirichlet}(\theta)$
- 2) For each document k,
 - I. Select a topic distribution $\alpha_k \sim \text{Dirichlet}(\eta)$;
 - II. For each sentence s in document k,
 - i. Obtain the topic $\lambda_{k,s} \sim \text{Multinomial}(\alpha_k)$ of sentence s from the polynomial distribution of text-topic;
 - ii. For each word n in the sentence s, the word $v_{k,s,n} \sim \text{Multinomial}(\beta_{\lambda_{k,s}})$ is obtained from the polynomial distribution of topic-words.

Before applying the Sent-LDA model to identify topics contained in large amounts of textual data, you need to set the number of topics. In this paper, the perplexity method is used to assist in choosing the range of suitable topics. Confusion is an indicator to evaluate the quality of models in the field of natural language processing. It is assumed that K_t , K, v_k and A_k represent the test set, the number of documents, the number of words in document k and the number of words in document k, respectively. The confusion of each document test set is defined as:

$$\text{Perplexity}(K_t) = \exp\left(-\sum_{k=1}^K \log p(v_k) / \sum_{k=1}^K A_k\right) \quad (1)$$

Formula (1) can be used to calculate the confusion degree corresponding to the number of topics in the model, and then draw the change curve of the confusion degree with the increase of the number of topics. In this paper, while calculating the degree of confusion, python software was used to get the effect diagram of topic division, and the best clustering results were obtained by synthesizing the two perspectives.

3. The Distribution Results of Chinese Medico-maintenance Combination Research Theme

3.1 Topic Number Selection

The Sent-LDA model was used to identify the key topic words of 1482 downloaded PDF texts. Firstly, the confusion degree was calculated to determine the range of the optimal topic number, and then the appropriate topic number was selected. Figure 4 shows the variation curve of the model's confusion degree when the number of topics is different. It can be seen that the confusion degree tends to be stable when the number of topics is 3 and 4, and begins to increase when the number exceeds 4. By checking the clustering results with the number of topics being 3, 4, 5 and 6, it is found that although the larger the number of topics is, the higher the degree of confusion is. At the same time, according to the clustering effect diagram, when the number of topics is 3, the farther away each topic circle is, the lower their similarity is, and the classification effect of topics is relatively good. Therefore, the number of topics is set to 3 in this study.

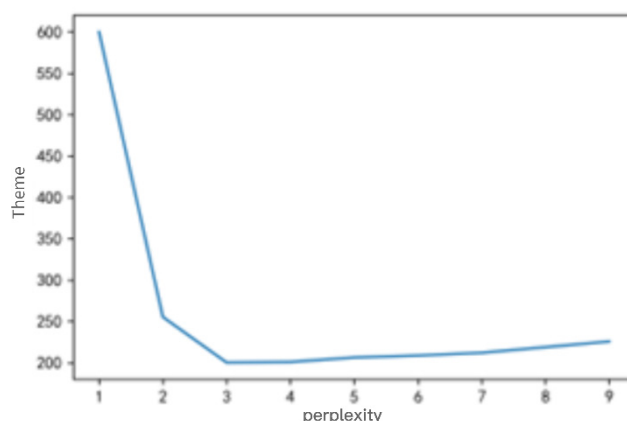


Fig 4. Confusion degree changes with different topic numbers

When using the Sent-LDA model to identify topics, we screened the words (stop words) that appeared frequently in each topic but were easy to cause trouble to the recognition of topics. Finally, 12 words were identified as stop words in the research field of the combination of medicine and care, including "combination of medicine and care", "pension", "old people", "old people", etc. After removing the stoppages, the Sent-LDA model is applied again to obtain three topics and a list of keywords for each topic from the PDF downloaded from CNKI. The results showed that, compared with the direct application of the Sent-LDA model to the recognition of topics in combination with medical care, the recognition of topics was no longer affected by ill-defined words after the screening and elimination of words, and the classification effect was improved.

3.2 Topic Result Analysis

To clarify the content of each topic, you need to name them. The high-frequency words clustered into each topic are displayed in the form of word cloud map. The higher the lexical weight ratio, the larger the word cloud map. See Figure 5 for details

It can be clearly seen from the word cloud map of the clustering results that "mode", "institution" and "development" in the existing research on the combination of medical and nursing care are the words with the highest weight ratio under the three themes. After comprehensive consideration of all the words under each theme, the three themes are finally determined as follows: The research on the mode of the combination of medical and nursing care, the research on the institutions of the combination of medical and nursing care and the other research on the combination of medical and nursing care with "development" as the core.



Fig 5. Word cloud map of clustering results with topic number of 3

(1) Study on the mode of combination of medical and nursing care

The research on the model of combining medical and nursing care is helpful to promote the further development of this field. The existing models of the combination of medical and nursing care mainly include three types: "care in medical care", "care in medical care" and "care in medical care". The difference lies in the expansion of old-age care function led by medical institutions or the expansion of medical care function led by old-age care institutions. As can be seen from the related clustering word cloud map, words such as "demand", "nursing", "health" and "resources" have a high weight, which means that for different elderly people in different modes, nursing needs and health needs should be comprehensively considered to complete the resource allocation under different supply and demand modes. For example, the elderly who are in good health occasionally need middle-end medical and health resources, so it is more necessary to strengthen the demand for front-end medical and health resources and improve the supply of health and prevention resources. However, for the elderly with potential disease risks and disabled or semi-disabled elderly, they are more dependent on the middle and back-end medical and health resources. For these elderly people, the needs of the back-end of medical and health resources should be analyzed while fully considering the connection with professional conservation resources.

Correspondingly, words such as "Internet", "wisdom", "radiation" and "platform" show that the wisdom of the combination of medical and nursing care mode has become a hot topic. By combining the wisdom with the medical industry and the elderly care industry, the platform information sharing can be realized, so as to improve the quality of medical treatment and service in the process of model operation. The model of the combination of intelligent medical care and old-age care is mainly composed of three levels: the first level is the service object, namely the elderly and their families. With the help of intelligent terminals connected to the intelligent pension platform, such as mobile phones and wearable devices, service subjects can upload the vital signs, geographical positioning and service demand information of the elderly in real time. The second level is technical support, namely the intelligent home-based elderly care platform based on the Internet of Things, cloud computing and other technologies, which plays the role of a hub bridge connecting service subjects and service providers. The third level is the service provider, namely community service centers, medical and health centers, pension institutions, catering companies, financial institutions, universities for the elderly, travel agencies, etc. In-depth discussion and research on intelligent mode will help to provide more professional and convenient medical and pension services for the elderly, and provide strong guarantee for every decision with the help of platforms and Internet big data technology, thus improving the quality of life of the elderly to a certain extent.

(2) Research on institutions combining medical and nursing care

According to the word cloud map with the theme of "institutions", the research of "medical institutions" and "community health institutions" is relatively high in the present research, which is also the hot spot of the combination of medical and nursing in our country. Let the medical institutions pay attention to the maintenance of talent investment, let the community health institutions strengthen the construction of medical resources, give full play to the professional advantages of medical services and elderly care services, make full use of social organizations, enterprises, volunteers and

other subjects, jointly support the quality of life of the elderly. It is found in the research of different levels of medical and nursing care institutions that high-end medical and nursing care institutions can create medical and nursing care services with the help of real estate industry or tourism, and these institutions can adjust themselves according to the law of market value and price changes, basically achieving long-term supply and demand balance, while some institutions with low positioning can obtain government subsidies through PPP operation. To improve medical equipment, environmental facilities, etc.

As can be seen from the distribution map of institutional topic words, a lot of research has been gradually carried out on the development of rural medical and nursing institutions. How to make the combination of rural medical and nursing institutions better and bigger is also a research hotspot at present. Due to the low level of economic development in rural areas compared with that in urban areas, the consumption capacity and acceptance degree of medical and nursing institutions for the elderly are not enough, coupled with limited government resources, leading to the low coverage rate of medical and nursing institutions in rural areas. Related research has found that most of the rural elderly, influenced by traditional concepts, are not willing to choose institutional care, they prefer to live in the familiar environment, which has become a major problem in our country combining medical care with institutional care. In addition, the management and nursing staff of rural medical and nursing institutions are not of high professional quality, most of their goals are to gain more benefits, and insufficient attention to the psychological needs and mental health of the elderly are also problems that need to be solved.

(3) The combination of medical and nursing studies with "development" as the core

The clustering results of the third topic are rather disorganized. Based on the distribution of high-frequency words of this topic, the topic is summarized here as the combination of medical care and other studies with "development" as the core. "Policy", "current situation", "dilemma" and "service" are the most frequently used discourse studies under this theme. In recent years, the state has introduced a series of policies to vigorously support the healthy and orderly development of the combined medical and nursing care market for the elderly. However, due to multiple management, the demarcation between powers and responsibilities of relevant departments is not clear, there are obstacles in information communication, and there are differences in resource allocation, which ultimately lead to the implementation of the policy. To this end, it is necessary to further clarify the responsibilities of all relevant departments, and establish a multi-department coordination mechanism led by the health department, so as to change the situation of "multi-management", which is also a major trend in the research field of the combination of medical and nursing care.

Other words that came first under the topic were "indicator", "expert", "assessment" and so on. This is because at present our country has not established a complete unified nursing standards for the aged, even the common medical institutions are also by disease diagnosis nursing classification, but such forms are often greatly different from the actual needs of the elderly, so that the life needs of the elderly can not be effectively met, resulting in a waste of nursing human resources. This is also a major research trend at present. It is suggested that the government should take the lead in establishing a scientific and unified evaluation system for elderly care in institutions combining medical and nursing care, train professional evaluation personnel, and encourage social capital to establish professional evaluation institutions to lay a solid foundation for graded care for the elderly.

4. Conclusion and Prospects

In terms of the number of studies, the research on the combination of medical and nursing care in China officially started in 2015, and on the whole shows an increasing trend year by year, and the research literature pays more and more attention to refinement and specificity while paying attention to high quality. In terms of research hotspots, the Send-LDA model was used to extract and cluster the key words of PDF text, and finally three main research hotspots were found in the fields of the combination of medical and nursing care, namely, the model of the combination of medical and

nursing care, the institution of the combination of medical and nursing care, and other studies of the combination of medical and nursing care with "development" as the core. In the keyword clustering results of each topic word cloud map analysis found that, Current research focuses on resource allocation under different supply and demand models of the combination of medical care and old-age care, intelligent model of the combination of medical care and old-age care considered from three perspectives of service subject, technical support and service provider, research on the combination of medical care and old-age care institutions dominated by "medical institutions" and "community health institutions", management of professional personnel in institutions and development of rural combination of medical care and old-age care institutions ; From the perspective of research trend, the following can be considered: 1) how to take "development" as the core as soon as possible to improve the relevant research on the implementation of the combination of medical and nursing policies, so that the government policy is not superficial, can fundamentally solve the problems of the elderly; 2) Clarify the responsibilities of relevant departments and improve the cross-departmental collaboration mechanism; 3) To establish a scientific and unified evaluation system of old-age care in medical and nursing institutions; 4) According to the different conditions of the elderly to implement the hierarchical diagnosis and treatment system to maximize the utilization of resources.

As this paper only takes 1,459 fund literatures in CNKI database as analysis samples, foreign academic platforms are not involved. Given that the Sent-LDA model is probability-based, it is necessary to expand the number and scope of relevant papers, which can be collected from multiple domestic and international platforms, and add more content to ensure the accuracy and appropriateness of the topic. The next step is to carry out text analysis of foreign literatures on the combination of medical and nursing care and explore their research trends.

References

- [1] Si Xiaohong. Statistical analysis and prospect of research hotspots on balanced development of compulsory education [J]. Journal of Education, 2015, 11(06): 49-59.
- [2] Wang Xiaoxiao, Guo Qing. Analysis on the Hot spot and development Trend of medical care Combination based on CiteSpace in the past 10 years [J]. Chinese Journal of General Practice, 201, 24(01): 92-97.
- [3] Zhang Zeyu, Tian Chong, Fang Pengqian. Analysis on the Development Trend of medical care Combination System during the "14th Five-Year Plan" based on healthy aging [J]. Chinese Journal of Public Health Administration, 201, 38(07): 493-495.
- [4] Li Zi, Qian Siwen, Pu Haixia. Analysis on the Hot spot and Development Trend of Chinese Medical Maintenance Combination based on CiteSpace [J]. Journal of Geriatric Research, 202, 2(05): 27-32.
- [5] Yang Jinghui. Research hotspot and evolution trend of Chinese Medical Care combined pension model: A visualization analysis based on CiteSpace software [J]. Scientific Research on Aging, 201, 9(02): 51-64.
- [6] Bi Dayan, Chu Qihuan, Cao Ran. Research on the Influencing Factors of Consumers' Intention to Negative Review based on Text Mining [J]. Information Theory & Practice, 20, 43(10): 137-143. (in Chinese)
- [7] Lu Jingjun, Hu Shunyi, Yu Jianguang. Method and demonstration of online public opinion Monitoring in Universities based on LDA-BiLSTM model [J]. Information Theory and Practice, 20, 43(11): 156-161.
- [8] Yan Duanwu, Yang Xiongfei, Li Tiejun. Product Review Sentiment Analysis based on Product Feature Tree and LSTM Model [J]. Information Theory & Practice, 2019, 42(12): 134-138. (in Chinese)
- [9] Sun Yusheng, Zhu Bo, Zhu Lijun. Research progress of information recommendation based on Knowledge graph [J]. Information Theory & Practice, 2019, 42(12): 163-169+149.
- [10] Li Bin, Wang Yinghui, Zhu Xiaoqian, Li Jianping. Identification and Evolution Analysis of important risk Points in Insurance Industry -- Based on text risk information disclosed in financial Report [J]. Systems Engineering Theory and Practice, 2021, (05): 1-15.
- [11] Travis Dyer, Mark Lang, Lorian Stice-Lawrence. The evolution of 10-K textual disclosure: Evidence from Latent Dirichlet Allocation [J]. Journal of Accounting and Economics, 2017, 64(2-3): 221-245.
- [12] Bao Y, Datta A. Simultaneously Discovering and Quantifying Risk Types from Textual Risk Disclosures [J]. Management Science, 2014, 60(6): 1371-1391.