Research Article

The Temporal Spatial Dynamic of Land Policy in China: Evidence from Policy Analysis Based on Machine Learning

Xiao Li,1 Yuting Yao,2 and Meihong Zhu1

1Zhejiang University of Water Resources and Electric Power, Zhejiang, Hangzhou 310018, China
2School of Public Affairs, Zhejiang University, Zhejiang, Hangzhou 310018, China

Correspondence should be addressed to Meihong Zhu; mhzhu200831@126.com

Received 9 July 2022; Revised 15 September 2022; Accepted 11 November 2022; Published 23 November 2022

Copyright © 2022 Xiao Li 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.

Extracting useful information from a large number of policy texts is a challenging and insufficiently discussed topic. Utilizing large sample policy texts and a method of machine learning, this study contributes to the research gap by systematically analyzing the temporal evolution and spatial differentiation of China’s land policy from 1998 to 2018. A framework comprising six major themes of land policy, namely, “land development, land acquisition and demolition, cultivated land protection, land planning, land consolidation and utilization, and land confirmation and transfer” is first established, according to the theoretical and institutional background of land management. Based on this framework, the Latent Dirichlet Allocation analysis of more than 20,000 policy documents at different levels of government shows that, (1) temporally, the priority of land policy evolves with the spirit of the central document and the macroeconomic and social conditions and, (2) spatially, there are significant differences in land policies among provinces. Overall, the analysis of land policy documents shows the tradeoff between cultivated land protection and land development and also the emphasize on other topics, with the changes in land policy priorities in different periods and regions.

1. Introduction

Over the past 40 years of reform and opening up, China has witnessed rapid economic growth and urbanization. According to the National Bureau of Statistics of China, the urbanization rate increased from 17.92% in 1978 to 60.60% in 2019. China’s unique land system has played a crucial role in its industrialization and urbanization. For example, George and Samuel pointed out that the liberalization of land rights initiated industrialization from 1981 to 1994, and land-based development has promoted urbanization [1]. Wang and Lu argued that the “land expropriation-land transaction” model determined by the dual urban-rural land structure has been an important impetus for China’s rapid urbanization [2].

China’s land policy has undergone continuous evolution since the Third Plenary Session of the 11th Central Committee of the Communist Party of China (CPC), which initiated the household responsibility system (HRS). For example, the Constitution of 1982 stipulated that “no organization or individual may occupy, sell, lease, or illegally transfer land in any forms.” However, the first public auction of land in Shenzhen in 1987 prompted an amendment to the 1988 Constitution to allow the transfer of land use rights. On the contrary, land development in the 1990s led to a sharp decline in arable land, and in 1998, the Land Administration Law was amended to emphasize the protection of arable land. A recent case is the latest Land Administration Law (2020), which emphasizes the reform of the land expropriation system and the protection of farmers’ rights and interests. Those cases imply that land policies have been changed over time. However, what is the evolution trend of land policy in China? What is the driving force for policy evolution? What is the future of China’s land policy?

Existing studies have made some attempts at these questions, mainly by the analysis of land policies. For example, Wang et al. used 192 central-level policy texts on idle land management from 1992 to 2015 to conduct an
econometric analysis and explored the policy tools used by the central government in managing idle land issues [3]. Lv et al. used 59 relevant policies as samples to analyze the evolution process of collective construction land transfer policies [4]. Those studies help to understand the evolution logic of China’s land policy, yet there are also limitations. One of the major limitations is that those studies mainly concentrates on a limited number of policy texts of the central government or local governments of certain region, which limits the application of analysis results and may lead to the problem of selection bias. Therefore, a larger number of policy texts with a wider coverage are needed for the analysis of land policies.

Based on this requirement, we collected all land-related policy documents issued by all levels of government in China from 1998 to 2018 (N = 22,659) and analyzed them with the Latent Dirichlet Allocation (LDA) model. With the policy analysis, this study explored the evolution process of China’s land policy over the past two decades, including the changes in policy priorities and the spatial differentiation of land policies in China.

The remainder of this study is organized as follows. Section 2 introduces the institutional background and lays the foundation for the topic classification in the following sections. Section 3 introduces the data and study methods used in this study. Section 4 presents the machine learning results based on the theoretical and institutional background. Section 5 further analyzes the characteristics of the temporal and spatial evolution of China’s land policy. Section 6 presents the conclusion and discussion. The abbreviations in this study are all listed in Table 1.

2. Theoretical and Institutional Background

2.1. Literature Review. Studies on government policies are important for their theoretical significance as well as their practical meaning [5, 6]. For example, based on studies on environmental protection policies, Bao et al. find that the prices of battery electric vehicles are more dependent on government subsidy than fuel vehicles [7]. It is also approved that government subsidies and carbon emission reduction have an effective influence on the development of electric vehicles [8–10]. However, policies of government subsidies can also have a negative effect, for instance, on the supply chain system’s stability [11]. Furthermore, policies are found to have a significant impact on firms’ production behavior [12, 13], economic returns [14], and even the fluctuation of international markets [15]. These empirical results all suggest the importance to extract information from policy text and to evaluate the effects of policy.

Therefore, it is essential to study land policy texts, to understand the evolution of China’s land policy and land system. However, due to the multilevel administrative system in China, the number of policy documents issued by the multilevel governments is very large, making it impossible to manually conduct textual analysis. Thus, previous studies on the analysis of policy texts have mainly focused on some important policy documents. For example, Wang et al. used 192 central-level policy texts to explore the policy tools for managing idle land issues [3]. Lv et al. used 59 relevant policies to analyze the evolution process of collective construction land transfer policies [4]. The manual screening of a limited number of policy documents may lead to the problem of “seeing trees but not the forest” or falling into the trap of selection bias. Moreover, previous studies mainly focus on the policy documents issued by the central government, with little notice of the spatial differences in policy priorities in different regions. Therefore, it is necessary to include a larger number of policy texts with wider coverage, which calls for more advanced text analysis methods.

The core challenge of text analysis is how to extract the required information accurately and classify and analyze them efficiently. Machine learning is widely used for the purpose of information extraction, classification, and analysis. For example, modified region growing algorithm and adaptive genetic fuzzy classifier are used in the process of noise removal, segmentation, feature extraction, and recognition, which are aimed to extract and recognize sign gesture language to facilitate gesture-based communication [16, 17]; to enhance communication efficiency, a federated communication framework named TKAGFL is proposed to deal with the problems of updates’ strategy and data heterogeneity, which is expected to benefit the application of federated learning in the industry [18]. Drawing on these studies, we use the method of machine learning to analyze land policies and to sum up the evolution logic of the land system in China.

2.2. Institutional Background of Land System in China. Under the dual land ownership system, the land system in China is an important part of linking urban and rural development and a necessary basis for coordinating urbanization, industrialization, and agricultural stability [19, 20]. Over the past 20 years, land use and development, especially the large-scale transformation from agricultural land to nonagricultural land, have provided the necessary conditions for rapid economic and social development [21]. At the same time, conflicts, such as the decline in arable land, mismatches in land supply structure, and land requisition conflicts, have prompted a series of changes in the land system.

The year 1998 was a milestone of land system reform in China. The second revision of the Land Administration Law basically laid the foundation for the current land system in

Table 1: Expression and abbreviation.

<table>
<thead>
<tr>
<th>Expression</th>
<th>Abbreviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Latent Dirichlet allocation</td>
<td>LDA</td>
</tr>
<tr>
<td>Natural language processing</td>
<td>NLP</td>
</tr>
<tr>
<td>Support vector machines</td>
<td>SVMs</td>
</tr>
<tr>
<td>K-nearest neighbor</td>
<td>KNN</td>
</tr>
<tr>
<td>Communist party of China</td>
<td>CPC</td>
</tr>
<tr>
<td>Household responsibility system</td>
<td>HRS</td>
</tr>
<tr>
<td>Optimal separating hyperplane</td>
<td>OSH</td>
</tr>
<tr>
<td>Empirical risk minimization</td>
<td>ERM</td>
</tr>
<tr>
<td>Term frequency inverse document frequency</td>
<td>TF-IDF</td>
</tr>
</tbody>
</table>
China, which is also the main reason why 1998 was used as the starting point for the analysis in this study. The revised *Land Administration Law* stipulates that land expropriation is the only way to convert agricultural land to nonagricultural use. In the same year, the State Council issued the *Notice on Further Deepening the Urban Housing System Reform and Accelerating Housing Construction*, which marks the beginning of the monetization of housing allocation. Since then, the system of paid use of state-owned land has been continuously strengthened. For example, in 1999, the *Ministry of Land and Resources issued the Several Opinions on Regulating the Lease of State-owned Land*, which allows the direct lease of state-owned land use rights; then, in 2003, the system of bidding, auction, and listing of commercial began to be implemented. These land systems are important tools for local governments to use land resources to promote development and generate revenue [22]. In the late 1990s, industrial parks developed rapidly, and the demand for land greatly increased. Following the tax-sharing reform of China in 1994, local governments have largely lost control of local taxes and become increasingly dependent on land revenue [23]. These economic and political backgrounds have been integrated with the land system to form a land urbanization development mode with China’s characteristics.

Land urbanization is largely depended on the large-scale transformation of agricultural land to nonagricultural land. Correspondingly, land expropriation and land development have become important topics of land policy. For example, the Ministry of Land and Resources issued the *Notice on Some Opinions on Land Development, Rehabilitation, and Measures for the Administration of the Examination and Submission for Approval of Construction Land* in 1999; the State Council issued *The Decision of the State Council on Furthering the Reform and Intensifying the Land Administration* and *The Guiding Opinions on Improving the Land Expropriation Compensation and Resettlement System* in 2004. These documents provide detailed provisions on land development, land expropriation procedures, compensation, and resettlement.

However, land urbanization has also led to a contradiction between “preserving rice bowls” and “promoting development.” Due to food security concerns, the *Land Administration Law* in 1998 established the basic national policy for the protection of arable land in the form of legislation for the first time. The first policy statement released by the central government in 2004 clearly requires governments at all levels to implement the strictest protection of arable land. According to the *2007 Government Work Report*, 1.8 billion mu of arable land is an insurmountable red line in China. The protection of arable land has always been the top priority of land policy [24].

Land-driven industrialization and urbanization have brought about rapid economic development. However, problems such as overexploitation, inefficient land use, and irrational structures have emerged in this process. In the past decade, the supply of state-owned construction land has increased rapidly, reaching a peak of 730,000 hm$^2$ in 2013. Such an immense amount of land supply is accompanied by the inefficient use of land. The total amount of construction land approved by the central and provincial governments from 2012 to 2016 was 1.97 million hm$^2$, and of this total, the amount of land on which construction had not been started or completed on schedule accounted for nearly one third (the data source is the “China Land and Resources Statistical Yearbook” and “China Land and Resources Bulletin” over the years). Such a background has prompted the government to strengthen regulations related to land planning, consolidation, and utilization. The central government has issued a series of land policies to regulate land development and agricultural land conversion. For example, in 2003, the Ministry of Land and Resources issued the *Urgent Notice on Clearing up Land Used in Various Parks and Strengthening Regulation and Control of Land Supply*; in 2008, the State Council issued the *Notice on Promoting Land Saving and Intensive Use*; and the Ministry of Land and Resources has revised the *Measures for Disposal of Unused Land* over the years. Various land regulation policies were established from time to time to promote the reasonable and efficient use of land.

The year 2014 was a turning point for urbanization in China over the past 20 years. The National New Urbanization Planning for 2015–2020 clearly points to the need to shift from land urbanization to population urbanization. Correspondingly, promoting the coordinated development of urban and rural areas has become the topic of the land system. Rural development relies on land development, and the confirmation and transfer of land rights are inevitable requirements of rural land development [25–27]. Accordingly, the first policy statement released by the central government in 2013 proposed to complete the confirmation, registration, and certification of rural land contractual management rights within five years and that of 2014 allowed rural collective construction land to be sold, leased, and pooled as shares have equal access to the market and with equal rights and prices as state-owned land. At the same time, the rural homestead system was reformed to promote the mortgage, guarantee, and transfer of housing property rights. How to stimulate the vitality of rural land through the reform of property rights has become an important issue in land management in China.

In summary, for the last two decades, urbanization in China has undergone a transformation from land urbanization to population urbanization, and the land system has also been transformed from urban development to urban-rural coordinated development. In this process, the land policy has been constantly adjusted with a focus on economic and social development and the contradictions that emerged during this period. Based on a simple review of the institutional background, this study identified six core topics of China’s land policy in the past 20 years, i.e., (1) land development, (2) land expropriation and demolition, (3) farmland protection, (4) land planning, (5) land consolidation and utilization, and (6) the confirmation and circulation of rural land rights.

2.3. Machine Learning for Text Documents’ Classification. Along with the transfer of text information from paper media to Internet media, the cost of text data collection and
transmission is greatly reduced, which provides application scenarios for Natural Language Processing (NLP). The core challenge of text analysis is how to extract the required information from the text accurately and efficiently. To achieve this, the computer is required to analyze and process language like a human being. One of the most important parts is to classify the text. Generally speaking, machine learning algorithms can be divided into three categories. They are supervised, semisupervised, and unsupervised methods. In this section, we briefly introduce some algorithms widely used, including Naive Bayes algorithms, support vector machines (SVMs), K-nearest neighbor (KNN), and neural networks.

\[
P(c_i | D) = \frac{P(c_i) P(D | c_i)}{P(D)},
\]

\[
P(D | c_i) = \prod_{j=1}^{n} P(d_j | c_i), \quad \text{where} \quad P(C = c_i) = \frac{N_i}{N}, \quad P\left(d_j \mid c_i\right) = \frac{1 + N_{ij}}{M + \sum_{k=1}^{M} N_{ik}}.
\]

A tiny amount of training data is needed to estimate the parameters for classification, which is an obvious advantage of the Naive Bayes classifier. In addition, the Naive Bayes classifier is proved to perform better on numeric and textual data, and the computation is easier and more efficient than other algorithms. However, the defects are also obvious. With real-world data, the conditional independence assumption will be violated. And it works poorly when features have high correlation and the frequency of a word is neglected; therefore, its applicability is seriously limited.

2.3.2. Support Vector Machines (SVMs). Support vector machines (SVMs) are a discriminative classification method, which is based on the structural risk minimization principle [31]. The core of this principle is to ensure the lowest true error by finding a hypothesis, which makes the SVMs more accurate. SVMs are used to find out the linear separating hyperplane which maximizes the margin between two datasets, i.e., the optimal separating hyperplane (OSH). The key lies in the calculation of the margin, which is based on the construction of two parallel hyperplanes, as shown in equations (2) and (3). The margin is on each side of the separating hyperplane, which is “pushed up against” the two datasets. The generalization error of the SVMs decreases with the increase in the margin. And to increase the margin, the hyperplane is required to have a larger distance to the neighboring data points of both classes.

We maximize the margin as follows:

\[
\begin{align*}
\text{minimize} \quad & \frac{1}{2} \omega^T \omega + C \left( \sum_{i=1}^{N} \zeta_i \right), \\
\omega b \zeta_i \text{ subject to} \quad & y_i (\omega^T x_i - b) + \zeta_i - 1, \quad 1 \leq i \leq N, \\
& \zeta_i \geq 0, \quad 1 \leq i \leq N.
\end{align*}
\]

2.3.1. Naive Bayesian Algorithms. The Naive Bayes classifier is a simple probabilistic classifier, which applied Bayes’ theorem with strong independence assumptions. Based on independence assumptions, the order of features is irrelevant; therefore, the existence of one feature has no influence on other features in the classification tasks [28]. With these over simplified assumptions, Naive Bayes classifiers have been proved to work unexpectedly well in many complex real-world classification applications [29, 30].

Below is the expression of the Naive Bayesian algorithm (equation group (1)):

\[
\begin{align*}
\xi (\omega, b, \zeta; \alpha, \beta) = & \frac{1}{2} \omega^T \omega + C \left( \sum_{i=1}^{N} \zeta_i \right) \\
& - \sum_{i=1}^{N} \alpha_i \left( y_i (\omega^T x_i - b) + \zeta_i - 1 \right) - \sum_{i=1}^{N} \mu_i \zeta_i, \\
= & \frac{1}{2} \omega^T \omega + \sum_{i=1}^{N} \left( C - \alpha_i - \mu_i \right) \zeta_i \\
& - \left( \sum_{i=1}^{N} \alpha_i y_i x_i^T \right) \omega - \left( \sum_{i=1}^{N} \alpha_i y_i \right) b + \sum_{i=1}^{N} a_i.
\end{align*}
\]

The SVMs is prominent for its classification effectiveness [32, 33], which makes it very suitable for theoretical understanding and analysis [34]. And also, it performs well on documents with high-dimensional input space; most of the irrelevant features in the documents are weeded out. However, the training and categorizing algorithms of the SVMs are more complex than other methods. Besides, in the training and classifying stage, more time and higher memory consumptions are required. Furthermore, as the similarity is typically calculated for each individual category, it may lead to confusions when documents are notated to several categories in the classification.

2.3.3. Neural Networks. Artificial neural networks are structured by a large amount of elements named as artificial neuron. Compared with the elements of traditional architectures, artificial neuron has larger input fan order of
magnitudes [35, 36]. Besides, they are made more sensitive to store items and more suitable for distortion tolerant storing and, therefore, can store a greater number of items displayed by high-dimensional vectors. Artificial neural networks interlink those neurons into groups with a mathematical model of information processing, as shown in equation (4). In this way, artificial neural networks have some obvious advantages. The main advantage is to perform well in complex domains, on documents with high-dimensional features and also on noisy, contradictory, discrete, and continuous data. Besides, a parallel computing architecture is employed to provide linear speed up in the matching process of computational elements. In such process, the input value of each element can be compared with the value of stored cases. However, the drawbacks are also obvious. Though the testing is very fast, the training is slow. And for users, learned results are more difficult to comprehend than learned rules. Also, empirical risk minimization (ERM) enables artificial neural networks to minimize training error, yet it may result in overfitting. For pattern p, the output from neuron j is $O_{pj}$:

$$O_{pj}(\text{net}_j) = \frac{1}{1 + e^{-\text{net}_j}},$$

$$\text{net}_j = \text{bias}^* W^\text{bias} + \sum_k O_{pk} W^k_{jk}. \quad (4)$$

### 3. Data and Study Methods

#### 3.1. Data Source and Brief Description

The data in this study were collected from the "Faxin" online platform (https://www.faxin.cn/) (more detailed introduction could be found in: https://www.faxin.cn/). The "Faxin" online platform was established and maintained by the Supreme People’s Court in 2012. After several years of development, it has become an advanced digital network platform that deeply integrates legal knowledge services and big data services for cases in China, including more than 1.4 million laws, regulations, and policy documents.

Using the search engine provided by "Faxin," all land-related regulations from the central and local governments (hereafter referred to as policy documents) from 1998 to 2018 in China were collected, accounting for a total of 22,659 documents. Among the 22,659 policy documents, 1,262 were issued by the central government, and the remaining 21,397 were issued by local governments. Figure 1 shows the distribution of land-related policy documents in the past 20 years. Obviously, the overall number of land-related policy documents grew significantly, indicating that governments have increasingly paid attention to land management issues.

#### 3.2. Study Methods

As mentioned above, the main task of this study is to classify more than 22 thousand of land policies in China according to their topics. Basically, text classification is a process of clustering, which groups a collection of objects into subsets or clusters that share similar topics [37]. In the last decades, the field of machine learning has emerged as plenty of algorithms in text mining and text clustering [38]. For example, Avalos reviewed a series of text mining methods, such as real-time data text mining based on a gravitational search algorithm, clustering approach using a combination of a gravitational search algorithm, and k-harmonic means [39]. Recently, Zabith and Osman propose a novel predictive analytics framework in the work of unstructured text classification and analysis [40]. In the existing studies, these methods have been widely applied in engineering management [41] and in bibliometrics [42]. Combining the consideration of robustness and accuracy of text mining [43, 44], this study used the Latent Dirichlet Allocation (LDA) model to analyze 22,659 policy texts. In the field of machine learning, the LDA model occupies a very important position in topic models and is often used for text classification. The essence of the LDA model is a Bayesian probabilistic model that contains a three-layer structure of the corpus, topic, and word. In this model, each word in a corpus is considered to select a certain topic with a certain probability, and a certain word is selected from this topic with a certain probability. The corpus, topic, and word follow the Dirichlet distribution. In the LDA algorithm, a corpus represents a probability distribution composed of some topics, and a topic is a probability distribution composed of many words. The text clustering results generated by the LDA model can reveal the keywords and specific probabilities of each topic, and researchers can interpret the meaning of the corpus accordingly [45]. Therefore, compared to manually reading and understanding a policy text, the LDA model can efficiently and accurately help researchers identify the topics of the corpus.

In brief, the LDA model assumes a hierarchical structure among words, topics, documents, and corpus. The documents and the words could be observed, but a latent structure of topics, topic distributions per document, and word distributions per topic exist [46]. Therefore, LDA could be viewed as an approach where multiple words are estimated to be associated with a few latent topics. In more formal terms, LDA is a model based on observed variables (words) and hidden variables (topics) that define a joint probability distribution. The joint probability distribution is then used to calculate, according to Bayes rule, a conditional or posterior distribution of the hidden variables given the observed variables [47].

The more formal and mathematical notation could be presented as follows. A word is denoted as $w$, and a document is a collection of $N$ words, $w = (w_1, w_2, \ldots, w_n)$, where $w_n$ is the $n^{th}$ word in the document. It is worth noting that the ordering of words is not important since the LDA model assumes the approach of a "bag of words," in which the co-occurrence instead of the ordering of words is used to identify the underlying topics. Finally, a corpus is a collection of $M$ documents denoted as $D = (w_1, w_2, \ldots, w_M)$. Following this notation, the generative process—the assumed process that generated the documents, topics, and words—can be described as follows [39]:

(i) choose $N \sim \text{Poisson}(\xi)$

(ii) choose $\theta \sim \text{Dir}(\alpha)$
(iii) For each of the $N$ words $w_n$,

(i) choose a topic $z_n \sim \text{Multinomial} (\theta)$

(ii) choose a word $w_n$ from $p(w_n | z_n, \beta)$ and a multinomial probability conditioned on the topic $z_n$

$\theta$ is used to measure the topic proportions, which are the sum of the probabilities for each topic in the $d$th document. In the theoretical issue definition model, $\theta$ was the summation of the issue dimensions weighted by salience and that concept is operationalized as the topic proportions estimated by LDA [46]. These proportions are drawn from a Dirichlet prior, $\theta \sim \text{Dir}(\alpha)$ where $\alpha$ is the Dirichlet distribution’s shape parameters. Tenumber of topics $K$, and by extension the dimensionality of Dir($\alpha$) and the topic variable $z$, is assumed a priori and is also assumed as fixed. Note that a proportion is estimated for each of the $K$ topics within each document. The standard Dirichlet prior of $\alpha$ is $50/K$. The possibility of density is presented as

$$p(\theta | \alpha) = \frac{\Gamma(\sum_{i=1}^{K} \alpha_i)}{\prod_{i=1}^{K} \Gamma(\alpha_i)} \theta_{1}^{\alpha_1-1} \cdots \theta_{K}^{\alpha_K-1}. \quad (5)$$

The distribution of words is parameterized by $\beta$. As noted, the observed and latent variables form a joint distribution that, given the parameters $\alpha$ and $\beta$, topic mixture $\theta$, a set of $N$ topics $z$, and a set of $N$ words $w$ can be expressed as follows:

$$p(\theta, z, w | \alpha, \beta) = p(\theta | \alpha) \prod_{n=1}^{N} p(z_n | \theta) p(w_n | z_n, \beta). \quad (6)$$

This joint distribution is used to calculate a posterior distribution of topic probabilities for each document, as expressed in the following:

$$p(\theta | w, \alpha, \beta) = \frac{p(\theta, z, w | \alpha, \beta)}{p(w | \alpha, \beta)}. \quad (7)$$

The numerator is the joint distribution of all random variables, and the denominator is the probability of obtaining the observed corpus under any topic model. These probabilities could be summed for overall possible topic structures. Therefore, this needs to be approximated using either sampling-based or variational approximations. Gibbs sampling is used to estimate the posterior distribution. By doing this, LDA could generate a cluster of words. The whole process of the LDA model could be summarized in Figure 2. And in the field of policy text analysis, researchers could infer the topics of each policy according to the co-occurrence of words, which is shown by the logic presented in Figure 3 [46].

Before using the LDA model, more than 20,000 policy documents need to be segmented. This study uses the Jieba library in Python to divide sentences into individual words (the Jieba library integrates two word segmentation methods based on rules and statistics, which effectively improve the accuracy of word segmentation. A more detailed introduction of the Jieba library could be found at https://pypi.python.org/pypi/jieba/. Based on the Chinese stop word list from the Harbin Institute of Technology, the relevant words in this study were updated to remove the stop words with no actual connotation (the stop vocabulary list of the Harbin Institute of Technology contains 1893 words with no actual connotation but a very high frequency, such as “oh” and “is.” Including these words in the word, frequency statistics will obviously affect the accuracy of the results. So, we need to eliminate them.). Next, the term frequency-inverse document frequency (TF-IDF) algorithm was used to optimize the word segmentation results, and the words that frequently appeared but had little impact on the
topics of the corpus were excluded. After these progress, the contents of the 22,659 policy texts were transformed into 189,582 important keywords, which laid the foundation for the construction of the LDA model. For the number of topics, this study first sets the number of topics to 15 (setting the number of topics to 15 is a result of the strategy of "more is better than less." That is, too many topics can be combined through the researcher's understanding, while too few topics may miss some important topics. In fact, as shown in Table 2, our preset number of themes is indeed too much: out of 15 themes, there are 2 themes whose meaning cannot be determined, and the content of 4 themes is repeated.)

4. Results of the Topic Model-Based on the Theoretical and Institutional Background

In this section, results of the topic model are presented based on the theoretical and institutional background of land policy in China. By those results, the keywords in the policy documents can be fitted into several topics; then, the proportion of each topic in all documents can be calculated based on the frequency of keywords. Those are the foundations for the analysis of the key point and the temporal and spatial evolution of China’s land policy in the next section.

First, we present the keywords with a word cloud, which is obtained with word segmentation and word frequency statistics. As shown in Figure 4, the larger the font size of the word in the word cloud, the higher the frequency of the word in the policy texts. In addition to the word “land,” the two most frequently used words are “construction” and “cultivated land,” indicating that a large portion of the land-related policy documents deals with the contradictions between “promoting development” and “preserving rice bowls.” Other high-frequency words include “levy,” “right to use,” and “planning,” indicating that issues such as land expropriation, land ownership, and land planning also frequently appear in policy documents. Those keywords and their frequency can roughly reflect the focus of land policies.

Second, the fitting results of feature words and topics by the LDA model are shown in Table 2. Among the 15 preset topics, 13 have clearly determined feature words corresponding to the topic content. Taking topic 1 as an example, according to words such as “agricultural land,” “arable land,” “conversion,” “approval,” and “reclamation,” it is easy to understand that this topic discusses a series of policies for the protection of arable land in the process of agricultural land conversion, which belongs to the topic of cultivated land protection. The fitting results of the two topics (No. 14 and 15) are unclear. Therefore, they are excluded from the subsequent analysis. With these fitting results, the key points of policy documents can be easily analysed by the classification of topics.

Third, we reclassified the topics to avoid repeats. Among the 13 topics presented in Table 2, some are duplicates (e.g., topics 1 and 5), and some have overlapping connotations...
Table 2: Land policy text LDA model results, subject inference, and reclassification.

<table>
<thead>
<tr>
<th>Serial number</th>
<th>Feature words</th>
<th>Topic inference</th>
<th>Reclassification</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>&quot;Land&quot; + &quot;agricultural land&quot; + &quot;people's government&quot; + &quot;cultivated land&quot; + &quot;plan&quot; + &quot;conversion&quot; + &quot;project&quot; + &quot;approval&quot; + &quot;annual&quot; + &quot;construction project&quot; + &quot;approval&quot; + &quot;indicator&quot; + &quot;city&quot; + &quot;agree&quot; + &quot;reclamation&quot;</td>
<td>Cultivated land protection</td>
<td>Cultivated land protection</td>
</tr>
<tr>
<td>2</td>
<td>&quot;Contracting&quot; + &quot;rural land&quot; + &quot;circulation&quot; + &quot;confirmation of rights&quot; + &quot;registration&quot; + &quot;operation rights&quot; + &quot;rural&quot; + &quot;agriculture&quot; + &quot;organization&quot; + &quot;farmers&quot; + &quot;disputes&quot; + &quot;city&quot; + &quot;People's government&quot; + &quot;certificate&quot; + &quot;collective&quot;</td>
<td>Land transfer</td>
<td>Land right confirmation and circulation</td>
</tr>
<tr>
<td>3</td>
<td>&quot;Project&quot; + &quot;city&quot; + &quot;development&quot; + &quot;land reserve&quot; + &quot;fund&quot; + &quot;unit&quot; + &quot;implementation&quot; + &quot;idle&quot; + &quot;disposal&quot; + &quot;relevant&quot; + &quot;responsible&quot; + &quot;land&quot; + &quot;organization&quot; + &quot;organization&quot; + &quot;people's government&quot;</td>
<td>Land consolidation</td>
<td>Land remediation and utilization</td>
</tr>
<tr>
<td>4</td>
<td>&quot;Transfer&quot; + &quot;right to use&quot; + &quot;land&quot; + &quot;listing&quot; + &quot;city&quot; + &quot;state-owned land&quot; + &quot;auction&quot; + &quot;transfer&quot; + &quot;method&quot; + &quot;people's government&quot; + &quot;allocation&quot; + &quot;tendering&quot; + &quot;purpose&quot; + &quot;transfer fee&quot; + &quot;should&quot; + &quot;Facility&quot; + &quot;standard&quot; + &quot;agriculture&quot; + &quot;cm&quot; + &quot;cultivated land&quot; + &quot;project&quot; + &quot;unit&quot; + &quot;quality&quot; + &quot;soil&quot; + &quot;planting&quot; + &quot;formula&quot; + &quot;engineering&quot; + &quot;fertilization&quot; + &quot;calculate&quot; + &quot;include&quot;</td>
<td>Land transfer</td>
<td>Land development</td>
</tr>
<tr>
<td>5</td>
<td>&quot;People's government&quot; + &quot;relocation&quot; + &quot;unit&quot; + &quot;relocation&quot; + &quot;building area&quot; + &quot;property&quot; + &quot;area&quot; + &quot;grant&quot; + &quot;state-owned land&quot; + &quot;law&quot; + &quot;price&quot; + &quot;city&quot; + &quot;implementation&quot; + &quot;residential&quot; + &quot;notice&quot;</td>
<td>Land acquisition and demolition</td>
<td>Land acquisition and demolition</td>
</tr>
<tr>
<td>6</td>
<td>&quot;Land&quot; + &quot;approval&quot; + &quot;road&quot; + &quot;people's government&quot; + &quot;land&quot; + &quot;project&quot; + &quot;agree&quot; + &quot;plan&quot; + &quot;guigang city&quot; + &quot;bureau&quot; + &quot;limited company&quot; + &quot;use right&quot; + &quot;city&quot; + &quot;avenue&quot; + &quot;residential&quot;</td>
<td>Land development</td>
<td>Land development</td>
</tr>
<tr>
<td>7</td>
<td>&quot;Registration&quot; + &quot;land resources&quot; + &quot;unit&quot; + &quot;application&quot; + &quot;certificate&quot; + &quot;collective land&quot; + &quot;change&quot; + &quot;cadastre&quot; + &quot;right to use&quot; + &quot;data&quot; + &quot;land&quot; + &quot;national land&quot; + &quot;materials&quot; + &quot;should&quot; + &quot;handle&quot;</td>
<td>Land right</td>
<td>Land right confirmation and circulation</td>
</tr>
<tr>
<td>8</td>
<td>&quot;Land&quot; + &quot;renovation&quot; + &quot;people's government&quot; + &quot;announcement&quot; + &quot;agricultural land&quot; + &quot;woodland&quot; + &quot;implementation&quot; + &quot;village&quot; + &quot;utilization&quot; + &quot;project&quot; + &quot;project&quot; + &quot;city&quot; + &quot;yuandi&quot; + &quot;industry and mining&quot; + &quot;standard&quot;</td>
<td>Land consolidation</td>
<td>Land remediation and utilization</td>
</tr>
<tr>
<td>9</td>
<td>&quot;Relocation&quot; + &quot;standard&quot; + &quot;People's government&quot; + &quot;compensation&quot; + &quot;street&quot; + &quot;village&quot; + &quot;group&quot; + &quot;plan&quot; + &quot;cultivated land&quot; + &quot;subsidies&quot; + &quot;personnel&quot; + &quot;approval&quot; + &quot;announcement&quot; + &quot;town&quot; + &quot;Young Miao&quot;</td>
<td>Land acquisition expropriation</td>
<td>Land acquisition and demolition</td>
</tr>
<tr>
<td>10</td>
<td>&quot;Planning&quot; + &quot;land&quot; + &quot;land use&quot; + &quot;overall planning&quot; + &quot;cultivated land&quot; + &quot;scale&quot; + &quot;control&quot; + &quot;district&quot; + &quot;town&quot; + &quot;indicator&quot; + &quot;basic farmland&quot; + &quot;protection&quot; + &quot;edit&quot; + &quot;approval&quot; + &quot;implement&quot;</td>
<td>Land planning</td>
<td>Land planning</td>
</tr>
<tr>
<td>11</td>
<td>&quot;City&quot; + &quot;land&quot; + &quot;cultivated land&quot; + &quot;protection&quot; + &quot;district&quot; + &quot;land resources&quot; + &quot;responsibility&quot; + &quot;People's government&quot; + &quot;planning&quot; + &quot;basic farmland&quot; + &quot;organization&quot; + &quot;land utilization&quot; + &quot;national land&quot; + &quot;system&quot; + &quot;notice&quot;</td>
<td>Cultivated land protection</td>
<td>Cultivated land protection</td>
</tr>
<tr>
<td>12</td>
<td>&quot;Land price&quot; + &quot;benchmark&quot; + &quot;level&quot; + &quot;update&quot; + &quot;subsidy&quot; + &quot;township&quot; + &quot;notification&quot; + &quot;urban area&quot; + &quot;people's government&quot; + &quot;publication&quot; + &quot;standard&quot; + &quot;district&quot; + &quot;table&quot; + &quot;city&quot; + &quot;revision&quot;</td>
<td>Grading</td>
<td>Land remediation and utilization</td>
</tr>
<tr>
<td>13</td>
<td>&quot;Enterprise&quot; + &quot;use tax&quot; + &quot;urban&quot; + &quot;people's government&quot; + &quot;city&quot; + &quot;unit&quot; + &quot;notice&quot; + &quot;land&quot; + &quot;method&quot; + &quot;tax amount&quot; + &quot;administrative&quot; + &quot;approval&quot; + &quot;standard&quot; + &quot;payment&quot; + &quot;use fee&quot;</td>
<td>N.A.</td>
<td>N.A.</td>
</tr>
<tr>
<td>14</td>
<td>&quot;People's government&quot; + &quot;land&quot; + &quot;VAT&quot; + &quot;unit&quot; + &quot;rural&quot; + &quot;approval&quot; + &quot;house site&quot; + &quot;standard&quot; + &quot;development&quot; + &quot;liquidation&quot; + &quot;collective&quot; + &quot;transaction&quot; + &quot;should&quot; + &quot;transfer&quot; + &quot;method&quot;</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Data source: the author compiled the study based on the results of the LDA model of the 1998–2018 land policy text.
Such a classification may lead to confusion in the subsequent analysis. Therefore, based on the analysis of the theoretical and institutional background in Section 2, the 13 topics were reclassified, and the results are shown in the last column of Table 3. After reclassification, the 22,659 policy documents can be summarized into six major topics that match the theoretical analysis. In the following sections, the six topics are used to analyze the evolution and spatial differentiation of land policies. This reclassification shows an understanding of the background of the land system in China.

And lastly, the number and proportion of each topic are shown in Table 3, which visually displays the key points of land policy documents. Clearly, the proportion of policy documents related to the topics of arable land protection and land development accounts for more than 50%. The proportions of land planning, land consolidation and utilization, and land expropriation and demolition all account for approximately 10%.

5. The Temporal and Spatial Evolution of China’s Land Policy

5.1. Temporal Evolution and Change Logic of China’s Land Policy. To explore the characteristics of the temporal evolution of China’s land policy, the number of policy documents related to each topic in each year from 1998 to 2018 and the proportion of each topic were calculated, and the results are shown in Figure 5. Figure 5 shows that, in the past two decades, arable land protection and land development have received the most attention in land policy documents, and there is a tradeoff between them. Prior to 2016, the frequency of topics on arable land protection was significantly higher than that on land development. In particular, in 1998 and 1999, when the Land Administration Law was revised and implemented, and in the three years, after the Third Plenary Session of the 16th CPC Central Committee in 2003, more than 50% of policy documents focused on arable land protection. In 2017 and 2018, land development surpassed arable land protection for two consecutive years and became the main focus of land policy.

Besides, the two topics of land consolidation and utilization and land expropriation and demolition are also worth noting. Clearly, these two topics gained much attention in 2013, which declined slightly in subsequent years, yet are still higher than that in the past decade. That is because the report of the 18th CPC National Congress emphasized land use efficiency and land expropriation and demolition, which proposed that “reforming the land expropriation system and increasing the proportion of farmers in the distribution of land value-added income” and “substantially reduce land consumption intensity and improve utilization efficiency and benefits.” The emphasis of the CPC Central Committee on land use efficiency and land expropriation and demolition has encouraged governments at all levels to pay more attention to these issues, thus promoting the introduction of more relevant policy documents. In contrast, the proportion of policy documents on the topic of land planning has remained at a relatively stable level in the past decade, approximately 10%, while that on the topic of land rights confirmation and circulation has declined since 2015.

5.2. Spatial Differences in China’s Land Policy and Causes. China has a vast territory, and the socioeconomic, cultural, and historical backgrounds varied largely among regions. Local governance is adapted to local conditions, which is reflected in the tremendous differences in policy implementation and regulations [48]. For example, government subsidies have different degrees of influence on the supply chain and the market according to different levels of subsidies in different regions [49]. Land policy priorities also differed spatially. According to the classification of the National Bureau of Statistics, the 31 provinces were divided into four major regions, i.e., the eastern, central, western, and northeastern regions, and the proportions of land policy documents on the various
topics in the four major regions in the past 20 years are shown in Figure 6. First, for the topics of arable land protection and land development, the degree of attention for arable land protection in the northeastern and central regions (0.45 and 0.44, respectively) is significantly higher than that in the eastern and western regions (0.35 and 0.32, respectively), while the land policy of the eastern and western regions focuses more on land development. We believe that the tradeoff between “promoting development” and “preserving rice bowls” is mainly determined by two factors. The first factor is the regional economic structure. If a region is more dependent on agriculture, more attention is on the protection of arable land. Among the four major regions, the proportion of agriculture in the total GDP in the northeastern and western regions is approximately 11%, followed by the central region at 8%, while the eastern region only accounts for 4%. Therefore, the proportion of policy documents related to arable land protection in the northeastern and central regions accounts for more than 40%. However, for the western region, the factor of the regional economic structure cannot provide a full explanation. Since agriculture output accounts for up to 11% of the GDP in this region, its land policy documents pay the least amount of attention to the protection of arable land. Therefore, the second explanatory factor is the pressure on arable land protection. As pointed out by Liang et al., the western region obtained a large construction land quota after 2003 [50]. With more construction land quota, the pressure on arable land protection in the western region has been obviously smaller than that of other regions, so it is not surprising that the land policy of this region emphasizes the “use of land for development.”

In terms of other topics, the eastern region, which has gradually moved toward intensive development, has placed more emphasis on land planning than the other three regions. While the western region, with its loose land controls, has paid less attention to land consolidation and utilization. Agriculture plays an important role in the northeastern region; therefore, the topic of rural land rights confirmation and circulation has received significantly more attention. As land expropriation and demolition are inevitable in the process of land development, the topic of land expropriation and demolition appears more frequently in land policy

<table>
<thead>
<tr>
<th>Topic</th>
<th>Number of policy texts</th>
<th>Proportion (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cultivated land protection</td>
<td>7851</td>
<td>36.69</td>
</tr>
<tr>
<td>Land planning</td>
<td>2057</td>
<td>9.61</td>
</tr>
<tr>
<td>Land consolidation and utilization</td>
<td>2547</td>
<td>11.90</td>
</tr>
<tr>
<td>Land right confirmation and circulation</td>
<td>1323</td>
<td>6.18</td>
</tr>
<tr>
<td>Land expropriation and demolition</td>
<td>1955</td>
<td>9.14</td>
</tr>
<tr>
<td>Land development</td>
<td>5664</td>
<td>26.47</td>
</tr>
<tr>
<td>Total</td>
<td>21,397</td>
<td>100.00</td>
</tr>
</tbody>
</table>

Data source: the classification results of the land policy texts in 1998–2018 are obtained by authors.

Figure 5: Characteristics of the temporal evolution of land policy priorities.
documents in the eastern and western regions, where land
development is more emphasized.

6. Conclusion and Discussion

The vast number of policy documents issued by multilevel
governments makes it impossible to analyze and extract
useful information from them manually. The existing
studies on policy analysis mainly focus on some important
policy documents, making it hard to have a detailed
analysis of the temporal change and spatial differences in
policies. This study tries to use the way of machine
learning to have a thorough analysis of the evolutionary
logic and spatial differentiation of China’s land policy in
the past two decades. That is an innovation to combine
policy analysis with the technology of machine learning
and may have a further application to more wide-ranging
text analysis.

The results show that land development and arable land
protection have been the main focus of land policy over the
years, the tradeoff between which depends largely on the
political and economic background. And the focus on other
topics is also determined by political and economic con-
ditions. For instance, since the 18th CPC National Congress,
topics such as land use efficiency and the rights of land-
expropriated farmers have gained more attention. From a
spatial point of view, the focus of the local land policy is
largely determined by the local socioeconomic character-
stics and is affected by the land development mode formed
under the central-local relationship.

In the future, with the end of the era of urban sprawl,
improving land use efficiency may get more attention.
And the contradiction between “promoting development”
and “preserving rice bowls” may continue to exist, thereby
challenging the land governance ability of governments.
Meanwhile, the land redevelopment will inevitably affect
the interests of some social groups. For example, land
expropriation and relocation may affect the land use rights
of the original holders. Balancing the interests of all
parties and effectively adhering to the “people-oriented”
principle in the process of land development may become
a new issue in land governance.

Data Availability

The data supporting the findings of this study are available
within the article.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

The research was supported by the National Natural Science
Foundation of China: research on the mechanism of arable
and quality guarantee and counter measures based on the
verification of arable land redline protection (project no.
21BGL163).

References

and land development processes in contemporary China,”
orative Governance: The Changing Role of Local Governments
nance from the perspective of policy tools—analysis of 192
policy texts (1992-2015),” Chinese Public Administration,
collective construction land transfer policy based on content
analysis method,” China Land Science, vol. 29, pp. 25–33,
2015.
Dynamic decision model for the solar photovoltaic supply
chain,” Applied Mathematical Modelling, vol. 89, pp. 1106–
1123, 2021.


