Research Article

Policy Evolution of Government Procurement of Public Services in China: A Text-Mining Perspective

Dongdan Zhu,1,2 Yuting Zhang,2 and Zhengnan Lu2

1Jingjiang College, Jiangsu University, Zhenjiang 212003, China
2School of Management, Jiangsu University, Zhenjiang, Zhenjiang 212003, China

Correspondence should be addressed to Yuting Zhang; 15751773779@163.com

Received 25 November 2022; Revised 27 March 2023; Accepted 29 April 2023; Published 5 June 2023

Copyright © 2023 Dongdan Zhu et al. 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.

The in-depth mining of policy text of Government Procurement of Public Services (GPPS) is helpful to distinguish stage characteristics and evolution logic of the policy. Through text-mining technology, the current research analyzes the policy text of GPPS from 1995 to 2021 in China. Firstly, the GPPS policy is divided into three stages according to the key policy nodes. Secondly, the TF-IDF algorithm is adopted to obtain keywords at each stage, and the static stage characteristics are summarized by constructing the complex network of the extracted keywords. Finally, the policy is clustered into several categories with the help of K-means cluster analysis, and the characteristic of each category is achieved through secondary word segmentation, so as to figure out the dynamic evolution logic of each policy category at divided stages. Results show that the development of the GPPS policy in China presents a point-to-face change feature, manifested in the evolution logic of “government purchase—government procurement of public service—all-round supporting policy.” And policy priorities at different stages, namely, policy tools, will change according to the development of economy and variation of demands.

1. Introduction

Government Procurement of Public Service (GPPS) is an important way to improve the quality of public services and satisfy the diverse demands of the public. China began to implement GPPS in the 1990s. As a means of public governance, the policy of GPPS has become an important way to promote the construction of a service-oriented government. Public governance activities must be guided through the public policy. Policy formulation of GPPS can provide legal and institutional guarantee for the development of GPPS. Exploring the new anchor point of GPPS, standardizing the policy implementation procedures, and further improving the quality of public services are largely dependent on the policy of GPPS [1].

Nowadays, the research on GPPS is mostly related to the practical operation and governance, such as the purchase boundary [2], purchase decision factors [3], purchase mode [4], risks [5], and performance evaluation [6]. There is little research on GPPS from the perspective of policy text. The policy is the basis for the implementation of GPPS, and the policy text is the core element of the policy. The in-depth mining of policy text of GPPS can find out the logic of policy evolution and objectively analyze the inconsistency problems between the policy content and practice. In the process of policy improvement, what is the evolutionary trend of GPPS? What are the phased characteristics? In order to answer the above questions, the current research selects 86 representative policies of GPPS from 1995 to 2021 as research samples. With the help of text-mining technology, the static stage characteristic and dynamic evolution logic of GPPS can be extracted and summarized, so as to provide references for improving the policy system of GPPS.

Different from the previous studies, the current research mainly achieves the following expansions and innovations:

1) From the perspective of policy text, this paper conducts an in-depth analysis of GPPS. Nowadays, academia pays much attention to the problems in the actual operation of GPPS, ignoring the policy value.
The current research excavates the policy texts deeply with the text-mining method and the qualitative interpretation method, presenting the stage characteristic and the evolution process of the GPPS policy.

(2) The current research combines the static interpretation and dynamic interpretation of the GPPS policy. In the existing research on policy evolution in other fields, the conventional practice is to obtain the key words of the policy at each stage and summarize the policy evolution with the achieved key words, which belongs to the category of static research. This paper is not limited to the static investigation. It explores the evolutionary logic of GPPS from the static and dynamic perspective. Firstly, the GPPS policy is divided into several stages with the representative policy as the key node. Secondly, the TF-IDF algorithm is adopted to extract the keywords at each stage, and then the word2vec algorithm is applied to calculate the similarity between keywords, so as to construct a complex network of keywords. Thus, static stage characteristics of GPPS can be summarized. Finally, the K-means cluster analysis method is used to cluster the policy texts of GPPS, observing the dynamic evolution law of each cluster at each stage. Therefore, the comprehensive interpretation of GPPS is analyzed from both static and dynamic perspective.

2. Literature Review


2.2. Policy Research. Policy research is beneficial to grasp the connotation and future trend of a policy from a macro perspective. There are abundant research achievements in the field of policy research, such as policy diffusion, policy tools, policy evaluation, and policy text mining.

Research on policy diffusion mainly focuses on the diffusion path and the diffusion mechanism. When it comes to the policy diffusion path, the representative research achievements abroad mainly include the national interaction model, the regional diffusion model, the lead-follow model and the vertical impact model, which are proposed by Berry and Berry [23]. Based on the practice of policy diffusion in China, Wang and Lai [24] divides the policy diffusion path into four types, namely, the top-down hierarchical diffusion model, the bottom-up policy promotion model, the diffusion model between regions and departments, and the diffusion model from policy leading areas to policy follow-up areas. The diffusion mechanism is another research hotspot of policy diffusion. Dobbin et al. [25] proposes coercive force,
the constructivism theory, the competition theory, and the learning theory to explain the phenomenon of policy diffusion.

The policy tool is the refinement and deepening of policy at the level of tool science. It has become one of the important branches in the field of policy research. The most classic classification of policy tools is put forward by Rothwell and Zegveld [26], which are supply type, demand type, and environment type. What is more, according to the government’s application of resources, policy tools can be divided into information type, organizational type, and financial type [27].

Policy evaluation is a comprehensive investigation of the policy system through scientific methods. According to the time when the evaluation occurs, policy evaluation can be divided into three categories: prepolicy evaluation [28], which is to evaluate the policy plan itself; whole-process evaluation [29], which is to evaluate the entire policy implementation process; and postpolicy evaluation [30], which is to evaluate the effect of the implementation of the policy. Common policy evaluation methods can also be divided into qualitative evaluation and quantitative evaluation [31]. As a quantitative evaluation method for the policy itself before implementation, the PMC index model has been widely used in the field of policy evaluation in recent years because of its rationality [32, 33].

In terms of policy text, scholars at home and abroad began to apply the text-mining technology into policy text. For example, the social change analysis based on the government work report [34], the theme analysis of the national science and technology innovation policy [35], and the evolution research of the standardization policy [36] are all the exploration into the policy text with the help of text mining. Relevant text-mining studies on policy documents are summarized in Table 1.

2.3. Summary of Literature Review. The existing research on GPPS mainly focuses on the practical process, risks, and performance evaluation of the policy, lacking the exploration into the policy text. Excavating the policy text of GPPS is beneficial to deeply understand the policy from a macro perspective.

3. Research Design

3.1. Research Method. The current research adopts text mining to analyze the policy texts of GPPS. Text-mining technology is a method that aims to excave large-scale data to meet users’ needs through information retrieval and machine learning. Text mining can explore the potential laws and associations from unstructured texts and then help to further extract valuable information [41]. As the policy texts of GPPS have the characteristics of large length, relying solely on human beings to read the policy texts will have a huge workload and lack of objectivity. Text mining can overcome problems such as huge amount of policy text. Therefore, with the support of text-mining technology, the current research comprehensively applies the TF-IDF algorithm, the complex network, and cluster analysis to explore the static stage characteristics and dynamic evolutionary logic of GPPS with Python software. The research process of exploring the policy of GPPS is shown in Figure 1.

3.1.1. Keyword Extraction. In order to investigate the stage characteristics of the GPPS policy, with the divided policy stages, the TF-IDF algorithm is adopted to extract the keywords of the policy texts at each stage. The TF-IDF algorithm does not measure the criticality of a word according to the word frequency [42]. The essence of TF-IDF can be summarized as follows: If a word appears in a text with a high frequency and seldom appears in other texts in the corpus, it is considered that the word is critical and important. The TF-IDF algorithm can not only conduct word segmentation but also calculate the TF-IDF value of the segmented words. The TF-IDF algorithm can not only be used to extract key words from policy texts but also be applied to calculate text similarity, analyze network public opinion, conduct topic mining, and other fields. Therefore, through the keyword extraction with TF-IDF, the static stage characteristics of GPPS can be achieved.

3.1.2. Complex Network of Keywords. Constructing the complex network of keywords can further clarify the correlation among keywords, enriching and improving the static stage characteristics of GPPS. On the basis of the obtained keywords at each stage, the similarity between keywords can be calculated with the word2vec algorithm. The word2vec algorithm is a distributed expression, which calculates the similarity among terms by considering the context of terms. In order to obtain the correlation among keywords and figure out the relationship among keywords of policy text at each stage, the edges are constructed among keywords with similarity greater than 0.8 and the corresponding keywords are determined as nodes, so as to build a complex network of keywords at each stage.

3.1.3. Cluster Analysis. The current research applies the K-means cluster analysis method to investigate the dynamic evolution process of GPPS. Firstly, the elbow method is used to determine the number of categories. Secondly, the package of K-means is imported from the library of sklearn.cluster, so as to cluster the policy texts. Finally, the cluster results are compared with secondary word segmentation and frequency statistics, so as to obtain the text characteristics of each category. Based on the above operations, the text characteristics of each category are acquired and the dynamic evolution logic of GPPS can be interpreted by calculating the keyword frequency of each category at each stage.

3.2. Research Data

3.2.1. Data Sources. Before collecting the policy texts, it is worth emphasizing that the research objects will not be limited to a specific service form, such as government procurement of elderly-care services or government procurement of preschool
TABLE 1: Text-mining studies on policy documents.

<table>
<thead>
<tr>
<th>Research object</th>
<th>Research contents</th>
<th>Specific methods</th>
<th>Authors</th>
</tr>
</thead>
<tbody>
<tr>
<td>National science and technology innovation policy in China</td>
<td>Policy evolution and policy theme analysis</td>
<td>K-means clustering analysis</td>
<td>Zhang et al. [35]</td>
</tr>
<tr>
<td>Subsidy policy for new energy vehicles in China</td>
<td>Policy tools and policy evaluation</td>
<td>TF-IDF algorithm and PMC method</td>
<td>Zhang and Zhou [37]</td>
</tr>
<tr>
<td>Climate policy in China, the US, and EU</td>
<td>Quantitative analysis of policy themes</td>
<td>LDA model and N-gram model</td>
<td>Yang et al. [39]</td>
</tr>
<tr>
<td>High-tech industrial policy in China</td>
<td>Research on hierarchical characteristics and evolution</td>
<td>Coword analysis and LDA model</td>
<td>Wang and Sun [40]</td>
</tr>
<tr>
<td>AI policy in China</td>
<td>Policy status and frontier trends</td>
<td>LDA model and NLP</td>
<td>Zhang et al. [40]</td>
</tr>
<tr>
<td>Monetary policy in the US</td>
<td>Policy evolution and stability</td>
<td>Text-mining techniques</td>
<td>Wischnewsky et al. [38]</td>
</tr>
</tbody>
</table>
education services. In other words, all policies that adopt the form of government procurement will be included in the research objects. The policy of GPPS in the current research refers to the policy texts related to government procurement in the form of laws, regulations, department rules, implementation measures, and so on.

The implementation of GPPS in China started in 1995 in Shanghai. The Shanghai Pudong New Area Social Development Bureau commissioned the Shanghai YMCA to manage the Pudong New Area Luoshan Citizens Leisure Club and undertake community public services. This is the earliest practice to explore GPPS in China. Therefore, the collection of policy texts on GPPS began in 1995. In order to exhaust the relevant policies of GPPS in China and to ensure that the policy data are comprehensive and true, the current research takes the official websites of local governments and the China Government Procurement Network as the main data sources. What is more, the software of PKULAW is also an important tool, which is a legal inquiry software. GPPS policies are collected through retrieving the keywords such as government procurement of public services, government procurement and service outsourcing in the relevant websites, and the software. It is noteworthy that when retrieving the relevant policies, “full text” search rules should be applied, which refers to search the keywords related to government procurement in the full text of the policy, so as to cover the comprehensive policies. In order to eliminate the interference of useless data, the whole policy text with the keyword cumulative hit frequency greater than 10 will be included in the research sample. If the keyword hit frequency is less than 10, only the paragraphs containing keywords are included in the policy text as a data sample. This method can effectively eliminate irrelevant information and improve the pertinence and accuracy of text mining.

Finally, 112 policies related to GPPS are collected. Authors exclude the weakly relevant and invalid texts such as website reports. Finally, a total of 86 policy texts of GPPS are selected as the sample data, which includes not only the general guidance of GPPS but also covers the policy tools such as government procurement of pension services.

3.2.2. Data Overview. Text preprocessing is the first step of text mining. The current research applies the software of Python and the third-party library of `jieba` to conduct the word segmentation and word frequency statistics on 86 policy texts. Before word segmentation, stop words must be removed and the custom dictionary should be loaded. The paper applies the existing stop words thesaurus of Harbin Institute of technology to remove the meaningless words. What is more, to improve the accuracy of text mining and eliminate the interference of commonly used words in GPPS, the current research includes the words of “Government,” “Department,” “Country,” “Implementation,” etc. into the stop words thesaurus. And a professional dictionary of GPPS is constructed by customizing and using Sogou cell Thesaurus. On this basis, the policy texts are segmented with the precise mode under the library of `jieba`.

3.3. Stage Division. Representative GPPS policies are taken as the key nodes, and GPPS policies in China from 1995 to 2021 are divided into three stages. The first phase is from 1995 to 2003. The second phase is from 2004 to 2012. And the third phase is from 2013 to present.

The division basis is as follows: (1) From 1995 to 2003, the policies related to GPPS focused on the government purchase system. “Government Purchase Law of People’s Republic of China” promulgated in 2003 provided a legal basis for governments to purchase. Although the policy issued in 2003 only attaches great importance on goods and projects, neglecting the public services, this policy is also an important turning point in GPPS in China. (2) After the promulgation of “Government Purchase Law of People’s Republic of China,” local governments have actively explored public service outsourcing and many local regulations and documents on GPPS have emerged, such as “Guidance of Wuxi Municipal Government Purchase of Public Services (for Trial Implementation)” in 2005 and “Implementation Opinions of Xinxiang Municipal Government Office on the Purchase of Public Services” in 2008. (3) In 2013, the document of “Decision of the Central Committee of the Communist Party of China on Comprehensively Deepening Reform” stressed that governments should promote GPPS and all transactional management services should be outsourced in principle. Issuance of this document marks the strong support of the state for the policy of GPPS. Under the guidance of the national documents, the central and local governments actively issue
policies about GPPS. For example, “Administrative Measures for Government Procurement of Public Services (Provisional)” was implemented on January 1, 2015. “Administrative Measures for Government Procurement of Public Services” was implemented on March 1, 2020. Compared with previous documents, the policies at this stage have detailed and thorough regulations about GPPS in terms of connotation definition, purchase scope, purchase subject, etc., which signifies that the policy of GPPS has been elevated to a higher position. Figure 2 shows the stage division of the GPPS policy.

4. Static Stage Characteristics of GPPS

4.1. Keywords Acquisition. After having the basic grasp of 86 policy texts and dividing the policy stages, the theme analysis of the policy texts under each phase is carried out through the TF-IDF algorithm. Compared with word frequency statistics, the TF-IDF algorithm can better reflect the distinguishing ability of the words.

4.1.1. Embryonic Stage (1995–2003). China started the practice of GPPS in 1995. GPPS was not incorporated into the national policy until the promulgation of “Government Purchase Law of People’s Republic of China” in 2003. However, at this stage, the national policy documents, including "Government Purchase Law of People’s Republic of China" focus on government purchasing goods or projects and only mention the concept of GPPS. The connotation, purchase subject, purchase object, and purchase scope of GPPS are all not clear at this stage. Policy texts collected at this stage (1995–2003) are taken as the research object, and the library of jieba.analyse is imported from Python. After filtering out some useless words, the TF-IDF algorithm is applied, which carries out the word segmentation and calculates the TF-IDF values of words simultaneously. The words sorted by the TF-IDF value at this stage are shown in Table 2.

With the TF-IDF method, the keywords of policy texts during 1995 to 2003 are extracted. Results show that TF-IDF value of “Purchase” entry is the largest and it is significantly higher than the TF-IDF value of other entries. Keywords such as “Bidding,” “Cargo,” “Projects,” and “Engineering” also indicate that the policies during this stage focus on government purchasing cargo or engineering, which have not been refined to the field of GPPS. The conceptual framework of government purchase includes three objects, which are cargo, engineering, and services, so GPPS is an integral part of government purchase. At this stage, although the policies or regulations related to GPPS are still imperfect, local governments attempt to outsource public services and explore GPPS. For example, in 1998, the Shanghai Pudong New District government adopted a noncompetitive purchase way and commissioned the elderly-care services to the Shanghai YMCA. In 1999, Shenzhen Luohu District also outsourced the cleaning work to social organizations. The current research named this stage as the embryonic stage. During the embryonic stage, the practice of GPPS is still immature and the policy of GPPS is imperfect, which restricts the development of GPPS.

4.1.2. Exploration Stage (2004–2012). Promulgation of “Government Purchase Law of People’s Republic of China” in 2003 makes GPPS in China have laws to abide by. Therefore, GPPS has entered a new era since 2003. Local governments have actively explored the practice of GPPS. Policy texts collected at this stage are imported to the library of jieba.analyse. Similarly, after removing the stop words, the command of jieba.analyse.extract_tags is performed, so as to carry out the word segmentation and calculate the value of TF-IDF. Thus, the top ten keywords at this stage and corresponding TF-IDF values are shown in Table 3.

At this stage, GPPS policies have sprung up everywhere and local governments actively explore the practice of outsourcing. It can be seen from Table 3 that the TF-IDF values of “Government procurement” and “Public services” are at the top of the ranking list. Compared with the keywords at the embryonic stage, the policies at this stage are targeted at GPPS, rather than be limited to government purchase cargo or engineering. In addition, keywords such as “Healthcare” and “Community service” not only reflect the intensity of market-oriented reform in the fields of medical treatment and community but also illustrate the highlight of purchase content. Keywords like “Multiform” and “Project” are the descriptions of purchase ways, indicating that local governments are encouraged to adopt multiple purchase ways and the project system is strongly supported. Keywords such as “Social organizations” and “Government-affiliated institutions” are the definition of service undertaker. It can be distinguished that the producers and providers of services are mainly nonprofit organizations at this stage. Finally, among the top ten keywords at this stage, there is only one verb, namely, “Guide,” indicating that the country is guiding the development of GPPS at this stage.

This phase is named as the exploration stage. During this stage, all localities actively explore the practice of GPPS and the relevant policies are constantly improving. It can be seen from the results of keyword extraction that policy texts at this stage focus on purchase content, purchase ways, service undertakers, etc., which belongs to the investigation and mining of the GPPS policy itself.

4.1.3. Comprehensive Development Stage (2013–2021). The document of “Decision of the Central Committee of the Communist Party of China on Comprehensively Deepening Reform” issued in 2013 stressed that all transactional management services should be outsourced to the markets through contract, entrustment, or other means. GPPS has gradually become the basic strategy of constructing a service-oriented government. Since then, the number of GPPS policies and regulations has been rising. Policy texts of GPPS at this stage are excavated, and the top ten entries of the TF-IDF value are shown in Table 4.

At this stage, “Government procurement” and ”Public services” are still the most critical terms. Significantly,
“Elderly-care” has become the entry with the highest TF-IDF value except for “Government procurement” and “Public services.” This result reflects the government’s attention to the aging problem and its determination to actively respond to it. According to the results of the 7th National Census in China, the ratio of population of 65 and above is 13.5% in 2020. China is going to enter a deeply aging society, and the problems of the population structure are becoming increasingly prominent. Under this background, it is an inevitable choice for governments to purchase elderly-care services and attract multiple subjects to participate in the supply of elderly-care services. In addition, healthcare is still the key purchase content in GPPS. As one of the keywords, “Culture” reflects the growing spiritual needs of people. Keyword “Supervise” appears for the first time and has a high ranking at this stage, indicating that policymakers have realized the importance of supervision. A sound supervision mechanism is the basis for improving the implementation effect of GPPS. As the definition of service undertaker, the keyword “Social forces” at this stage naturally correspond to the keyword “Social organizations” and “Government-affiliated institutions” at the exploration stage. Compared with “Social organizations” and “Government-affiliated institutions”, “Social forces” has a broader scope, including profit-making organizations and nonprofit organizations. This change reflects the government’s greater tolerance in service undertakers and the support for GPPS. “Directory,” as the top ten keywords of TF-IDF, is actually the embodiment of purchase scope. Only by clarifying the purchase scope can GPPS be effectively promoted. Finally, the keyword “Capital” reflects the importance of fund use and budget management.

This phase is named as the comprehensive development stage. Compared with the exploration stage, the policy texts at this stage are no longer confined to purchase content, purchase object, and purchase way. Policies at the comprehensive development stage pay more attention to solve the problems of effective procurement.

4.2. Construction of the Complex Network. Based on the extracted keywords at each stage, the similarity among keywords can be calculated by applying the word2vec algorithm. The following steps are required to implement the word2vec algorithm in the library of genism.

(1) Step 1: original corpus should be preprocessed and conducted the word segmentation.
(2) Step 2: instantiate the model through the package of word2vec.
(3) Step 3: train the instantiated model.
(4) Step 4: similarity between any two terms can be calculated through applying the function of wv.similarity.

<table>
<thead>
<tr>
<th>Keywords</th>
<th>Purchase</th>
<th>Supplier</th>
<th>Bidding</th>
<th>Cargo</th>
<th>Projects</th>
<th>Agent</th>
<th>Institution</th>
<th>Tender</th>
<th>Engineering</th>
<th>Supervise</th>
</tr>
</thead>
<tbody>
<tr>
<td>TF-IDF value</td>
<td>0.623</td>
<td>0.542</td>
<td>0.512</td>
<td>0.506</td>
<td>0.432</td>
<td>0.421</td>
<td>0.412</td>
<td>0.322</td>
<td>0.309</td>
<td>0.271</td>
</tr>
</tbody>
</table>
Table 3: The TF-IDF value of words in policies from 2004 to 2012.

<table>
<thead>
<tr>
<th>Keywords</th>
<th>Government procurement</th>
<th>Public service</th>
<th>Healthcare</th>
<th>Multiform</th>
<th>Social organizations</th>
<th>Guide</th>
<th>Community service</th>
<th>Government-affiliated institutions</th>
<th>Project</th>
<th>Marketization</th>
</tr>
</thead>
<tbody>
<tr>
<td>TF-IDF</td>
<td>0.4912</td>
<td>0.436</td>
<td>0.426</td>
<td>0.413</td>
<td>0.406</td>
<td>0.387</td>
<td>0.365</td>
<td>0.341</td>
<td>0.332</td>
<td>0.312</td>
</tr>
</tbody>
</table>
Table 4: The TF-IDF value of words in policies from 2013 to 2021.

<table>
<thead>
<tr>
<th>Keywords</th>
<th>Government procurement</th>
<th>Public service</th>
<th>Elderly-care</th>
<th>Supervise</th>
<th>Encourage</th>
<th>Healthcare</th>
<th>Culture</th>
<th>Social forces</th>
<th>Directory</th>
<th>Capital</th>
</tr>
</thead>
<tbody>
<tr>
<td>TF-IDF</td>
<td>0.723</td>
<td>0.711</td>
<td>0.652</td>
<td>0.621</td>
<td>0.576</td>
<td>0.554</td>
<td>0.465</td>
<td>0.452</td>
<td>0.448</td>
<td>0.397</td>
</tr>
</tbody>
</table>
Significantly, the word2vec algorithm is based on the word segmentation results of original corpus, that is, the policy texts of GPPS. At the same time, the keywords obtained by the TF-IDF algorithm must exist in the word segmentation results. Therefore, according to the above steps, the similarity among the keywords at each stage can be achieved. Corresponding edges are established between two nodes with similarity greater than 0.8. Thus, the complex network of keywords at each stage can be constructed with the software of Gephi, which is shown in Figures 3–5. In the process of constructing complex networks through Gephi, the betweenness centrality of nodes is used as the indicator to measure the importance of nodes. Betweenness centrality means that if the shortest path between any nodes passes through the same node, then the node is important. In Figures 3–5, the greater the betweenness centrality of the node, the larger the node label and the darker the color of the node. With the complex network of keywords, the correlation among keywords at each stage can be observed and the stage characteristics of policies can be further portrayed.

Figure 6 shows the complex network of keywords at the embryonic stage. Keywords such as “Purchase,” “Cargo,” “Supplier,” “Bidding,” and “Tender” have large labels and dark color, indicating that these keywords’ value of betweenness centrality is great. The complex network at the embryonic stage further verifies that the policies at this stage focus on government purchase, such as government purchasing cargos or engineering, without refining to GPPS. According to Figure 3, the policy texts at the embryonic stage present the following characteristics: (1) the concentration is government purchase, rather than GPPS; (2) the purchase objects are mainly cargo and engineering; and (3) the importance of supervision has been paid attention to.

Figure 3 shows the complex network of keywords at the exploration stage. At this stage, keywords such as “Government procurement,” “Public services,” “Marketization,” and “Social organizations” have large labels and dark colors, indicating that policy texts at the exploration stage attach importance to GPPS. According to Figure 3, the characteristics of policy texts at the exploration stage can be summarized as follows: (1) The focus is GPPS, realizing the leap from government purchase cargo or engineering to GPPS. (2) Keywords such as “Multiform” and “Project” represent the purchase mode. “Community service” and “Healthcare” represent purchase content. Keywords such as “Government-affiliated institutions” and “Social organizations” represent purchase objects. Policy texts at this stage emphasize the basic operation of GPPS, concerning the policy itself. (3) The country is guiding the development of GPPS.

Figure 4 shows the complex network of keywords at the comprehensive development stage. It can be learned from Figure 5 that keywords “Social forces,” “Supervise,” and “Elderly-care” have large labels and dark color except for “Government procurement” and “Public services.” According to Figure 4, static stage characteristics at the comprehensive development stage can be summarized as follows: (1) The policy theme is still GPPS, the same as the second stage. (2) It can be seen from keywords such as “Supervise,” “Capital,” and “Directory” that, policymakers have paid attention to the effect and efficiency of the GPPS policy. Policymakers attempt to improve the implementation effect of GPPS by means of supervision, fund management, and purchase directory setting, realizing the transformation from “what to buy,” “how to buy,” and “whom to buy from” at the second stage to “implementation effect” at the comprehensive development stage. (3) Purchase content at this stage mainly includes elderly-care, culture, and healthcare. The keyword “Elderly-care” indicates that the problem of aging population in China is becoming increasingly urgent. The keyword “Culture” shows that people demand for the plentiful cultural activities with the development of economy. As the same keyword as the secondary stage, the keyword “Healthcare” explains the public’s long-term demand for healthcare. (4) The country is encouraging the implementation of GPPS.

5. Dynamic Evolution Logic of GPPS

Keyword extraction and the complex network of keywords at each stage help to clarify the static stage characteristics of
GPPS. Collected policies of GPPS have a rich source, including not only central regulations but also local documents. What is more, when setting the retrieval criteria, both accurate search and the fuzzy search are implemented. Therefore, the collected policy texts are complicated. This chapter attempts to cluster the policy texts by K-means clustering analysis, so as to classify the complicated policies and observe the dynamic evolution logic of the GPPS policy.

5.1. Policy Clustering Based on K-Means. Text clustering based on K-means is an unsupervised machine learning method that does not need to train and labeling the texts in advance. The method classifies the documents based on the value of text similarity. The main steps of clustering the GPPS policies based on K-means are as follows.

5.1.1. Vectorization of Document Information. The basis of K-means clustering analysis is still the word segmentation and word frequency statistics. However, the result of word frequency statistics cannot be modeled in K-means clustering analysis directly. Further vectorization of policy texts is needed, that is, the word frequency matrix of the policy is supposed to be constructed. Specific operations of vectorization are as follows: Firstly, the package of CountVectorizer should be imported from the sklearn library, and then instantiation is carried out. Secondly, to realize the vectorization of policy texts, the command of count.fit_transform is applied to the policy texts which have been performed with word segmentation, so as to obtain the word frequency matrix.

5.1.2. Determination of Clustering Category Number k. In this paper, the elbow method is applied to determine the optimal category number k. The core index of the elbow method is the sum of squares of errors (SSE). The basic principle of the elbow method denotes that when the value of k is smaller than the real cluster number, the value of SSE will
drop greatly because the increase in $k$ will greatly amplify the
aggregation of each cluster. When the value of $k$ is close to
the real cluster number, the decrease degree of SSE will
narrow sharply and the value of SSE tends to be flat with the
continuous increase in $k$. Therefore, the graph of the rela-
tionship between SSE and $k$ is the shape of an elbow, and
the value $k$ that corresponds to the position of the elbow is
the final clustering category number $k$ [43]. The calculation
formula of SSE is as follows:

$$\text{SSE} = \sum_{i=1}^{k} \sum_{p \in C_i} |p - m_i|^2,$$

(1)

where $C_i$ indicates the $i$th cluster, $p$ is the sample point of $C_i$,
$m_i$ is the centroid of $C_i$, SSE is the clustering error of all
samples, which represents the clustering effect. With the aid
of Python software and the application of the elbow method,
it is found that the SSE curve has an obvious turning point
when $k = 7$. Therefore, the collected policies of GPPS are
divided into 7 categories.

5.2. Results of K-Means Clustering. The package of K-means
is imported from the library of sklearn.cluster. Then, based
on the word frequency matrix of the policy that is obtained
in the chapter of 5.1.1, model instantiation is carried out
where the parameter of n_clusters is set as 7. The interface of
fit is used to implement the clustering analysis on the word
frequency matrix. According to clustering analysis results,
the number of policies corresponding to 7 clusters is 32, 15,
18, 7, 5, 5, and 4, respectively.

In order to make the clustering results clearer and
further explore the internal evolution law of GPPS policies,
the current research carries out the following operations: (1)
Firstly, the policy texts are classified into 7 documents
according to their corresponding categories. (2) Secondly,
the merged 7 policy documents carried out the word seg-
mentation again. Based on the word segmentation results,
word frequency of policy texts under each category is cal-
culated and word cloud of each category is drawn.

Cluster1 contains the keywords such as “Government
procurement,” “Public services,” “Directory,” and “Social
forces.” Compared with other clusters, keywords in Cluster1
stress the GPPS policy itself and have the characteristic of
guiding. Therefore, this category is summarized and named
as “General rules.” General policies aim to deploy the
implementation of GPPS by providing top-level design such as
the basic principles, purchase subjects, purchase contents,
purposes, and tasks. In Cluster2, “Government proc-
urement” is still the most prominent. In addition, keywords
such as “Budget,” “Capital,” “Financial capacity,” and
“Expenditure based on events” show that policies in this
category are mainly related to budget and fund management.
Therefore, Cluster2 is named as “Budget management.”

Policies in Cluster2 aim to improve the use efficiency of
financial funds in GPPS. Keywords such as “Performance,”
“Evaluation,” “Third-party,” and “Quality” are in Cluster3,
which indicate that the policies in this category aim to
improve the quality of GPPS and pay attention to perfor-
manance evaluation. Thus, Cluster3 is named as “Performance
evaluation.” In Cluster4, there are keywords such as “In-
formation,” “Publicity,” “Transparency,” “Disclosure,” and
“Social supervision.” Policies in Cluster4 are intended to
ensure the openness and transparency of information in
GPPS. The work mechanism and the operation process of
information disclosure in GPPS should be improved, so as to
achieve the fair and just implementation of GPPS. Therefore,
Cluster4 is named as “Information disclosure.” In addition
to “Government procurement,” keywords such as “Health,”
“Public hygiene,” and “Medical” are particularly striking in
Cluster5. This category is mainly related to the government
purchasing medical health. Thus, Cluster5 is named as
“Healthcare.” Keywords such as “Elderly-care,” “Pension,”
and “Aging” in Cluster6 indicate that this category em-
phasizes government purchasing elderly-care service.
Cluster6 is designated as Elderly-care. Keywords such as
“Culture,” “Sports,” “Sports facilities,” and “Library” are in
the Cluster7, which attaches importance to culture and
sports. Thus, Cluster7 is named as “Culture & Sports.”

5.3. Dynamic Evolution of Different Clusters. Through K-
means clustering analysis, policies of GPPS are classified to
7 categories. In order to further explore the dynamic evolution
logic of the GPPS policy, the current research selects the top 20
keywords in terms of word frequency under each cluster. Then,
word frequency of selected keywords in each cluster is counted
under three stages, respectively, so as to observe the changes
of 7 clusters of the GPPS policy at three stages. Cluster3, which is
the performance evaluation policy, is taken as an example to
illustrate the operations. Specific steps are as follows: Step 1:
Top 20 keywords in terms of word frequency are selected in
Cluster3. Step 2: Policies at the embryonic stage, the explo-
ration stage, and the comprehensive development stage are
taken as the original corpus, respectively. Command fdist
under NLTK in Python is applied to calculate the word fre-
quency of the selected 20 keywords at each stage. Step 3: The
total word frequency of 20 keywords is calculated at each stage.
According to the above steps, total word frequency of keywords
selected in each cluster can be calculated at three stages. Finally,
the calculated results are normalized by the maximum and
minimum method.

Governments usually achieve one policy objectives by
combining different policy tools [44], which is suitable in the
policy of GPPS. 7 clusters of GPPS policies can be regarded
as different policy tools to provide guarantee for the
implementation of GPPS. Dynamic change logic of 7 policy
tools at three stages is shown in Figure 5.

It can be seen from Figure 5 that the word frequency of
keywords in 7 clusters has shown an upward trend from the
embryonic stage to the comprehensive development stage,
indicating that the number of 7 policy tools have been in-
creasing from 1995 to 2021. Generally speaking, the GPPS
policy has been continuously improved and standardized.
Meanwhile, 7 policy tools have different evolution logic at
three stages.

(1) The number of policies in Cluser1, which are mainly
general rules of GPPS, is increasing steadily from the
embryonic stage to the comprehensive development
stage. And the policies in Cluster1 maintain a relatively high number at three stages, indicating that the general rules of GPPS have always been valued and paid attention to. General rules play a vital role in GPPS policies. Through clarifying the policy purposes, purchase subjects, purchase contents, and so on, general rules standardize the implementation of GPPS. General rules are rich in content and have strong applicability to different policy tools. Therefore, general rules of GPPS are indispensable at all stages. For example, the policy “Measures for the Administration of Government Purchase of Services” issued in 2020 belongs to the general rules. The policy redefines the purchase subjects and undertakers of GPPS, and the purchase directory is updated. To sum up, general rules provide the basis for the development of GPPS.

(2) For healthcare, elderly-care, and culture & sports in Cluster5, Cluster6, and Cluster7, they are specific policy tools of GPPS in different fields. It can be learned from Figure 5 that the number of three policy tools presents a steady increasing trend at three stages. Among three policy tools, policies about government purchasing healthcare services have the largest number at the initial embryonic stage and the number remains relatively high at three stages. The evolution logic of Cluster5 indicates that the demand for healthcare in China has been persistently urgent. Governments increase the supply of healthcare service through implementing the policy of GPPS. Furthermore, the number of elderly-care and culture & sports policy tools in GPPS is low at the initial stage. Until recent years, the number of relevant policies about government purchasing elderly-care services and government purchasing culture & sports services has been raising. The change in policy numbers is closely associated with the aggravation of aging problems and the improvement of people’s spiritual and cultural needs in China. China entered the aging society in 1999. At the embryonic stage, governments explored the practice of purchasing elderly-care service, but there are still few relevant policy tools. In recent years, the problem of aging in China has become increasingly serious. In 2019, the growth rate of aging is the highest in the world and China is going to enter a deep aging era. Under the background, Chinese governments attach great importance to the issue of aging and actively confront the aging problem through promulgating the policy of GPPS. Similarly, at the embryonic period, economic development in China was slow. With the help of Maslow’s hierarchy of needs theory, people paid more attention to medical and health care which is more closely related to their own survival during this stage. With the rapid growth of China’s economy and the improvement of the living standards, public culture and sports have become the demand focus. Therefore, governments increase the supply of culture & sports by means of purchase.

(3) Cluster2, Cluster3, and Cluster4, which are policy tools of budget management, performance evaluation, information disclosure in GPPS, present the similar evolution logic at three stages. Three policy tools are necessary supporting systems for the implementation of GPPS. At the embryonic stage, policy texts about GPPS are mainly related to the basic content of the policy such as purchase content, purchase objective, and purchase subject, and there are few policies about supporting policies. However, with deepening of the reforms and development of the GPPS, new problems have emerged. For example, governments neglect the importance of performance evaluation, resulting in the low-quality level of public services purchased by governments. Thus, to provide the guarantee for the high-quality development of GPPS, number of policies about budget management, performance evaluation, and information disclosure surged during the exploration stage and the comprehensive development stage.

6. Conclusions

Exploring the static stage characteristics and dynamic evolution logic of GPPS can provide references for improving the GPPS policy. Based on the text-mining technology, the current research conducts keywords extraction, complex network construction, and cluster analysis on 86 GPPS policies from 1995 to 2021. The following conclusions are obtained:

(1) The TF-IDF algorithm is applied to extract keywords of GPPS policies at each stage, and then based on the Word2vec algorithm, complex networks of keywords are constructed, so as to summarize the stage characteristics of GPPS policies. Results show that the development of the GPPS policy can be divided into the embryonic stage, the exploration stage, and the comprehensive development stage. Policy characteristics at each stage are also obtained. The GPPS policy in China presents a point-to-face change feature, manifested in the evolution logic of “government purchase—government procurement of public service—all-round supporting policy.” At the embryonic stage, GPPS lacks targeted and detailed legal basis. There are only documents aimed at government purchasing cargo or engineering. During the exploration stage, the GPPS policy has developed rapidly, no longer focusing on government purchasing cargo or engineering. Policies at this stage pay attention to the basic processes and subjects involved in GPPS, stipulating the purchase content, purchase ways, purchase objects, and so on. At the comprehensive development stage, the GPPS policy still focuses on the top-level design for the policy itself. In addition, policy texts at this stage begin to involve the supporting regulations to improve the quality of GPPS, such as budget
management policies, information disclosure policies, and performance evaluation policies. In other words, policies during this stage are more detailed and refined.

(2) Based on K-means cluster analysis, clustering results are compared with secondary word segmentation and word frequency statistics, so as to figure out the dynamic evolution logic of the GPPS policy. Results show that policy priorities at different stages, namely, policy tools, will change according to the development of economy and variation of demands. Specifically, (1) general rules are of great importance in the development of GPPS and have maintained a stable development at all stages. (2) The development of policy tools which are oriented to specific service areas fluctuates with the variation of the public’s demand at different stages. For example, with the improvement of economy, people’s demand for public culture and sports is swelling. To increase the supply of culture & sports services, government purchasing culture & sports services has become an important policy tool. Moreover, government purchasing elderly-care service is also a vital measure to deal with the problem of aging. (3) The number of supporting policies in GPPS presents a gradual upward trend, indicating that the focus of GPPS has shifted from “what to buy,” “how to buy,” and “whom to buy from” to “how to improve the quality of GPPS”.

The current work is applicable to other regions, and the proposed method also can provide the reference to the studies on policy evolution in other countries. Different from the previous research, this paper combines the static stage characteristics analysis with dynamic evolution analysis through adopting text-mining technology, so as to figure out the evolution logic of policy. Thus, when we aim to investigate the policy evolution in other regions, the TF-IDF algorithm is firstly applied to obtain the keywords of policy and the static stage characteristic can be summarized by constructing the complex network of keywords. Then, K-means clustering analysis is adopted to summarize the dynamic evolution logic of the policy. In summary, the proposed method and research logic are applicable for the policy evolution research in other regions.

Data Availability
The datasets generated during the current study are available from the corresponding author on reasonable request.

Conflicts of Interest
The authors declare that there are no conflicts of interest.

Authors’ Contributions
All authors contributed to the study conception and design. Material preparation, data collection, and analysis were performed by Yuting Zhang, Lan Xu, and Zhengnan Lu. The first draft of the manuscript was written by Yuting Zhang, and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Acknowledgments
This study was funded by the National Nature Science Foundation of China under grant no. 72243005, the China Scholarship Council (202208320284) and the Jiangsu Provincial Department of Education Fund of Philosophy and Social Science (2017SJB2176).

References


