Quantitative Evaluation of Big Data Development Policy: Text Data Analysis Based on Coword Network and Policy Tools

Xuefei Hong
School of Internet Finance and Information Engineering, Guangdong University of Finance, Guangzhou 510521, China

Correspondence should be addressed to Xuefei Hong; hongxuefei2006@163.com

Received 18 September 2021; Revised 11 July 2022; Accepted 22 September 2022; Published 11 October 2022

Academic Editor: B. Rajanarayan Prusty

Copyright © 2022 Xuefei Hong. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In the context of the continuous promotion of China’s big data development strategy, this paper quantitively analyses China’s existing national-level big data policies from the perspective of policy instruments and coword networks, discusses the rationality of existing policies, explores ways to improve policies, and provides a reference for the innovation of China’s big data policies. This paper carries out a quantitative textual analysis of China’s national big data policy from the perspective of policy instruments using word frequency analysis to obtain a keyword coword matrix and visualization analysis tools to obtain a coword network. This paper further studies the network node characteristics and structure using social network analysis methods, including degree centrality, clustering analysis, and multidimensional scale analysis, to identify the policy structure and characteristics. Improving big data policy requires improvements in policy instruments on the supply side, resolving existing policy gaps, and strengthening coordination with other policies.

1. Introduction

With the development of cloud computing, the Internet of Things (IOT), mobile Internet, and other technologies, big data can uncover new knowledge from massive datasets and thus create new value in the information age [1]. Big data has a significant impact on economic development, scientific and technological innovation, national security, lifestyles, and so on [2]. The United States, Japan, France, South Korea, Australia, and other countries have issued a series of policies to promote the development of big data [3]. The Chinese government has also started to formulate relevant policies related to data centres, cloud computing, and other aspects. For example, it has promulgated policies such as “Decision on Accelerating the Cultivation and Development of Strategic Emerging Industries” and “Platform for Action to Promote the Development of Big Data.” These policies position cloud computing as a key technology and key development direction for building national information infrastructure and realizing integration and innovation [4]. They propose supporting the research, development, and industrialization of mass data storage and processing technologies to promote the development of cloud computing and Internet of Things technologies.

The development of big data depends on the promotion of national policies [5]. A perfect big data policy system helps to correct market failure, prevent government failure, compensate for system failure, and rectify ethical anomic. Despite the development of big data in China, the construction of laws and regulations continues to lag. Therefore, it is increasingly important to improve big data policies to ensure that China can seize the opportunity of big data development to upgrade its industrial structure and modernize government governance.

Policy documents are the material carriers of policy, and analysing policy documents helps scholars observe policy content, processes, and tools [6]. These policy documents include planning, notices, bulletins, announcements, resolutions, speeches, and research reports. Quantitative analysis has become an important means of studying policy documents, and existing studies have been carried out based on external attributes, language elements, semantic analysis, network analysis, and other aspects. The textual elements of a policy document include the policy title, issuing institution,
policy number, year of publication, and so on. Ritter and Lancaster [7] systematically searched references to IDR and EDR in policy documents to analyse policy impact. Chowdhury and Koya [8] conducted thematic analysis on key United Nations policy documents. Lee et al. [9] analysed Korea’s data privacy system in the context of medical big data and performed a comparative analysis of the legal and regulatory environment for managing health and medical big data. El-Taliawi et al. [10] used bibliometric analysis and subject modelling to analyse the popularity of big data in public policies. Governments around the world are trying to accelerate social and technological change and achieve sustainable development by introducing a policy combination of technology promotion and demand pulling tools [11]. Benchimol et al. [12] reviewed various text analysis methods and introduced several text mining applications for analysing texts. Potter et al. [13] investigated policy documents related to ICT/digital communication and conducted thematic analysis and quantitative research on policy documents.

By consulting existing studies, the particularity of policy documents can be considered. For example, the relationship between the upper and lower levels of policy-issuing agencies and the standardization of policy texts are taken into account. Policy instruments are used to achieve policy objectives. Policy instrument analysis is an important analysis method in big data policy research. Most research on China’s big data policy is purely qualitative, content-based, or quantitative. There is a lack of research on further mining text characteristics using coword analysis under the framework of policy objectives and policy tools. What policies has the Chinese government adopted to develop big data at the national level? What are the basic characteristics of these policies? What policy tools are used in China’s national big data policy? How can scientific policy guide the optimization and improvement of big data development policy? From the perspective of policy tools, this paper uses content analysis and coword network analysis to quantify the text of China’s national-level big data policy to discuss the rationality of this policy and ways to improve it.

First, this paper focuses on China’s national big data development policy and the use of the quantitative textual analysis to systematically analyse the main content and basic characteristics of the policy. Second, this paper adopts the policy target-policy instrument framework to investigate the limitations of China’s national big data policies and propose suggestions and directions for further optimization, which has certain significance as a reference for guiding the Chinese government’s policies on big data development. Finally, this paper constructs a policy analysis research framework to provide research ideas for this research field and to aid in further qualitative and quantitative comprehensive research on big data policy.

2. Literature Review and Theoretical Analysis Framework

2.1. Literature Review. Policy documents are the basis of big data policy analysis and the material carriers of policies, thus providing planning, governance layout, and code of conduct [14]. Scholars have conducted several studies on the big data policy environment, policy description and analysis, policy framework, policy comparative analysis, and references. Zuiderwijk and Janssen [15] constructed a comparative research framework for government data opening policies, including policy environment factors, policy content, performance indicators, and public value, and analysed the similarities and differences in the Dutch government’s open data strategies at different levels of government under this framework. Bertot et al. [16] analysed the challenges of the US information policy framework in terms of big data access and dissemination, digital asset management, archiving and preservation, and privacy and security and proposed suggestions for revising the policy framework. In a country comparison study, Nugroho et al. proposed a transnational comparison framework of government data opening policies to provide a country comparison research framework for the development of open data policy [17]. Ma and Wang [18] compared the main characteristics, obstacles, progress, and effects of big data opening strategies in five countries, namely, Australia, Denmark, Spain, Britain, and the United States, and identified the obstacles and driving factors concerning the implementation of data opening policies.

In examining China’s big data policy, scholars have focused on the content and learning experience of foreign governments’ big data policies, and based on the results, research has gradually deepened to incorporate the specific situation of China’s big data policy. The first category of these studies is policy environment analysis [18]. The second category is policy synergy analysis [19]. The third category is econometrics and the analysis of policy texts [20]. The fourth category is policy comparative analysis and policy system construction [21]. Measurement and analysis based on the policy text will help to clarify the evolution mechanism of cloud manufacturing and provide guidance on the implementation and application of cloud manufacturing [22].

In recent years, scholars have made new progress on this issue, especially with respect to research methods. Existing research on big data policy mainly adopts the perspective of policy instruments. The policy target-policy instrument research framework has been widely applied in the textual analysis of policy documents. For example, Huang et al. used the policy target-policy instrument framework to examine China’s nuclear energy policies [6]. Big data policy tools are the means to achieve big data policy goals. Policy tool analysis is based on policy structural development. Scholars classify policy tools into information, authoritative, organizational, and fiscal tools [23]. Scholars in China have adopted the classification method of Roy Rothwell and Walter Zegveld and classified policy tools into the supply side, demand side, and environment side [24]. In terms of big data policy analysis methods, more achievements have been made [25]. First, research on big data policy using quantitative textual analysis methods, such as Zhou et al. [26], has adopted content analysis methods to analyse policy instruments and discuss the rationality of big data policy in China to explore the path to a perfect policy. Second, the case study method has been adopted to conduct comparative
studies on big data policies. For example, Nugroho et al. adopted the case study method to compare the open data policies of the UK, the US, the Netherlands, Kenya, and Indonesia. Third, text mining technology and social network analysis have been introduced to study big data development policies from multiple perspectives, thereby extending quantitative policy analysis [27]. Coword analysis is an important method for social network analysis [28]. Leng and Han [29] took China’s medical big data policy as the research object and used the social network analysis method and complex network analysis software to measure the node attributes of major policymakers. Zhang and Liu [30] evaluated the big data industrial policy, adopted a quantitative method to examine policy content, and used the PMI index model to conduct a regional quantitative evaluation of China’s provincial big data industrial policy. Wang et al. [31] found that among all policy tools, demand-based policy tools have the most significant impact on innovation efficiency. Yang and Huang [32] described the development and evolution of China’s AI policy based on the research framework of bibliometrics. Cao [33] built a two-dimen- sional analysis framework of policy text from the dimensions of policy tools and the value chain. Zhou et al. [34] conducted policy research from three perspectives: policy intensity, policy tools, and policy objectives.

Text mining techniques also contribute to quantitative research on policy documents [35]. Luhn [36] first proposed automatic classification by word frequency statistics. Maron and Kuhns [37] studied automatic classification. Hotho et al. [38] mapped words into concepts for text mining and proposed a variety of concept mapping strategies. Zhang et al. [39] collected big data from policy texts and conducted network analysis and core periphery analysis based on the keyword dataset. El Haddadi et al. [40] used text mining techniques to analyse and evaluate texts. In recent years, the application of text mining in management has increased exponentially [41]. Topic modelling is the frontier of text mining. Machine learning enables machines to learn rules from a large amount of data input through algorithms to conduct identification and judgement [42]. Weiss and Nemeczek [43] proposed an automatic text analysis method based on unsupervised latent Dirichlet assignment (LDA) topic modelling and dictionary-based emotion analysis. Xing et al. [44] used social network analysis (SNA), machine learning, text clustering, content analysis, and other methods for text mining. Győdi et al. [45] proposed an innovative text mining method to support policy analysts in identifying, defining, and selecting problems.

Research on coword networks can be divided into two main perspectives. One stream of research adopts the level of a single policy document to build a text graph representation model to facilitate text mining, such as extracting keywords and text structure features. The other stream focuses on the set of policy documents to explore text features and themes of the document set. As China gradually proceeds with the development of big data, relevant policies are being constantly created. It is increasingly difficult to analyse the text in policy documents using traditional methods only. Therefore, to compensate for the deficiencies of traditional analysis methods, this paper applies text mining technology combined with social network methods to perform big data policy analysis. With the gradual development of research on text knowledge mining, many scholars have proposed coword analysis as a method of text knowledge mining and have carried out relevant research. However, further study of the coword networks of big data policy is needed.

In summary, scholars have invested substantial effort in constructing a research framework for big data policy, analysed existing big data policies, and gradually expanded the scope of research methods for studying big data policy from policy content to quantitative research, thus providing a reference for further analysis and improvement. However, a systematic analysis of China’s big data policy and the construction of a holistic theoretical research framework are lacking. Research has focused on the big data policies of a certain country or region that warrant further study. With the promotion of China’s national big data strategy, China’s national big data policy plays an important role in guiding and exemplifying the big data policies of local governments, and it needs to be further studied by academic researchers. In terms of research methods, previous studies have focused mainly on the description and analysis of big data policies. Most studies of big data policy have focused on qualitative normative discussions, with quantitative research being used to provide only a simple summary of the nature of the policy text, without in-depth analysis of its content and structure. With the development of social network analysis methods, big data policy documents can be effectively mined, and the use of social network analysis methods to evaluate big data policy needs to be expanded.

To address the above problems, this article performs textual quantitative analysis of China’s national big data policy from the perspective of policy instruments, employs a textual analysis method and coword network analysis method by using word frequency analysis of keywords, and constructs a big data policy word network to discuss the rationality of and ways to perfect China’s big data policy.

2.2. Theoretical Basis and Analytical Framework

2.2.1. Policy Tool Analysis Framework and Policy Target-Policy Instrument Matching Model. Based on the research of Fan and Tan [20], this paper introduces technology promotion and demand pull to establish an interpretive model based on the policy target-policy instrument matching model (PTPTMM), as shown in Figure 1. In this model, policy objectives are the starting point of a policy. In the context of policy-driven innovation, the policy objectives include technological innovation and application innovation. The establishment of policy objectives reflects the degree of policymakers’ preference for technological innovation and applied innovation. When the policy goal focuses on technological development, the policy goal is technological innovation [46]. If the focus is the application of technology in products and markets, the policy goal is application innovation. After the goal is set, the policymaker will develop a policy plan and choose the corresponding
policy tool. Technology-driven policies adopt mostly supply-side policy tools, while demand-pull policies are dominated by demand-side policy tools.

From the perspective of the PTPTMM, supply-side policy tools act mainly on technological innovation, demand-side policy tools act mainly on application innovation, and environment-side policy tools indirectly affect both technological innovation and application innovation [47]. The intensity of the effect of environment-side policy tools is weaker than that of supply-side and demand-side policy tools, and they are represented by dotted lines in the model, as shown in Figure 1.

2.2.2. Quantitative Analysis Framework for Big Data Policy Documents. Social network analysis can help provide an understanding of the overall network architecture and reveal the importance of nodes in the network through the study of degree centrality and network density. Clustering analysis is helpful to describe the clustering effect of nodes in the network. Multidimensional scaling analysis is helpful to analyse the attributes of nodes.

Coword analysis usually uses keyword coword analysis to analyse big data policy topics. The appearance of two keywords expressing the policy theme in the same policy document shows that there is a certain internal relationship between them. The more often these keywords appear in the same policy document, the closer the distance is and the closer the relationship is. Then, based on the relationship between keywords, the structural changes to the big data policy theme are analysed. First, the number of keywords that appear together in the same policy in the big data policy set is counted, and then, a cooccurrence matrix is constructed. In fact, the similarity index is used to calculate the correlation between keywords and generate a correlation matrix. Finally, we use multivariate statistical methods to visualize the results of coword analysis, classify cooccurrence keywords, and summarize and refine the theme of big data policy to reveal the theme structure, hot topics, and development trends of the big data policy samples.

Therefore, on the basis of clustering analysis, multidimensional scaling analysis, and social network analysis, this paper constructs a quantitative textual analysis framework for big data policy, as shown in Figure 2. This comprehensive analysis framework overcomes the shortcomings of the above analysis methods: clustering analysis cannot reflect the status of nodes in the network, multidimensional scale analysis cannot describe the relationship between nodes, social network analysis cannot describe the clustering effect of nodes in the network, and the attribute characteristics of nodes cannot be analysed in depth.

3. Research Methods and Data Sources

3.1. Quantitative Analysis of Policy Documents. This paper examines the degree of matching between policy instruments and policy targets through coding and frequency statistics of policy instruments and policy targets for the big data development policy documents of China’s central government and conducts coword analysis and network analysis of sample policy documents.

Big data policy documents reflect the thinking and behaviour of government departments in developing big data. Scholars always consider how to scientifically analyse big data policy documents. Quantitative research methods for policy documents include content analysis, bibliometrics, statistics, and methods from other disciplines. Quantitative analysis is carried out considering the content and external structural elements of policy documents to reveal public policy issues such as changes in policy theme, the choice and combination of policy tools, and the main cooperation network of the policy process. Specific analysis techniques of policy literature quantification include content analysis, coword analysis, and network analysis. This new policy research method and paradigm can be used for descriptive

![Figure 1: Logic diagram of the big data policy target-policy instrument matching model (PTPIMM) [20].](image-url)
Figure 2: Research framework for big data policy textual quantitative analysis.

analysis of many policies. Academics could also use this method to carry out research in different fields.

In this paper, BIBEXCEL and UCINET are used for quantitative analysis [48]. BIBEXCEL is used to analyse the frequency of keywords and coword analysis. Coword analysis originated in the middle and late 1970s and belongs to a kind of content analysis. The coword analysis method counts the number of times that two words in a group appear in the same text at the same time, reflects the correlation degree between these words with this cooccurrence number, and then analyses the topic structure with the clustering method. The higher the coword intensity is (the number of times two words appear in the same text at the same time), the closer the relationship between the two words is. The number of cowords between keywords forms a coword matrix. Then, we import the word frequency cooccurrence matrix into UCINET for drawing. It visually presents the results of cooccurrence analysis and divides the cooccurrence keywords into groups. Using this method, we can summarize, refine, and abstract the big data policy topics to reveal the theme structure, hot topics, and development trend of the big data policy samples.

Specifically, we repeatedly read the research samples and the corresponding policy interpretation documents and extract high-frequency keywords related to “big data” from the policy text. The sentences in the policy text are processed according to the policy theme-policy purpose-policy tool framework introduced in Section 2. The policies of the research sample are open coded and axis coded.

Then, the keyword frequency is determined, the core keywords are selected from each policy text keyword, and after manual processing and judgement, keywords with broad and unclear meanings are deleted to obtain the core keywords of the overall policy text. A common word matrix is generated according to the frequency of occurrence. According to the concept of the Ochiai coefficient, the correlation coefficient between any two keywords can be calculated to measure the similarity between two keywords. Degree centrality is used to measure the importance of nodes in the network.

Finally, cluster analysis is used to set the partition distance of cluster analysis and obtain the keyword aggregation set. Multidimensional scale analysis uses the idea of dimension reduction, combined with subjective discrimination and classification of subject categories. The research samples are classified and analysed to determine the centre and theme.

3.2. Data Sources. The big data policy texts selected in this paper come from databases such as the website of the People’s Government of China, the websites of provincial people’s governments of China, the websites of the State Council and ministries of China, and the Chinese government’s open information integration service platform. Searching for the keyword “big data” on the website platform, this paper obtains the original policy text and selects 20 documents with big data as the policy target after screening, as shown in Table 1 [26].

4. Unit Coding and Content Analysis of Big Data Policy Documents

4.1. Coding Rule

4.1.1. Policy Targets. Specific policy targets correspond to links among government data governance, big data industry development, and scientific and technological innovation resources. Soares holds that big data governance is part of the broader information governance plan because of the adjustment of multiple functional objectives and formulation of policies for data optimization, privacy protection, and data realization related to big data [49]. Given the above definition, government data governance refers to government departments as the main body formulating data optimization, privacy protection, and data realization policies related to government big data by adjusting a variety of functional objectives [50, 51]. The sharing of scientific and technological innovation data resources refers mainly to the coconstruction, opening, and sharing of scientific and technological innovation data resources through policy regulation and effective management systems and operation mechanisms to promote the effective integration of scientific and technological innovation data resources, improve the comprehensive data utilization, and continuously reduce the cost of scientific and
technological innovation [52]. The big data industry chain refers to “all activities related to the generation and aggregation of big data (data source), organization and management (storage), analysis and discovery (technology), transaction, application, and derivative industry” [53].

The US National Institute of Standards and Technology subdivides the industrial chain of big data according to the process of realizing value through data. Fan and Tan constructed a framework for the industrial chain of big data [20]. Based on the above studies, the present paper constructs a framework for the development of big data and uses it as the basis of classification of policy target coding. According to the literature [50], the industrial chain of big data includes six parts: IT infrastructure, data management, data analysis, data transaction, data application, and data security. This classification framework reflects the development of big data from data to data derivatives and the evolution of big data technology innovation towards big data application innovation.

4.1.2. Policy Instruments. Rothwell and Zegveld’s classification criteria for policy tools on the supply side, environmental side, and demand side are adopted [24]. Coding rules are formulated on the basis of Fan and Tan and Cui et al. [20, 46], as shown in Table 2.

4.1.3. Big Data Policy Content Analysis Unit Coding. In the government’s national-level big data policy documents, content under “Important Tasks” describes the corresponding policy targets. Content under “Policy Guarantee” or “Guarantee Measures” describes the instruments to be adopted by the government. This paper randomly extracts parts of the “Platform for Action to Promote Big Data Development” for precoding. The precoding results show that the coding rules are operable. Then, considering that statements separated by periods in the policy documents have independent meanings, this paper takes the statement as the coding unit and carries out preliminary coding [20, 26]. Finally, we eliminate interference from unrelated sentences by comparing the coding content with the original policy text and reorganizing the coding content. The coding form is Policy Number-Chapter-Code Number, with a total of 60 items, as shown in Table 3.

4.2. Statistics and Analysis of Policy Targets. According to the frequency statistics of China’s national big data policy targets, IT infrastructure is the main policy target of China’s national big data development policy. Text discourse accounts for 50% of all text expressions, followed by data applications, which account for 16%, and data management, which accounts for 14%. By comparison, the frequency of data analysis, data transaction, data security, and other policy objectives together account for 40%, as shown in Figure 3. Thus, IT infrastructure, data application, and data management are the main policy targets of China’s national big data development policy.

4.3. Statistics and Analysis of Policy Tools. The statistics for the frequency of policy instruments show that the proportion of supply-side policy instruments is 14%, that of
environment-side policy tools is 44%, and that of demand-side policy tools is 42%, as shown in Figure 4. Thus, environment-side policy tools are the main means of developing big data at the national level in China, followed by demand-side policy tools and supply-side policy tools. The ratio of demand-to-supply-side policy instruments is 3.

Among secondary indicators, the top six indicators are technology application (demand-side tools), technology support (supply-side tools), technology standards (environment-side tools), regulation (environment-side tools), public services (supply-side tools), and major projects (demand-side tools), as shown by the frequency statistics in Figure 5. The distribution of secondary indicators of supply-side policy instruments used by the government is relatively average and focuses mainly on two aspects: public service

![Table 3: Coding of big data policy content analysis unit (part).](image1)

| Number | Document name | Content analysis unit | Type | Coding |
|--------|---------------|-----------------------|------|--------|
|        |               |                       |      |        |
| 12     | Platform for Action to Promote Big Data Development | Environment side | 12-2-2 |
|        |               |                       |      |        |

**Figure 3:** Frequency distribution of big data policy targets.

**Figure 4:** Proportional map of policy instruments.
and human support. The distribution of environment-side policy instruments is quite different. Target planning accounts for 55%, and there are no relevant policy tools related to “intellectual property.” Thus, although the government uses more supply-side and environment-side policy tools, according to the distribution of its secondary indicators, the six types of supply-side tools are more evenly distributed, while environment-side policy tools focus mainly on regulation. Notably, demand-side policy instruments are concentrated in the two categories of “technology application” and “major projects,” and there are few prices subsidies, trade controls, and external contracting.

In summary, first, quantitative research on the objectives and tools of big data development policy shows that China’s national big data policy focuses mainly on strengthening infrastructure construction and promoting data application, and the status of data application is increasingly prominent. In terms of policy tools, the Chinese government’s national big data policy focuses mainly on demand-side and environment-side policy tools, especially the latter, with limited attention given to supply-side policy tools. Second, the results of the analysis of the matching degree model of policy objectives and policy tools show that the goal of China’s state-level big data development policy is to promote the applied innovation of big data at the product and market levels through environment and demand-side policy tools. Therefore, at present, China’s national-level big data policy tools and policy objectives are well matched. China should also pay attention to the use of demand-side policy tools.

5. Coword Network Analysis of Big Data Policy Keywords

5.1. Keyword Extraction from Big Data Policy Documents. Repeated intensive reading of the 20 policy documents and corresponding policy interpretation documents of the research sample is performed, and high-frequency keywords related to “big data” are extracted from the policy text. First, according to the policy theme-policy target-policy instrument framework established in the second section of the theoretical framework of this paper, the sentences in the policy text are processed, and the policies in the research sample are coded in an open and axial manner. Open coding is the process of assigning concepts and classifying original data into categories. Then, the frequency of each keyword is determined, and the core keywords are selected from each policy text. After manual processing and judgement, keywords with a broad or an unclear meaning are deleted, and the core keywords of the overall policy text are obtained, as shown in Table 4. For example, the frequency of “big data” in policy 12 is 201. In this paper, absolute word frequency notation is adopted, and the value of the corresponding component in the text vector is the number of times the keyword appears in the policy text. First, the high-frequency keywords used in the coword analysis are selected, and then the low-frequency terms are removed.

The threshold value for eliminating low-frequency words is determined by the high-frequency and low-frequency word boundary dividing formula put forward by Donohue in 1973 according to Zipf’s second law, as shown in the following equation [54]:

\[
T = \frac{-1 + \sqrt{1 + 8I_1}}{2},
\]

where \(T\) denotes the threshold value and \(I_1\) denotes the number of keywords that appear only once.

Table 5 lists the threshold values for excluding low-frequency words in each big data policy text.

Figure 5: Composition of three policy tools: (a) supply side; (b) environment side; (c) demand side.
according to the frequency of occurrence in each policy document. According to the concept of the Ochiai coefficient, the correlation coefficient between any two keywords can be calculated to measure the similarity between them. The corresponding coefficients in the common network are calculated according to the following equation [27]:

\[ W(j, k) = \frac{\sum_i H_{ik} H_{ij}}{\sum_i H_{ik} \cdot \sum_j H_{ij}} \]  

(2)

The similarity matrix is shown in Table 6:

5.3. Degree Centrality Analysis of Big Data Policy High-Frequency Keywords. High-frequency keywords in policy texts reflect the value orientation of policies to a large extent [55]. The frequency of keyword occurrence refers to how many policy texts the keyword appears in, which reflects the popularity of the keyword. Point degree centrality refers to how many nodes corresponding to a keyword are connected with other nodes in the keyword cooccurrence network, which reflects the degree of correlation between the keyword and other keywords. Degree centrality is used to measure the importance of nodes in the network. The higher the degree centrality of nodes in the network is, the more important the nodes are in the network [27]. The absolute degree centrality of the nodes in the undirected network is the sum of the number of connections of each node, which is greatly affected by network size. Therefore, to eliminate the influence of network size on degree centrality, the relative centrality of nodes can be obtained by dividing the sum of the numbers of node links by the maximum possible number of nodes. The degree centrality value of keywords in the big data policy text is shown in Table 7.

As shown in Table 7, “security,” “innovation,” “industry,” and other keywords are considerably more important in the network than other keywords. “Strategy,” “data,” and other keywords in the network have approximately the same degree of importance. A visualization of the keyword network and its degree centrality is provided in Figure 6.

The value orientation of China’s big data policy is as follows. (1) First, security is an important premise. The frequency of the keyword “safety” is very high, ranking first, with a frequency of 10, and the corresponding point degree centre is also the highest. This conclusion is consistent with Zhang et al. [55]. On February 27, 2014, China’s government established the central leading group on network security and informatization. The network security law promulgated on November 7, 2016, emphasizes that China’s informatization work and big data development must be based on security guarantees. (2) Second, data fusion is a means of development, as reflected by keywords such as “public,” “resources,” “data,” “Internet,” and “network.” These corresponding points are also very central. Data technology can realize data fusion and opening through information system integration, intensive construction, and overall management across regions and departments. (3) Third, industry convergence is the development path, which is reflected by keywords such as “industry” and “market.” (4) Fourth, as a development idea, technological innovation manifests in keywords such as “technology” and “innovation.” Technological innovation helps promote the development of the big data industry and government data governance.

---

**Table 4: Statistical analysis of big data policy keywords and word frequency.**

| Policy number | Keywords     | Frequency | Policy number | Keywords     | Frequency |
|---------------|--------------|-----------|---------------|--------------|-----------|
| 1             | Industry     | 113       | 12            | Big data     | 201       |
| 1             | Strategy     | 52        | 12            | Security     | 73        |
| 1             | Innovation   | 47        | 12            | Share        | 59        |
| 2             | Network      | 189       | 14            | Standard     | 370       |
| 2             | Resources    | 167       | 14            | Technology   | 129       |
| 2             | Security     | 166       | 14            |             | 79        |
| ...           | ...          | ...       | ...           | ...          | ...       |
| 11            | Internet     | 155       | 20            | Data         | 105       |
| 11            | Innovation   | 60        | 20            | Scientific data management | 19 |
| 11            | Financial    | 38        | 20            | Security     | 18        |

**Table 5: Threshold values for excluding low-frequency words.**

| Policy number | Threshold of excluding low frequency | Policy number | Threshold of excluding low frequency |
|---------------|-------------------------------------|---------------|-------------------------------------|
| 1             | 17                                  | 11            | 24                                  |
| 2             | 33                                  | 12            | 23                                  |
| 3             | 33                                  | 13            | 44                                  |
| 4             | 14                                  | 14            | 36                                  |
| 5             | 21                                  | 15            | 27                                  |
| 6             | 41                                  | 16            | 15                                  |
| 7             | 24                                  | 17            | 23                                  |
| 8             | 24                                  | 18            | 65                                  |
| 9             | 21                                  | 19            | 23                                  |
| 10            | 24                                  | 20            | 16                                  |

---
5.4. Cluster Analysis of Big Data Policy. Cluster analysis divides the research sample set into multiple clusters constructed by similar objects [56]. According to their own data state, research objects will automatically aggregate and form a class without any prior artificial setting. The research objects of each kind of collection have similar characteristics, and there are certain differences between each kind of collection. Clustering analysis is conducted through the matrix of the Chinese big data policy keyword co-word network. Based on the research content of this paper and referring to existing research [27], 5 categories of keywords are obtained. According to the keyword connotations, the five categories are system construction, industrial integration, innovation environment, technological development, and external environment.

5.5. Multidimensional Scaling Analysis of Big Data Policy. Multidimensional scaling analysis uses the idea of dimensionality reduction to classify and analyse research samples. Multidimensional scaling analysis is used to understand the internal relationship between samples while retaining the statistical analysis method of the original relationship between samples [55, 57]. Big data policy keywords are presented in a two-dimensional atlas in the form of an aggregation. According to the connotation and extension of the keywords in China’s national big data policy, starting from the first quadrant, the four sets of key words are summarized as follows: external support and overall layout, industrial development and application, infrastructure construction, and data management. Overall, the keywords of China’s national big data policy are evenly distributed. Based on the above analysis, the big data policy common words network is summarized into “one centre, five themes.” A core area is formed around big data with keywords such as data security, application, and innovation. The five themes are innovation environment, infrastructure construction, big data application, institutional guarantee, and industrial integration.

| No. | Keywords   | Degree | Nrmdegree | Share |
|-----|------------|--------|-----------|-------|
| 1   | Security   | 93     | 17.547    | 0.039 |
| 8   | Innovation | 93     | 17.547    | 0.039 |
| 5   | Industry   | 85     | 16.038    | 0.036 |
| 54  | Enterprise | 69     | 13.019    | 0.029 |
| 81  | Information| 66     | 12.453    | 0.028 |
| 107 | Resources  | 63     | 11.887    | 0.027 |
| 32  | Technology | 61     | 11.509    | 0.026 |
| 67  | Market     | 56     | 10.566    | 0.024 |
| 14  | Engineering| 55     | 10.377    | 0.023 |
| 16  | Public     | 49     | 9.245     | 0.021 |
| 72  | Network    | 48     | 9.057     | 0.02  |
| 100 | Quality    | 42     | 7.925     | 0.018 |
| 48  | Energy     | 40     | 7.547     | 0.017 |
| 90  | Application| 40     | 7.547     | 0.017 |
| 40  | Economy    | 39     | 7.358     | 0.016 |
| 61  | Facilities | 38     | 7.17      | 0.016 |
| 26  | Environmental| 37    | 6.981     | 0.016 |
| 51  | Agriculture| 36     | 6.792     | 0.015 |
| 19  | Plan       | 35     | 6.604     | 0.015 |
| 95  | Strategy   | 34     | 6.415     | 0.014 |
| 62  | Sociology  | 34     | 6.415     | 0.014 |
| 68  | Data       | 33     | 6.226     | 0.014 |
6. Conclusions

At present, the Chinese government’s national big data development policy aims to promote the application of big data in industrial transformation, upgrade and strengthen the integration of big data, and improve national governance by constantly opening the data resources owned by the government. In short, from the perspective of policy objectives, the government aims to promote continuous innovation in the application of big data based on the openness and management of data sources. In terms of policy tools, central policymakers rely more on environment and demand-side policy tools than the other two types of tools. Therefore, there is a basic match between policy tools and policy objectives.

This paper draws three main conclusions. First, quantitative analysis of the goals and tools of big data development policy shows that China’s national big data development policy adopts strengthening infrastructure construction and data application innovation as the policy objectives, with data application playing an increasingly prominent role. Second, in terms of policy tools, the Chinese government’s national big data policy is based mainly on demand-side and environment-side policy tools and takes the them as the main means to achieve the goal. Environment-side policy tools are particularly prominent, while there are few supply-side policy tools. Third, analysis of the matching degree model of policy objectives and policy tools shows that the target-tool matching of China’s state-level big data development policies is basically consistent. Therefore, at present, China’s national big data policy tool selection and policy objectives are well matched, but China should also pay attention to the use of demand-side policy tools.

Based on the above conclusions, this paper proposes suggestions to improve China’s big data policy. First, existing national big data policy gaps must be addressed. At present, China’s national big data policy focuses mainly on infrastructure construction. Although there are policies for specific industries, there are still policy gaps. Therefore, the formulation of future big data policy should start by filling in policy gaps in issued policies. Second, the frequency of supply-side policy tools should be increased. The purpose of supply-side policy tools is to improve the supply factors of big data development, such as increasing financial support and human resource support, to promote the development of big data technology. This financial support is for basic research and generic technology research and development in scientific and technological innovation. Supply-side policy tools account for only 14%, leaving considerable room for improvement. Therefore, future big data policymaking should improve supply-side policy. Third, the coordinated development of big data industry development policy, scientific data management policy, and government data governance policy should be strengthened. Big data is closely related to cloud computing, the Internet of Things, mobile Internet, and artificial intelligence. The development of big data is also inseparable from data disclosure, data sharing, and data governance by the government and enterprises. Therefore, big data policies should take big data as the core and coordinate big data industrial development policy, scientific
data management policy, and government data governance policy to avoid overlooking or overlapping policy elements and thus wasting government resources.

The research contributions of this paper are threefold. First, this paper focuses on China’s national big data development policy and uses a quantitative research method to systematically analyse the main content and basic characteristics of the policy. Second, this paper uses the target-tool matching degree of the policy as the analysis framework to investigate the limitations of China’s national big data development policies and provides suggestions and directions for further optimization, which has certain significance for guiding China’s national government’s big data policy development. Finally, this paper constructs the policy toolword network research framework, which provides new research ideas for quantitative comprehensive research on big data policy.

The limitations of this paper are as follows. First, the process of identifying policy objectives and policy tools requires a large amount of work, and coding and classification take a long time. Second, given the workload of text analysis and the process of policy formulation and implementation in China, as well as the early stage of the development of big data policy, our greatest concern at present is national big data policy. Therefore, as local governments gradually improve the big data policy system under the guidance of the central government, we can carry out further research to develop a classification matching method based on a machine learning algorithm to automatically identify the policy objectives and policy tools of big data. In addition, empirical investigations of the impact of policy objectives and policy tools on the development of the big data industry can open up a new research field.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The author declares that there are no conflicts of interest regarding the publication of this paper.

References

[1] C. Pontones-Rosa, R. Pérez-Morote, and J. F. Santos-Peña1ver, "ICT-based public policies and depopulation in hollowed-out Spain: a survey analysis on the digital divide and citizen satisfaction," Technological Forecasting and Social Change, vol. 169, Article ID 120811, 2021.
[2] R. K. Goel and S. Vishnoi, “Urbanization and sustainable development for inclusiveness using ICTs,” Telecommunications Policy, vol. 46, no. 6, Article ID 102311, 2022.
[3] A. Ingrams, “Big data and dahl’s challenge of democratic governance,” The Review of Policy Research, vol. 36, no. 3, pp. 357–377, 2019.
[4] J.-S. Lacam and D. Salvetat, “Big data and Smart data: two interdependent and synergistic digital policies within a virtuous data exploitation loop,” The Journal of High Technology Management Research, vol. 32, no. 1, Article ID 100406, 2021.
[5] S. Athey, “Beyond prediction: using big data for policy problems,” Science, vol. 355, no. 6324, pp. 483–485, 2017.
[6] C. Huang, C. Yang, and J. Su, “Policy change analysis based on “policy target–policy instrument” patterns: a case study of China’s nuclear energy policy,” Scientometrics, vol. 117, no. 2, pp. 1081–1114, 2018.
[7] A. Ritter and K. Lancaster, “Measuring research influence on drug policy: a case example of two epidemiological monitoring systems,” International Journal of Drug Policy, vol. 24, no. 1, pp. 30–37, 2013.
[8] G. Chowdhury and K. Koya, “Information practices for sustainability: role of iSchools in achieving the UN sustainable development goals (SDGs),” Journal of the Association for Information Science and Technology, vol. 68, no. 9, pp. 2128–2138, 2017.
[9] D. Lee, M. Park, S. Chang, and H. Ko, “Protecting and utilizing health and medical big data: policy perspectives from Korea,” Healthcare informatics research, vol. 25, no. 4, pp. 239–247, 2019.
[10] O. G. El-Taliawi, N. Goyal, and M. Howlett, “Holding out the promise of Lasswell’s dream: big data analytics in public policy research and teaching,” The Review of Policy Research, vol. 38, no. 6, pp. 640–660, 2021.
[11] A. Nuñez-Jimenez, C. Knoeri, J. Hopmann, and V. H. Hoffmann, “Beyond innovation and deployment: modeling the impact of technology-push and demand-pull policies in Germany’s solar policy mix,” Research Policy, vol. 51, no. 10, Article ID 104585, 2022.
[12] J. Benchimol, S. Kazinnik, and Y. Saadon, “Text mining methodologies with R: an application to central bank texts,” Machine Learning with Applications, vol. 8, Article ID 100286, 2022.
[13] R. E. Potter, M. Dollard, S. Pignata, A. Zadow, and K. Lushington, “Review of practice & policy strategies for managing digital communication and ICT use in Australian universities,” Computers in Human Behavior Reports, vol. 5, Article ID 100160, 2022.
[14] C. R. Featherston and E. O’Sullivan, “Enabling technologies, lifecycle transitions, and industrial systems in technology foresight: insights from advanced materials FTA,” Technological Forecasting and Social Change, vol. 115, pp. 261–277, 2017.
[15] A. Zuiderwijk and M. Janssen, “Open data policies, their implementation and impact: a framework for comparison,” Government Information Quarterly, vol. 31, no. 1, pp. 17–29, 2014.
[16] J. C. Bertot, U. Gorham, P. T. Jaeger, L. C. Sarin, and H. Choi, “Big data, open government and e-government: issues, policies and recommendations,” Information Polity, vol. 19, no. 1-2, pp. 5–16, 2014.
[17] R. P. Nugroho, A. Zuiderwijk, M. Janssen, and M. de Jong, “A comparison of national open data policies: lessons learned,” Transforming Government: People, Process and Policy, vol. 9, no. 3, pp. 286–308, 2015.
[18] H. Ma and H. Wang, “SLEPT analysis and strategic deployment of open government data strategy in China,” Information Science, vol. 34, no. 3, pp. 3–8, 2016.
[19] M. Lang, “The matching degree between function of local government and central government under big data perspective: a research based on the LDA model of gansu province,” Journal of Intelligence, vol. 37, no. 9, pp. 78–85, 2018.
[20] Z. Fan and H. Tan, “Big data development strategies of Chinese local governments based on documents quantitative methods— compatibility of policy goals and policy tools,” Chinese Public Administration, vol. 390, no. 12, pp. 46–53, 2017.
[21] M. Hu and L. Sun, “Open data policy of the research funded by public science institutes in UK and its enlightenments,” Science Foundation in China, vol. 32, no. 05, pp. 93–98, 2018.
[22] J. Wang, “Research on sustainable evolution of China’s cloud manufacturing policies,” Technology in Society, vol. 66, Article ID 101639, 2021.
[23] J. Hopfmann, M. Peters, M. Schneider, and V. H. Hoffmann, “The two faces of market support—how deployment policies affect technological exploration and exploitation in the solar photovoltaic industry,” Research Policy, vol. 42, no. 4, pp. 989–1003, 2013.
[24] R. Rothwell and W. Zegveld, Reindustrialization and Technology, pp. 83–104, Logman Group Limited, Harlow, UK, 1985.
[25] G. Galimberti, A. Manisi, and G. Soffritti, “Modelling the Role of Variables in Model-Based Cluster Analysis,” Statistics & Computing, vol. 28, pp. 1–25, 2017.
[26] J. Y. Zhou, H. N. Zhang, Y. R. Huang, M. H. Li, and R. Liu, “Policy tool based textual quantitative analysis on big data policy in China,” Information Research, vol. 230, no. 12, 2016.
[27] L. Xu, Q. Li, and X. Xiao, “The Co-word network analysis of big data policies based on grounded theory,” Journal of Modern Information, vol. 38, no. 6, pp. 157–164, 2018.
[28] R. N. Kostoff, “Co-word analysis,” in Proceedings of the 2021 5th International Conference on Education, Management and Social Science (EMSS 2021), Suzhou, China, March 2021.
[29] L. Zhang and M. Liu, “Regional comparative research on China’s big data industry policy—empirical analysis based on 39 policy texts,” Scientific Journal of Economics and Management Research, vol. 3, no. 5, 2021.
[30] J. Wang, X. Ma, Y. Zhao, J. Zhao, and M. Heydari, “Impact of scientific and technological innovation policies on innovation efficiency of high-technology industrial parks—a dual analysis with linear regression and qca,” International Journal of Innovation Studies, vol. 6, no. 3, pp. 169–182, 2022.
[31] C. Yang and C. Huang, “Quantitative mapping of the evolution of AI policy distribution, targets and focuses over three decades in China,” Technological Forecasting and Social Change, vol. 174, Article ID 121188, 2022.
[32] M. Cao, “Policy analysis and optimization method of Chinese animation industry management based on policy tools,” Mathematical Problems in Engineering, vol. 2022, pp. 1–13, Article ID 1903710, 2022.
[33] X.-X. Zhou, Z.-H. Zhu, and C. Feng, “The evolution of land policies in China from 1980 to 2019: a policy-text based analysis,” Environmental Science and Pollution Research, vol. 29, no. 36, pp. 54902–54915, 2022/03/21.
[34] C. Huang, X. Yue, J. Chen, W. Xu, and J. Li, “The effect of open access” on journal impact factors: a causal analysis of medical journals,” Physica A: Statistical Mechanics and Its Applications, vol. 533, Article ID 122043, 2019.
[35] H. P. Luhn, “The automatic creation of literature abstracts,” IBM Journal of Research and Development, vol. 2, no. 2, pp. 159–165, 1958.
[36] M. E. Maron and J. L. Kuhns, “On relevance, probabilistic indexing and information retrieval,” Journal of the ACM, vol. 7, no. 3, pp. 216–244, 1960.
[37] A. Hotho, S. Staab, and G. Stumme, “Ontologies improve text document clustering,” in Proceedings of the Third IEEE International Conference on Data Mining, pp. 541–544, Melbourne, FL, USA, November 2003.
[38] D. Zhang, F. Fan, and S. D. Park, “Network analysis of actors and policy keywords for sustainable environmental governance: focusing on Chinese environmental policy,” Sustainability, vol. 11, no. 15, pp. 4068–4129, 2019.
[39] T. El Haddadi, O. El Haddadi, T. Mourabit, A. El Allaoui, and M. Ben Ahmed, “Automatic analysis of the sustainability of public procurement based on Text Mining: the case of the Moroccan ICT markets,” Cleaner and Responsible Consumption, vol. 3, Article ID 100037, 2021.
[40] G. Song, J. Wu, and S. Wang, “Text mining in management research: a bibliometric analysis,” Security and Communication Networks, vol. 2021, pp. 1–15, Article ID 2270276, 2021.
[41] Y. Xu and H. Ma, “Research and implementation of the text matching algorithm in the field of housing law and policy based on deep learning,” Complexity, vol. 2021, Article ID 31, 9 pages, 2021.
[42] D. Weiss and F. Nemeczek, “A text-based monitoring tool for the legitimacy and guidance of technological innovation systems,” Technology in Society, vol. 66, Article ID 101686, 2021.
[43] Y. Xing, X. Wang, C. Qiu, Y. Li, and W. He, “Research on opinion polarization by big data analytics capabilities in online social networks,” Technology in Society, vol. 68, Article ID 101902, 2022.
[44] K. Gyödi, L. Nawaro, M. Paliński, and M. Wilamowski, “Informing Policy with Text Mining: Technological Change and Social Challenges,” Quality & Quantity, 2022.
[45] H. Cui, X. Lei, Z. Xiaotian, and J. Su, “A framework of integrated policy-technology roadmap (P-trm) and its use: examples of wind turbine and solar PV industries,” China Soft Science, vol. 5, pp. 76–84, 2014.
[46] C. Huang, T. Ren, and J. Zhang, “Policy documents quantitative research: a new direction for public policy study,” Journal of Public Management, vol. 12, no. 2, pp. 129–137+158, 2015.
[47] O. Persson, R. Danell, and J. W. Schneider, “How to Use Bibexcel for Various Types of Bibliometric Analysis,” Celebrating Scholarly Communication Studies a Festschrift for Olle Persson at His Birthday, ISSI, Bern, Switzerland, 2009.
[48] S. Soares, Big Data Governance. An Emerging Imperative, MC Press, USA, 2012.
[49] B. Liu, W. Wei, and X. An, “Analysis of government data governance policy in the era of big data,” Journal of Intelligence, vol. 38, no. 1, pp. 142–147, 2019.
[50] Y. Zhang, X. An, and G. Zhang, “Construction of Maturity Evaluation Index System for Government Big Data Governance,” Information and Documentation Services, vol. 40, no. 1, pp. 28–32, 2018.
[51] L. Wan and C. Huang, “Influencing factors and model for the regional science and technology innovation resources sharing:—a research based on theories of path dependence and innovation diffusion,” Journal of Intelligence, vol. 35, no. 1, pp. 182–187+181, 2016.
[52] Y. Gui, “White Paper on China’s Big Data Trading Industry,” 2016, http://www.cbdi.com/BigData/2016-06/02/content_4965656.html.
[53] J. C. Donohue, Understanding Scientific Literature: A Bibliometric Approach, MIT Press, Cambridge, Mass, 1973.
[55] H. Zhang, N. Guo, and X. Tang, "Promoting logic and future directions: a textual analysis of China’s government big data policy," *Journal of Intelligence*, vol. 37, no. 3, pp. 46–53, 2018.

[56] Y. Sun and L. Feng, "The logical evolution of public policy on internet governance in China from 1994 to 2014," *Journal of Intelligence*, vol. 36, no. 1, pp. 87–91, 2017.

[57] A. Wu, M. Dong, and H. Li, “Analysis on policy structure of ’internet + and’ mass entrepreneurship and innovation”: based on grounded theory and Co-word analysis,” *Science and Technology Management Research*, vol. 38, no. 10, pp. 44–52, 2018.