Textual Clustering Analysis of River Chief System Policy in China

Zhimo Zhao[(✉)] [iD]

Department of Public Administration, University of Electronic Science
and Technology of China, Chengdu, China
`627682339@qq.com`

Abstract. The river chief system is an innovation of China's water resources management system, the core of which is to control and protect the water environment of rivers and lakes. The optimization of the river chief system cannot be achieved without the support and guarantee of policies. In this paper, the author adopts big data text clustering methods, such as the Linear discriminant analysis (LDA) model, to collect the topics of China's river chief system policies and identify the policy texts' potential topic information, intensity, and structural features. It is found that the topic of China's river chief system policy has low topic similarity and little difference in topic intensity. In addition, policies enacted focus more on law enforcement supervision, ecological governance, performance appraisal, information management, and other fields, while policies on public participation and watershed coordination are insufficiently supplied. In the future, policy supply in this respect should be strengthened.

Keywords: river chief system · text clustering · Linear discriminant analysis

1   Introduction

Only about one-third of the water in China's pilot water system is suitable for drinking, and China has suffered from serious water pollution and growing social problems related to such pollution. In 2008, there was a severe algae crisis in China's Taihu Lake. Local governments have boldly implemented the river chief system and achieved remarkable results. Hunan, Jiangxi, Guizhou, and other provinces have followed suit and implemented the river chief system. In 2016, The General Office of the State Council issued the Opinions on the Full Implementation of the River Chief System, marking the official promotion of the river chief system as a policy system nationwide. The core tenet of the river chief system is that the party and government leaders of the local government are simultaneously responsible for the optimal allocation of water resources and pollution prevention and control within their jurisdiction.

In this paper, the author evaluates and analyzes the policy topic of the river chief system in China based on text cluster analysis. This paper contributes in the following aspects. First, based on the analysis of the Chinese policy text, this paper analyzes the text of China's river chief system policy to make up for the shortage of the previous

academic analysis of this policy text. Secondly, based on the big data analysis method, this paper analyzes the text of the river chief system policy to make up for the adverse problems, such as the strong subjectivity of the previous qualitative analysis of policies.

2 Literature Review

2.1 Text Clustering

Text clustering analysis is a process of automatically grouping text sets and a typical unsupervised machine learning process. Text clustering has a certain automatic processing ability because it does not need a training process or prior manual labeling of document categories. It has become an important means to extract and organize text information effectively and has attracted attention from more and more researchers in recent years.

In the beginning, researchers used the vector space model for text analysis. The vector space model simplifies the processing of text content into vector space operation and expresses semantic similarity through spatial similarity. When a document is represented as a vector of document space, the similarity between documents can be measured by calculating the similarity between vectors. However, this model cannot solve the phenomenon of polysemy and synonym in documents and can only be retrieved by a simple lexical matching pattern, affecting the accuracy and recall of information retrieval. Biel et al. (2003) proposed the LDA topic model in response to some defects of the vector space model. This model carries out a prior distribution of topics and special genes by introducing textual topic distribution, significantly reducing data dimensionality. Meanwhile, the parameter size of the model is fixed, which is suitable for large-scale text sets.

Text clustering analysis has been widely used in information retrieval, mathematical modeling, and other fields. For example, it can be used to improve the information retrieval accuracy of the retrieval system and predict the user's tendency to comment online. Text clustering analysis is also reflected in the field of public policy research. For example, Liu et al. (2021) explored China's SME support policies based on PMC thematic model, LDA thematic model, and policy co-word analysis. Li et al. (2021) used this model to conduct a quantitative study on China's hog price regulation policy. Zhang et al. (2022) used the latent Dirichlet allocation model, TF-IDF feature extrac-tion algorithm, and Euclidean distance measurement method to identify and judge whether there are similarities between the insurance, banking, taxation, and financial policy texts of the Hebei Provincial government and Shijiazhuang Municipal government.

2.2 River Chief System

The academic research on the river chief system has focused on implementation framework and policy performance. The implementation framework is based on the analysis of river basin scope, shoreline scope, and management demarcation in Anji County. Zhou et al. (2020) pointed out the principles and methods of river shoreline demarcation in Anji county, which is helpful to the implementation of the river chief system policy in

Anji county. Wang et al. (2019) argued that the operation of the river chief system does not match well with the three dimensions of authority, responsibility, and public participation, which may be successful in the short term but may lead to the failure of water pollution control in the long term. In terms of policy performance, Li et al. (2021) used the method of breakpoint regression to empirically explore the influence mechanism of the river chief system on water pollution control and found that the river chief system could promote the improvement of river water quality. Zhang et al. (2022) empirically investigated the impact of the river chief system on water pollution and its influencing factors using the econometric analysis model and found that the river chief system has a positive impact on pollution control, but it is also restricted by local government environmental expenditure and other aspects.

Through the review of domestic and foreign literature, the author finds that the academic research on the river chief system mainly analyzes the implementation performance of specific policies of the river chief system through quantitative methods or discusses the operational framework of the river chief system from the perspective of histology and sociology, and rarely systematically sorts out and organizes the policy texts from the macro level. Secondly, studies on the river chief system mostly use econometric analysis models to explore the relationship between the river chief system and pollutants and water quality content and rarely use big data methods to study the policy content of the river chief system. Therefore, this paper uses the LDA model to investigate the text of China's river chief system policy, study the topic of this policy, systematically sort out the problems of China's current river chief system policy, and provide a new perspective on the study of river chief system.

3 Methodology

3.1 General Research Framework

This paper will use the big data text clustering method to analyze the construction of China's river chief system from the policy level as the research content. Firstly, the relevant policy texts of the river chief system in China are collected. Secondly, the Python program is used for word segmentation and pause processing, and then the policy topic intensity and regional analysis are realized through relevant operations such as topic visualization, word cloud image, and word frequency comparison. Finally, this paper analyzes the topic structure through LDA and discusses the policy topic of China's river chief system.

3.2 LDA Research Model

LDA model is a typical bag of words model, which considers a document as a collection of words with no order or sequence relationship between words. The LDA model is one of the most common topic models, which is an unsupervised learning algorithm. It does not need manually labeled training sets in the training process and only requires the document set and a specified number of K's for the model construction. A document in the LDA model is generated as follows:

The topic distribution θ_i of the generated document is sampled from the Dirichlet distribution α.

The word distribution $\ddagger_{i,j}$ of the topic is generated by sampling the polynomial distribution of the topic θ_i.

Sampling $\ddagger_{i,j}$ from the polynomial distribution of the word generates the final generated word.

Sampling from the polynomial distribution $\phi_{\ddagger_{i,j}}$ of the word generates the word $\omega_{i,j}$.

4   Experimental Analysis

4.1   Data Collection and Pre-processing

The research data of this study comes from the Law Database of Peking University, which is the most professional database in China and covers the largest number of legal documents and policies. In addition, the author collects policy texts related to the river chief system policy from the portal websites of government departments. The publishers include The State Council, the Ministry of Water Resources, the National Development and Reform Commission, and the Ministry of Environmental Protection. The forms selected for the policies mainly include notices, reports, opinions, suggestions, and guidelines. Meanwhile, only policy texts at the national level are selected, and policies issued by governments at other levels are not within the scope of this study. The national-level policy text is chosen because they are formulated and implemented by the CPC Central Committee and various Ministries and commissions of the State Council, which is the embodiment of collective authority. At the same time, these policies are highly reliable and accurate. In the end, 89 policy texts are collected.

After that, the collected policy text is segmented. This paper uses Regexp Tokenizer to perform word segmentation with regular expressions to remove invalid symbols and non-Chinese characters density words from the document. The frequent occurrence of meaningless words in documents will affect the quality of the collected text data. Therefore, the removal of stop words should be further carried out. This study uses the Chinese stopword list developed by the Harbin Institute of Technology to filter out the Chinese stopwords.

Next, the cleaned policy text is trained through the LDA model, and the initially processed policy text is used as input. First, the parameters that affect topic sparsity and word sparsity are set. Some scholars have compared the advantages and disadvantages of different values of these two variables and concluded that β = 0.1 works better when α = 50. Therefore, this study will refer to this setting method. Once the hyperparameters are set, the optimal number of topics is determined. There are three academic settings for the optimal number of topic selections: using the Bayesian model, the complexity evaluation model, and calculating the consistency of selected topics. This paper uses the complexity method to determine the number of selected topics. The calculation results of the Perplexity degree are shown in Fig. 1. It can be seen that when k = 4, the chart's perplexity is the lowest, so the number of topics is determined to be 4.
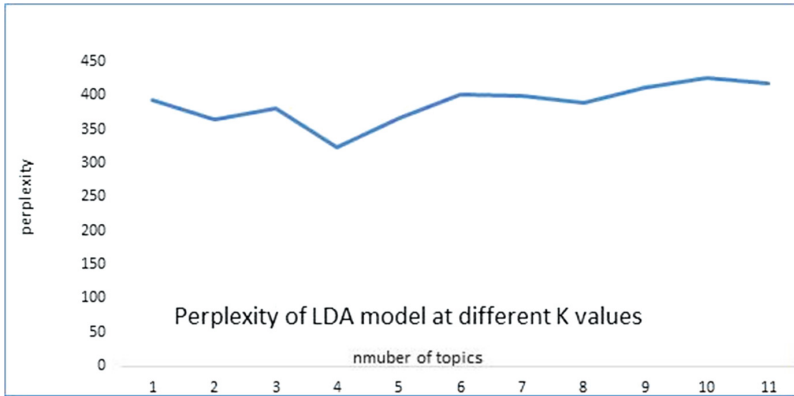
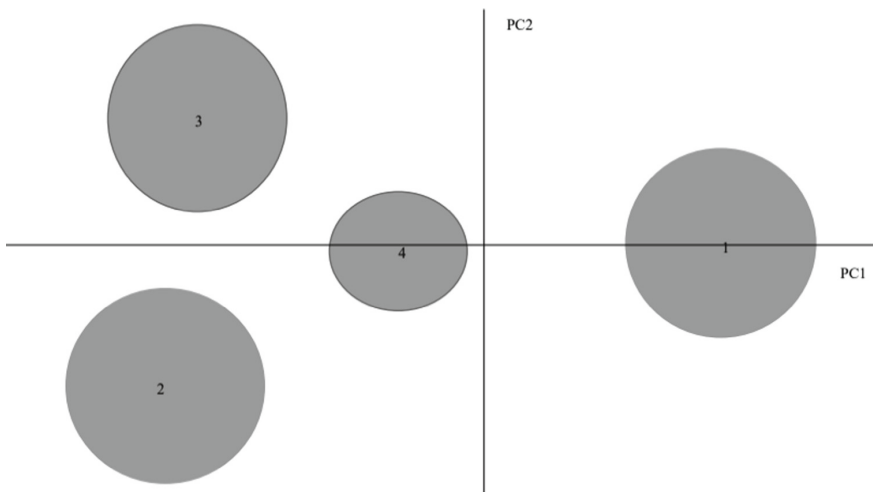**Fig. 1.** Perplexity of LDA models at different K values

Fig. 2. LDA topic model analysis results of China's river chief system policy text

5   Result

5.1   Policy Topic Intensity Analysis

This article calls Python's Ldavis package to analyze policy text visually. Figure 2 shows the inter-office distance map after the multidimensional scaling of Ldavis. In the figure, the distance between the center of each circle indicates the similarity between topics, and the central area indicates the importance of each topic. As shown in.

Table 1, the ranking of the four types of topics from high to low is 30%, 29.8%, 24.8%, and 15.3%, respectively. Topic 1 and 2 have similar strengths. Except for topic 4, which is similar to topics 2 and 3, the circular distance represented by each classified topic is relatively large, indicating a great difference in the topic of river chief system policy.

Table 1. Proportion of topic words in the policy text

Topic	percentage
1	30%
2	29.8%
3	24.8%
4	15.3%

5.2   Policy Topic Area Analysis

This paper uses the exact word segmentation mode to perform text segmentation on the collected policy text based on Python software. After removing invalid characters through the regular expression, the preliminary text filtering is finished. After that, the Chinese stopword list developed by the Harbin Institute of Technology is selected for processing. The author then uses the Counter program under the Collections toolkit to statistically analyze the policy text's word frequency, and the result is shown in Table 2. Then the word cloud can be obtained, which is a form of text visualization. High-frequency words can be displayed more clearly through word cloud images. In a word cloud, the frequency of a word is indicated by the font size. The more frequently the words appear in the policy text, the larger font size, which can visualize the relationship between the high-frequency words and the policy text. The extracted terms are sorted in descending order, and the results are shown in Fig. 3. It can be seen from the word cloud picture that:

"River" and "lake" are the core words.

The words such as "supervision" and "assessment" appear more frequently. The river chief system policy emphasizes supervising and assessing officials' governance performance. River chiefs with outstanding performance will be rewarded, while Those who neglect their duties and responsibilities will be punished. The assessment system with clear rewards and punishments is conducive to motivating them to work.

At the same time, "duty," "responsible," and other words frequently appear in the policy, indicating that the policy of the river chief system emphasizes the implementation of official responsibility, avoids the phenomenon of shirking responsibility, and ensures the maximization of the policy.

The implementation of the river chief system policy is the key to solving the problem of water pollution control and realizing the set policy goals. As can be seen from the word cloud diagram, words such as "coordination" and "implementation" account for a high proportion in the main line of the policy text.

5.3   Policy Topic Structure Analysis

The LDA topic model is used to classify the topics of China's river chief system policy text, and the number of topics is set as 4, and each topic contains ten topic words. The structure of each potential topic can be obtained through the LDA model. This structure can be represented by the probability of each supported subject word. The higher the

Fig. 3. Policy text word cloud image

Table 2. Policy topic of River Chief System in China

Topic	law enforcement and supervision	ecological governance	Performance appraisal	information management
Policy topic of River Chief System in China	Term of words			
	department	pollution	assessment	information
	enforcement	water	chief	service
	system	ecological	situation	support
	shoreline	resource	evaluation	supervisor
	patrol	lake	performance	service
	way	governance	establish	database
	supervision	protection	implement	basic
	report	construction	effect	provide
	oversee	savage	coordinate	dynamic

probability, the higher the word's relevance to the topic. Table 3 shows the combination of the features corresponding to the four topics. Each topic can be regarded as the focus of the river chief system policy. By reviewing the data, the four topics are specified and finally named as "law enforcement and supervision," "ecological governance," "performance appraisal," and "information service."

The featured word in topic one, "law enforcement supervision," is the most relevant word in that topic. The results show that the policy text of the river chief system focuses on the supervision and law enforcement of river and lake pollution. For example, the Opinions on the Full Implementation of the River Chief System issued by The General Office of the State Council in 2016 clearly states that it is necessary to strengthen the

management, protection, and supervision of rivers and lakes, establish and improve the joint law enforcement mechanism between departments, improve the linkage mechanism between administrative law enforcement and criminal justice, and severely crack down on illegal and criminal activities involving rivers and lakes. This shows that China's river chief system policy has maximized its effectiveness by strengthening law enforcement and supervision, maintaining the ecological environment of rivers and lakes, and promoting river ecological restoration and protection.

In the second topic, "ecology," "pollution," and "water" are the most relevant words. It can be seen that the topic is about environmental protection and pollution control. In the related policy text of the river chief system, it is repeatedly emphasized that we should strengthen the water environment, reinforce the protection and comprehensive improvement, improve the river pollution control mechanism, strengthen comprehensive management, strictly control the industrial and mining enterprises, urban non-point source agricultural non-point source, port, pollution, strengthen the treatment of black smelly water bodies. By adjusting measures pro-water ecological shoreline according to local conditions, and ultimately achieve a clean and beautiful overall environment of rivers and lakes, clear water, and green shore.

In the third topic, words such as "evaluation" and "performance" are highly relevant to the topic. Combined with the policy text of the river chief system, the third topic is "performance appraisal." The main purpose of performance appraisal is to enhance the sense of responsibility of leading cadres and promote the quality of leading cadres. The river chief system evaluates the river ecological environment through differential performance evaluation. The evaluation mainly includes river illegal sand discharge regulation, conservation system management, shoreline protection, promotion, etc. The formation and perfection of the performance appraisal system provide an institutional guarantee for the environmental governance of the watershed.

In the fourth topic, the word "information," "database," and "support" are also highly related to the topic, so the topic is about information. The specific approach is to improve the governance performance through inter-departmental information system integration and inter-departmental collaboration in managing systems and other business systems for security applications. Name the topic "information management" with the associated policy text. For example, the General Office of the Ministry of Water Resources issued the Guiding Opinions on the Construction of the River and Lake Information Management System in 2018, emphasizing the need to improve the management of information management and information services while constructing and developing the database system of river and lake management.

6 Conclusion

LDA model has strong expansibility, and its application to public policy research has certain scientific guiding significance. Taking China's river chief system policy as an example, this paper explores the topic strength, field, and structure of the policy through the topic clustering study of LDA. It is found that the policies of the river chief system in China are divided into four topics, which are law enforcement and supervision, ecological governance, performance appraisal, and information management in descending order of

topic intensity, and a complete set of policy systems has been initially formed. However, the policies still have shortcomings and need to be improved.

Public participation is an effective means to improve the performance of environmental governance. In the policy text of China's river chief system, the author can see that although the policy emphasizes the important role of public participation, officials resist public participation by setting a high threshold and refusing to omit key information to avoid responsibility. Moreover, the policy does not explicitly emphasize specific channels of public participation. Therefore, it is difficult to mobilize the enthusiasm of the masses to participate in the supervision of the policy performance.

In 2018, the Ministry of Water Resources issued the policy Opinions on Promoting the River Chief System from "Nominal" to" Practical," which pointed out that the communication and coordination within the river basin should be strengthened. But there are no qualifiable, measurable work requirements. As a result, the joint prevention and control mechanism among basins cannot be properly implemented.

It should be noted that this study has some limitations that need to be improved in the further research process. First, this paper conducts a cluster analysis on the topics of the policy text of China's river chief system. Without a dynamic analysis over time, it is difficult to judge the changes, additions, and deletions of policy topics, and the development trend of future research needs further improvement. Second, in the research process, only the policies at the central level are selected, and there is no comparative analysis of the river chief system policies at all levels. Future research should be further supplemented and improved.

References

Blei, D. M., Jordan, M. I., Griffiths, T. L., & Tenenbaum, J. B. (2004). Hierarchical topic models and the nested Chinese restaurant process. Advances in neural information processing systems, 16.

Li, W. H., Y. D. Zhou, & Z.J. Deng. (2021). The Effectiveness of "River Chief System" Policy: An Empirical Study Based on Environmental Monitoring Samples of China. Water, 13(14)

Liu, W., Gu, J., Zhang, R., & Yang, Y. (2021). Supply and demand matching of financial support policies for private enterprises based on text measurement. Discrete Dynamics in Nature and Society

Liu, Y., and H. Zhao. (2022).Quantitative Evaluation of Policy Based on PMC Index Model: A Case Study of China's Textile Industry Policy. Mathematical Problems in Engineering, 11(2)

Wang, L., Tong, J., & Li, Y. (2019). River chief system (RCS): an experiment on cross-sectoral coordination of watershed governance. Frontiers of Environmental Science & Engineering, 13(4), 64.1–64.3.

Zhang, J.H., W.B. Gui., and J.Q. Zhou, J., Chen, Q., & Han, H. (2020). Discussion on the delimitation of the main river in Anji county under the system of river chief. IOP Conference Series Earth and Environmental Science, 455, 012184.

Zhou, J., Cen, Q., & Han, H. (2020). Discussion on the de-limitation of the main river in Anji county under the river chief system. IOP