Features and Comparative Research on Ecological Civilization Vocabularies in the Five-Year Plan of China: An Analysis Based on Semantic Phrases

Shaoli He, Shanghai Maritime University, China
Xianhua Wu, Shanghai Maritime University, China*
Ji Guo, Shanghai Maritime University, China

ABSTRACT

The Five-Year Plan (FYP) in China guides the social and economic development. The features of ecology-related vocabularies in national and regional FYPs help to study China’s emphasis and focus on ecological civilization. In this paper, a multi-phrase dictionary with syntactic and semantic features is constructed to analyze the features and trends of ecology-related vocabularies through social network and phrase topic model. Then, based on studies on the 13th and 14th FYPs of China and 31 provinces and cities, it is found that remarkable different stage definitions and tasks of ecological civilization construction are proposed according to different description vocabularies, with more pragmatic measures to fulfill its “carbon” target in 2030 as a duty-bound responsibility and mission to cope with climate change. The research ideas and methods in this paper can provide references for similar policy text analysis, and the conclusions from empirical studies are helpful to grasp the focus and trends of ecological civilization-related policies in China and various provinces and cities.

KEYWORDS

1. INTRODUCTION

The “Outline of the Five-Year Plan for National Economic and Social Development” is a medium and long-term plan for economic and social development in China (Lin, 2021), which sets clear goals and directions for economic and social development (Group, 2020). The “Five-Year Plan” was started in 1953. So far, thirteen five-year plans have been implemented in China. It is driving China out of poverty, building a moderately prosperous society, and marching towards the dream of the great rejuvenation of the Chinese nation. The “Thirteenth Five-Year Plan” is a key stage for building...
a well-off society in an all-round way and an important node for achieving the first centenary goal. The “Fourteenth Five-Year Plan” will embark on a new journey of comprehensive modernization and building a grand blueprint for China’s development in a new era. In the “Five-Year Plan”, the ecological environment construction plays an important role, such as the research on the driving factors of energy consumption and intensity changes (Yan & Su, 2020), the energy system structure adjustment and renewable energy (Hong et al., 2013; Lo & Wang, 2013). In 2007, the 17th National Congress of the Communist Party of China (CPC) took the construction of ecological civilization as one of the goals of building a well-off society. The 13th Five-Year Plan promotes an innovative, coordinated, green, open, and shared development. During these five years, great achievements have been made in ecological environment, and the construction of ecological civilization has been integrated into all aspects of economy and society and the whole process of development. The 14th Five-Year Plan takes promoting high-quality development as the theme, and promotes the overall layout of economic, political, cultural, social and ecological civilization construction as a whole. It unswervingly implements the new development concept of innovation, coordination, green, openness and sharing. During the 13th Five-Year Plan and 14th Five-Year Plan, the construction of ecological civilization has been carried out in an all-around way and deepened continuously. Through a comparative study, we can understand the changing trends and laws of ecological civilization construction in different periods. Therefore, this paper chooses these two periods for research (Onan et al., 2016; Vega-Oliveros et al., 2019). Five-Year Plans are development plans with Chinese characteristics. Due to the development of massive texts and text mining tools, the analysis of the Five-Year Plans focuses on news commentary and expert opinions. Text mining tools can help to get the main idea of texts through keywords according to language features. Starting from the keyword frequency and the correlation between keywords, this paper makes an objective and quantitative analysis of the policy content. Through analysis of the text content, we can clarify the core and key points of the plans, and grasp the development theme and focus. This paper compares the focus change in different periods, explores the changing trends, and provides an objective and scientific basis for the implementation of policy formulation.

2. LITERATURE REVIEW

2.1 Overview of Quantitative Research on Ecological Civilization in Policy Texts

There are narrow and broad definitions of ecological civilization. In a narrow sense, ecological civilization involves ecosystem protection, environmental quality, and resource utilization (Lu et al., 2020; Wu et al., 2020). In a broad sense, ecological civilization covers the layout of land space, the ecological economy development, the ecological environment protection, the ecological culture construction (Cai et al., 2021; Huang & Wu, 2021), strategic emerging industries (Meng et al., 2021), economic development concept modes (Zhang et al., 2016). At present, the quantitative analysis of ecological civilization policy includes environmental pollution control, energy resource utilization and new energy development, and sustainable development policy effect evaluation (Thuestad, 2020). The green governance of the government has undergone logical transition in governance concepts, governance modes, governance supervision, and ecological civilization construction, showing that ecological protection and environmental governance policies keep pace with the times (Ran, 2020). Environmental governance needs the leadership of the government, the guidance of the news media, and the active participation of enterprises and the public in order to improve the overall environmental governance effect (Lam et al., 2019). Resources and energy industry policies involve resource utilization, new energy, energy structure transformation and international cooperation (Lu et al., 2019; Tong et al., 2021). Ecological economy not only promotes the upgrading and transformation of energy, but also improves the environment and help to realize targeted poverty alleviation (Zhang et al., 2018).
In the quantitative research on ecological civilization in policy texts, scholars at home and abroad often take the institutions, executors and the policy intensity as one of the weights or factors of quantitative analysis, and then conduct in-depth quantitative analysis combined with statistical data after coding conversion. There are few studies on the quantitative analysis of the policy text. The current research focuses ecological civilization in a narrow sense, or is limited to a certain field, such as resources (W. Du et al., 2021), pollution (Lu et al., 2020), emission reduction (Baldwin et al., 2018; Jean-Philippe, 2019), etc. Combined with the research of scholars, taking the Guide for the Ecological Civilization Construction Standard System (2018-2020) as the framework, this paper focuses on the ecological civilization-related ecological environment, resources and energy, green economy and pollution control to conduct research.

### 2.2 Overview of Phrases and Topic Models

Key-phrase extraction can be divided into three categories: supervised, semi-supervised and unsupervised. Unsupervised automatic extraction technology includes statistical methods, machine learning algorithms, deep learning algorithms, and a combination of multiple algorithms. The statistical methods are mainly probabilistic models and Markov (Chang & Choi, 2006; Feng & Croft, 2001). Compared with the supervised manual classification method, it improves the efficiency of phrase generation and reduces labor costs. The development of artificial intelligence and the application of big data have accelerated the application of machine learning and deep learning algorithms, and the introduction of the attention mechanism has improved the accuracy of the contextual semantic features of phrases (Lei et al., 2021). Combined with the semantic network and the polycentric network, the keyword extraction method based on the hierarchical structure obtains keywords that can effectively reflect the internal feature and topic (Sung & Kim, 2019). It focuses on context-relevant keyword analysis, and continuously improves the accuracy of keyword extraction (Eirini & Grigorios, 2018), which promotes the development of text analysis. The existing research focuses on comprehensive methods, and continuously improves the accuracy of keywords and the relevance between contexts through the combination of various algorithms (Eirini & Grigorios, 2018). For example, the phrase extraction method based on part-of-speech rules, combined with statistical features and graph models, obviously improves the accuracy and recall rate of keywords (Haifhua et al., 2022). In the unsupervised keyword algorithm, the title and abstract words in research papers are referred to as reference word vectors. The quality of keywords embedded in the content analysis is better than that of global representation (Papagiannopoulou & Tsoumakas, 2018).

Topic model is one of the most advanced techniques in text mining, and it was evolved from the earliest news topic tracking (Allan, 2002). The LDA (Latent Dirichlet Allocation) proposed by (Blei et al., 2003) is a classic model for text analysis. Based on the unsupervised “bag of words”, the topic is explained by using the most frequent word in each topic to analyze the simple relationship between the document and the topic. Based on this, the author expands the topic model, hierarchical topic model (Xu et al., 2018), and mixed topic model (Michal et al., 2010). With the popularity of applications, there are limitations in the probabilistic reasoning topic models. Scholars attempt to build a neural topic model of generative confrontation network (Wang et al., 2019), a word-pair topic model suitable for social media (He et al., 2017), and a neural network topic model that combines word vectors and entity vectors, etc. (Zhao et al., 2021). During the formulation and implementation of environmental policies, through text analysis of stakeholders (government, environmental protection organizations, news media), the results demonstrate that the government and environmental protection organizations are more concerned about pollutant emission and control, while the media are concerned about the impact of air pollution on health (Lam et al., 2019). Using the LDA theme model, the effectiveness and quantity intensity index of air pollution prevention and control policy are constructed. The results show that economic incentives and supportive policies have limited effect on emission reduction, and it is necessary to improve the legal effect of policies (H. Du et al., 2021). To understand the impact of COVID-19 space prevention policy on passengers’ experience, the space policy review texts of 64
major hubs in the United States find that passengers have good experience of staff and shop services, but poor experience of space services. This result provides reference for improving the health policy of transportation hubs in the future (Park et al., 2022).

In summary, the key-phrase extraction statistical methods lack semantic relevance, ignore the contextual information, so the phrases obtained are not representative and accurate. The phrase extraction of machine and deep learning are mostly applied into search engines, translation, robot services, etc, which are professional and high in development costs. At the same time, it highly depends on the quality and scale of the training set, and the training set generation is marked manually, with high cost and low portability. There are limitations of lacking language information and topic quality and data sparsity in “bag-of-words” hypothesis of the LDA topic model. The fusion of word embedding and neural network improve semantic accuracy, but the overall model efficiency is reduced. Therefore, the topic quality is low, and there are problems such as small information granularity, low topic recognition, and fuzzy meaning of topic words. Meanwhile, due to the language features of Chinese, the above-mentioned research methods based on English cannot be directly used. In this paper, through the word pair method, the key-phrases are extracted through grammatical, semantic rules and information entropy knowledge. A phrase database is set by using the text data, which is transferable in the semantic categories and normative texts. With the key-phrase as the basis of the LDA model, the topic noise and semantic ambiguity are reduced, and the accuracy of keywords are improved.

3. MODEL ESTABLISHMENT

3.1 Research Ideas

Figure 1 shows the analysis of the “Thirteenth Five-Year Plan” and “Fourteenth Five-Year Plan”. First, construct a phrase dictionary, preprocess the data and extract policy phrases. By calculating point-wise mutual information (PMI), grammatical rules, and similarity rules, the phrases are selected to generate a policy phrase dictionary. In this way, the unregistered new phrases can be obtained, and the newly generated phrases with contextual semantic features can satisfy the grammatical collocation rules of Chinese. Second, co-word analysis. The phrase dictionary is taken as a word segmentation library to obtain co-words in text data, and perform co-word clustering and evolution analysis. Perform co-word clustering through the community detection algorithm based on modularity, and the relationship among phrases is explicitly shown in a structured, clear and intuitive way. Then the co-word evolution in the two periods is analyzed in depth by using the clustering results. Third, topic analysis. With the phrases as the basic vocabulary unit of the model, the topic is analyzed through the LDA topic model that is widely used at present, improving the accuracy of the semantic expression of phrases. The high frequency words are filtered out through the word frequency ratio to improve the topic model quality.

3.2 Phrase Dictionary

According to the features of the Chinese language, the text can be divided into characters, words, phrases, sentences, paragraphs and chapters based on the granularity. Due to the vagueness of the meanings expressed by single character, words are often used as the basic analysis unit in text mining. Using words may lose the semantic relevance among words, and the same words can express different meanings in different contexts. Phrases not only retain the contextual semantic information to a certain extent, but also express the text content more clearly, with finer granularity and higher flexibility than sentences and paragraphs. A good phrase dictionary can reduce semantic ambiguity, have better text description capabilities, and provide an accurate word segmentation basis for in-depth mining. Based on the bigram proposed by (Cheng et al., 2014), the phrases are extracted by using the statistics and part-of-speech rules. The phrase extraction process is shown in Figure 2.
The construction of phrase dictionary includes data preprocessing and phrase extraction. Data preprocessing includes PDF document conversion, content extraction from image text, table content reading, and text preprocessing (word segmentation, deleting stop words, part-of-speech tagging, etc.). Phrase extraction includes calculating word relevance by using PMI value, filtering by grammatical rules, merging cross elements and deleting overlapping elements by similarity rules. Finally, a policy phrases dictionary is generated.

① Word relevance. PMI is used to measure the degree of interdependence between words in binary phrases and multiple groups, and its calculation formula is as follows:

$$ \text{PMI}(w_i, w_j) = \log \frac{p(w_i, w_j)}{p(w_i)p(w_j)} \quad \text{(Binary phrases)} \quad (1) $$

$$ \text{PMI}\left(\sum_{i=1}^{n} w_i\right) = \log \frac{p\left(\sum_{i=1}^{n} w_i\right)}{\sum_{i=1}^{n} p(w_i)} \quad \text{(Multiple phrases)} \quad (2) $$

Where $\text{PMI}(w_i, w_j)$ in formula (1) is the PMI value of phrases $w_i$ and $w_j$, $p(w_i, w_j)$ is the probability that the phrases $w_i$ and $w_j$ appear at the same time. $p(w_i)$ and $p(w_j)$ represent the probability that the phrases $w_i$ and $w_j$ appearing in the same article, respectively. From binary phrases to multiple groups, as shown in formula (2).

② Rule-based filtering. The word rules follow the Modern Chinese dictionary, combined with scholars’ research on Chinese participles (Jiwei & Dongfan, 2010), key phrase extraction, and phrase structure rules (Haihua et al., 2022; Weidong, 2017). Which include common nouns, verbs, adjectives, pronouns and quantifier phrases, and specific phrase collocation is shown in Table 1. “a” represents adjectives, including adjective morphemes, adjective idioms, distinctive word idioms; “n” represents nouns, including adnouns, names, place names, transliterated place names, organization names, other special nouns, and ancient names; “v” represents verb, including...
noun-verbs, adverbs, formal verbs, and directional verbs; "r" represents pronouns, including personal pronouns and demonstrative pronouns.

③ Delete overlapping elements. The lower-order tuple can be either a single high-frequency word or a subset of a group of higher-order elements. In the former case, the low-order tuple needs to retain, and in the latter case, the low-order tuple is deleted as overlapping elements. Discriminant rule: If the frequency of low-order tuples was close to high-order tuples, it means that low-order
tuples in the text almost always appear as part of high-order tuples, and high-order tuples were the common or specialized words. Therefore, delete the low-order tuples from the dictionary. For example, the triple \((w_i, w_j, w_k)\) contains two binary groups \((w_i, w_j)\), \((w_j, w_k)\). If \(Fre(w_i, w_j) \approx Fre(w_j, w_k)\) or \(Fre(w_i, w_j) \approx Fre(w_j, w_k)\), delete \((w_i, w_j)\), otherwise retain it.

Similarly, a subset of a unitary in a binary group can follow the word method and remove the overlapping elements.

④ Merge cross-elements. Unrelated low-order tuples have a low probability to generate related high-order words with a merger. Therefore, combining overlapped words can efficiently generate new strongly associated high-order tuples. Therefore, combining overlapped words can efficiently generate new strongly associated high-order tuples. At the same time, high-frequency and highly relevant new phrases are composed by borrowing the idea of linear programming for an optimal solution. The word overlap degree, lexical rules, and frequency statistical errors were used as constraints.

Similarly is calculated as follows:

If \(A = \left\{ \sum_{m=1}^{n} w_{i,m} \right\}; B = \left\{ \sum_{j=1}^{n} w_{j,p} \right\}; C = \left\{ \sum_{q=1}^{n} w_{k,q} \right\}\)

When \(\text{len}(A) = \text{len}(B) = \text{len}(C) = 3\)

\[
\begin{align*}
A \cap B &= 2 \\
B \cap C &= 2
\end{align*}
\]

If \(A \cap B \cap C = 1\) and \(A \cap B \cap C = \emptyset\), similarly=1, or similarly=0. (3)

\[
\begin{align*}
w_{i,m+1} &= w_{j,p} \\
w_{j,p+1} &= w_{k,q}
\end{align*}
\]

When \(\text{len}(A), \text{len}(B), \text{len}(C)\) is not exactly 3,

\[
\begin{align*}
A \cap B &= 2 \\
B \cap C &= 1
\end{align*}
\]

If \(A \cap B \cap C = \emptyset\), \(A \cap B \cap C = \emptyset\), \(A \cap B \cap C = \emptyset\), similarly=1, or similarly=0. (4)

\[
\begin{align*}
w_{i,m+2} &= w_{j,p} & w_{i,m+1} &= w_{j,p} & w_{i,m+1} &= w_{j,p} \\
w_{j,p+1} &= w_{k,q} & w_{j,p+2} &= w_{k,q} & w_{j,p+1} &= w_{k,q}
\end{align*}
\]
When \( \text{len}(A), \text{len}(B) \) is not exactly the same,

\[
\begin{align*}
\text{If } & A \cap B = 2, \\
& \text{or similarly}=1, \text{ or similarly}=0. \\
\text{max}(w_1, w_2, \ldots, w_n) & \text{ similarly } = 1; \\
\text{s.t. } & \left| \text{Fre}_A - \text{Fre}_B \right| + \left| \text{Fre}_A - \text{Fre}_C \right| + \left| \text{Fre}_B - \text{Fre}_C \right| < 3 \times 0.1;
\end{align*}
\]

where the optimal qualified phrase is solved in formula (6) based on the idea of obtaining the optimal solution through linear programming. The optimal semantic phrase is formed through the overlapping between multiple phrases and the frequency constraint between phrase groups. The similarly represents the similarity between phrases, and it is calculated by similarity formula (3, 4, 5). When similarly is 1, the similarity condition is satisfied. \( \text{Fre}_A, \text{Fre}_B, \) and \( \text{Fre}_C \) are the frequencies of sets A, B, and C.

### 3.3 Co-Word Analysis

According to the amount of text data, co-word analysis can be divided into co-word network analysis and cluster analysis. Co-word network analysis can visually present the relationship network and density of keywords when dealing with a small amount of data (Su et al., 2013). For a large amount of data, the cluster analysis can quickly obtain the classification results. Co-word analysis includes co-word clustering and co-word evolution. The community detection algorithm based on modularity optimization is adopted in co-word clustering (J, 2006; Javed et al., 2018). As one of the network clustering algorithms, it has good accuracy and stability. A collection of nodes with a class of the same characteristics is realized with the community as a module. Suppose that the complex network is undirected, that is, \( G = (V, E) \), \( V \) is the set of nodes in the entire network, and \( E \) is the set of sides in the entire network. Divide network G into N communities, \( P = \{C_1, C_2, \cdots, C_n\} \), and

\[
Q(C) = \frac{1}{2m} \sum_{vw} \left( A_{vw} - \frac{k_v k_w}{2m} \right) + \sum_{i=1}^{c} \left( e_{ii} - a_i^2 \right)
\]

\[
e_{ij} = \sum_{vw} \frac{A_{vw}}{m}
\]

\[
a_i = \frac{k_i}{2m} = \sum_{j} e_{ij}
\]

where \( A_{vw} \) is the sides between nodes v and w. When \( A_{vw} = 1 \), there are sides connected. \( k_v \) is the degree of node v, and m is the number of sides in the entire complex network. \( e_{ij} \) represents sides of the node in the community i and j. \( a_i \) is the degree of node in the community i.

### 3.4 Topic Model

Based on the LDA model, it is improved from the dimension of semantic association, and the Phrase-LDA model (Phrase-Latent Dirichlet Allocation Topic Model) is established. Suppose that there are multiple topics in each article, each word in the article selects a topic with a certain probability. A
word is selected from this topic with a certain probability to mine the underlying semantics of the vocabulary in the corpus, so the probability of each phrase can be expressed as:

\[
p(\text{Ph} | \text{Document}) = \sum_{\text{Topic}} p(\text{Ph} | \text{Topic}) \times p(\text{Topic} | \text{Document})
\] (10)

The words in the LDA model (Figure 3) are replaced with phrases by Phrase-LDA. The model structure is composed of two processes: “document-topic” and “topic-phrase”. In the process of “document-topic”, \( \theta_m \rightarrow z_{m,n} \) is Dirichlet-Multinomial structure, that is, when the m-th document is generated, the document topic is obtained from the document-topic multinomial distribution \( \theta_m \). The topic number of the n-th phrase in the extracted document topic is \( z_{m,n} \). In the “topic-phrase” process, \( \phi_k \rightarrow \text{ph}_{m,n} | k = z_{m,n} \) is also a Dirichlet-Multinomial structure. In the multinomial distribution \( \phi_k \) of topic-phrase in k, the phrase \( \text{ph}_{m,n} \) is obtained from \( k = z_{m,n} \), that is, the n-th phrase in the m-th document in the generated corpus.

The joint probability of the SS-LDA model is shown in formula (11):

\[
P(\text{ph}, z, \pm^2) = P(\text{ph} | z, \pm) P(z | \pm) = \prod_{k=1}^{K} \frac{\Delta(n_k + 2)}{\Delta(2)} \times \prod_{m=1}^{M} \frac{\Delta(n_m + \pm)}{\Delta(\pm)}
\] (11)

The documents and the phrases that make up each document are known, and the unknown variables \( \theta_m \) and \( \phi_k \) are obtained by utilizing the Collapsed Gibbs Sampling (Griffiths & Steyvers, 2004). For the specific phrase t, the unknown variable obtained is:

\[
P(z_i = k | z_{-i}, \text{ph}) \propto P(z_i = k, \text{ph}_i = t | z_{-i}, \text{ph}_{-i}) = \hat{\phi}_{kt}
\] (12)

\[
\hat{\phi}_{kt} = \frac{n_{k-1}^{(t)} + 2 \alpha}{\sum_{t=1}^{V} n_{k-1}^{(t)} + 2 \alpha}
\] (13)

Figure 3. Phrase-LDA model
Formula (13) and (14) are substituted into formula (12) to obtain final Collapsed Gibbs Sampling formula:

\[
P(z_i = k | z_{\neq i}, \text{ph}) \propto \frac{n_{m_{i-i}}^{(k)} + \pm_k}{\sum_{k=1}^{K} n_{m_{i-i}}^{(t)} + \pm_k} * \frac{n_{k_{i-i}}^{(i)} + 2^2}{\sum_{i=1}^{V} n_{k_{i-i}}^{(t)} + 2^2}
\]

4. EMPIRICAL ANALYSIS

4.1 Text Data Source

In this paper, “Outline of the Thirteenth Five-Year Plan for National Economic and Social Development” and “Outline of the Fourteenth Five-Year Plan for National Economic and Social Development” (hereinafter referred to as “Development Plan”) of the central, provincial, municipal and prefecture-level cities are taken as the research object. The selected policy text comes from the central, provincial, municipal and local public text materials. As of May 17, 2021, there are 395 “Thirteenth Five-Year Plan” and 347 “Fourteenth Five-Year Plan” texts, covering the national, provincial, and prefecture levels.

4.2 Phrase Dictionary

The key phrase extraction process is as follows: (1) Build a dictionary of source text phrases. The text is preprocessed with the help of Chinese word segmentation library jieba1, and then the phrase dictionary of the source text is generated according to the phrase dictionary algorithm. (2) Extract source text keywords. The source text phrase dictionary is used for word segmentation, and a co-word matrix file is generated. The high-frequency phrases related to the topic are considered as keywords for co-word analysis and topic classification and are further analyzed in depth. According to the phrase dictionary extraction process, the generation of the phrase dictionary is divided into two key steps.

First, data preprocessing includes text classification and conversion, text preprocessing and content reading. ① Text classification and conversion. According to the title of the document, it is divided into “Thirteenth Five-Year Plan” and “Fourteenth Five-Year Plan”. Part of the PDF document in the original document is converted to word format through Python. Read the contents of the table separately. Some of the original PDF documents are pictures, which need to be recognized and extracted. ② Text preprocessing and content reading. Unify the format of the generated text, such as Enter, space, page number etc. The word segmentation is conducted through jieba to generate unary, binary, and ternary phrases, and remove stop words and punctuation.

Second, the phrase extraction includes word relevance calculation, rule-based filtering, deleting overlapping elements, and merging cross elements. ① The word relevance is calculated by utilizing PMI (Formula 1, 2). The first 90% of binary and ternary phrases are retained and saved to carry out the rule-based filtering step of the rule. ② filter the retained binary and ternary phrases, and delete the binary and ternary phrases that do not satisfy the part-of-speech rules (Table 1). ③ the cross relationship between binary and ternary phrases, and ternary phrases is obtained through formula (6), and merge the qualified phrases to generate new phrases. Read from the table with more phrases, and overlapping test is conducted for the unary and binary phrases, respectively. If the low-element
phrase is completely contained in the higher-element phrase, the lower-element phrase is deleted. Multiple phrases are compared, and they are tested in descending order to iteratively generate a new phrase dictionary.

Table 2 is the extraction results from phrase dictionary in the periods of “Thirteenth Five-Year” and “Fourteenth Five-Year”. With “ecology” as the key word, only 10 groups of phrases are extracted from the end of the dictionary for display. Compared with phrase dictionaries, phrase dictionaries have higher topic description capabilities, clearer semantic expressions, and short-distance contextual relevance, improving the expression accuracy. With the phrase dictionary as the basis of word segmentation, word frequency distribution interval of phrase segmentation is gentle. Comparing the “Thirteenth Five-Year” and “Fourteenth Five-Year” dictionaries, the generated phrase dictionaries are different due to the differences in text data. It shows the portability of the phrase dictionary construction. A phrase dictionary with strong relevance can be obtained according to the specific text content, which provides a more accurate word segmentation basis for subsequent analysis.

4.3 Co-word Analysis

In the analysis of co-word phrases, the modular optimization community discovery algorithm is used for clustering. In the co-word network diagram, it is necessary to distinguish keyword topics and mark the keyword frequency. Therefore, the dot size is used to indicate the keyword frequency, and different colors are used to indicate the theme of the keyword. The webpage files of co-word network diagram in the two periods are 3.31MB and 4.25MB, respectively. The color block diagram represents the whole keyword network, and the topic tables are differentiated and compared. The co-word networks during the 13th Five-Year Plan and 14th Five-Year Plan are shown in Figure 4 and 5, including four categories: green environment, resources and energy, economic development concept, and pollution prevention and control. These four categories are discussed based on the analysis results. Tables 3, 4, 5, and 6 show the changes of key phrases of various topics in the two periods. Co-word analysis is the interpretation of related high-frequency words in the two periods. The 13th Five-Year Plan and 14th Five-Year Plan are two closely related stages, and the plan of one period is continuity as well as development of the other. Therefore, the analysis of key phrases in the two periods is elaborated from two aspects: thematic importance (word frequency of the same phrase) and development (differentiated phrase). Differentiated phrases are marked in italics in Tables 3, 4, 5 and 6, and detailed in the paper. However, the low-frequency words with great differences are not shown in this paper because they are less important.

4.3.1 Green Environment

Green environment is a central cluster of co-word network clustering, including ecology, environmental protection and green. Phrases such as restrictive ecological civilization and restrictive ecological environment during the “Thirteenth Five-Year Plan” period indicate that at this stage, ecological indicators are more restrictive than that during the “Fourteenth Five-Year Plan” period. The binding index is the targets that cannot be breakthrough or must be achieved to achieve the planning goals within the planning period. The watershed ecological protection (1.12) of the “Fourteenth Five-Year Plan” is an important water ecological barrier. The Yangtze River basin and the Yellow River Basin are the two horizontal economic zones in China, which are the key areas for population gathering, agricultural cultivation and industry. Paying attention to the ecological protection of river basins is conducive to promoting coordinated regional development, promoting the harmonious coexistence between people and water, and achieving high-quality development. “Eco-industrialization (0.53)” and “improving the quality of ecosystem (0.39)” emphasize the integrity and systematization of the ecology. Green transformation (1.13), green travel (0.61), and green production and lifestyle (0.78) embody that green has become an important ecological concept of social production and life during the “Fourteenth Five-Year Plan” period. The 13th Five-Year Plan period focuses on improving the ecological environment, so this period sets more binding targets, including “binding ecological
Table 2. Phrase extraction (partial)

<table>
<thead>
<tr>
<th>Phrase Dictionary</th>
<th>Jieba</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Thirteenth Five-Year</strong></td>
<td><strong>Fourteenth Five-Year</strong></td>
</tr>
<tr>
<td>Ecological and Environmental Protection Planning</td>
<td>Ecosystem Protection And Restoration</td>
</tr>
<tr>
<td>Eco-environmental Protection Construction</td>
<td>Ecosystem Protection Effectiveness</td>
</tr>
<tr>
<td>Ecosystem Restoration And Management</td>
<td>Living Eco-Space Layout</td>
</tr>
<tr>
<td>Ecological Civilization System Reform</td>
<td>Continuous Improvement Of Ecological Environment</td>
</tr>
<tr>
<td>Publicity And Education On Ecological Civilization</td>
<td>Protecting The Ecology Of Basic Farmland</td>
</tr>
<tr>
<td>Environmental Governance And Ecological Protection</td>
<td>Implementation Of Ecological Environmental Protection</td>
</tr>
<tr>
<td>Eco-environmental Damage Compensation System</td>
<td>Green Ecological Unit GDP</td>
</tr>
<tr>
<td>Restrictive Ecological Construction Environmental Protection</td>
<td>Green Ecological Forest Cover</td>
</tr>
</tbody>
</table>

Figure 4. “Thirteenth five-year plan” co-word network
civilization”, “binding ecological environment”, “binding resources and environment”, etc. The 14th Five-Year Plan, with the theme of promoting green development, promotes the height of ecological development from a holistic perspective. For example, “eco-industrialization”, “upgrading ecosystem management”, “ecological priority green development”, etc. The construction of ecological civilization has achieved new progress, from quantitative accumulation to qualitative transformation.

4.3.2 Resources and Energy

The ecological environment is closely related to the development and application of energy resources. The increasing demand for energy intensifies ecological damage, constantly exceeding the carrying capacity of the environment, and the shortage of important energy sources will further restrict economic development concept. Therefore, ecological environment and energy constraints should be resolved to realize sustainable economic and social development. These problems can be alleviated by using clean, renewable energy and optimizing the energy structure. Through analyzing the characteristic words of ecological civilization in the two periods, it is found that the Fourteenth Five-Year Plan” has continued the utilization and promotion of clean energy, renewable energy, and new energy during the “Thirteenth Five-Year Plan” period. Compared with the “Thirteenth Five-Year Plan” period, energy internet (0.3) and smart energy (0.64) appeared during the “Fourteenth Five-Year Plan” period. Meanwhile the proportion of primary energy consumption has decreased. The “13th Five-Year Plan” period focuses on new energy and related industries. At the same time, improve resource utilization efficiency to optimize resource allocation. The “proportion of disposable energy consumption” highlights the energy structure optimization to “reduce” and “control” as the main direction. The 14th Five-Year Plan, on the other hand, emphasizes systemic stability. For example, the “Energy Internet” can effectively integrate the industrial chain, and create interactive supply and demand transactions, thus realizing real-time dynamic optimization of energy. “Smart Energy” with
4.3.3 Economic Development Concept

The high input, high energy consumption, and high pollution characteristics presented in the “factor accumulation mode” and “extensive mode of growth” greatly threaten the ecological environment (Lin et al., 2016). Therefore, the high-quality development of the economy is a win-win path to realize the development of economic and ecological civilization construction. During the “Thirteenth Five-Year Plan” period, the concepts of green development (3.18), coordinated development (1.93), innovative

Table 3. Ecological green

<table>
<thead>
<tr>
<th>Fourteenth Five-Year Plan</th>
<th>Thirteenth Five-Year Plan</th>
</tr>
</thead>
<tbody>
<tr>
<td>Phrases</td>
<td>Counts¹</td>
</tr>
<tr>
<td>Ecology</td>
<td>7581</td>
</tr>
<tr>
<td>Ecological Environment</td>
<td>1553</td>
</tr>
<tr>
<td>Ecological Protection</td>
<td>1068</td>
</tr>
<tr>
<td>Ecological Restoration</td>
<td>979</td>
</tr>
<tr>
<td>Ecological Civilization Construction</td>
<td>745</td>
</tr>
<tr>
<td>Eco-environmental Protection</td>
<td>659</td>
</tr>
<tr>
<td>Eco-friendly</td>
<td>620</td>
</tr>
<tr>
<td>Ecological Civilization</td>
<td>585</td>
</tr>
<tr>
<td>Ecological Priority</td>
<td>404</td>
</tr>
<tr>
<td>Ecological Priority Green Development</td>
<td>99</td>
</tr>
<tr>
<td>Ecological Barrier</td>
<td>424</td>
</tr>
<tr>
<td>Key Ecological Function Areas</td>
<td>176</td>
</tr>
<tr>
<td>Eco-industrialization</td>
<td>183</td>
</tr>
<tr>
<td>Improving the quality Of Ecosystem</td>
<td>134</td>
</tr>
<tr>
<td>Watershed Ecological Protection</td>
<td>391</td>
</tr>
<tr>
<td>Environmental Protection</td>
<td>1092</td>
</tr>
<tr>
<td>Energy Conservation And Environmental Protection</td>
<td>576</td>
</tr>
<tr>
<td>Energy Conservation And Environmental Protection Industry</td>
<td>156</td>
</tr>
<tr>
<td>Energy Conservation And Emissions Reduction</td>
<td>226</td>
</tr>
<tr>
<td>Environmental Protection</td>
<td>499</td>
</tr>
<tr>
<td>Environmental Governance</td>
<td>424</td>
</tr>
<tr>
<td>Green</td>
<td>4637</td>
</tr>
<tr>
<td>Green Development</td>
<td>1162</td>
</tr>
<tr>
<td>Green Transformation</td>
<td>392</td>
</tr>
<tr>
<td>Green Travel</td>
<td>212</td>
</tr>
<tr>
<td>Green Production And Lifestyle</td>
<td>271</td>
</tr>
<tr>
<td>Smart Green</td>
<td>174</td>
</tr>
<tr>
<td>Significant Progress Has Been Made In The Green Transformation</td>
<td>114</td>
</tr>
</tbody>
</table>

Notes: 1. Counts: The number of occurrences of the phrase in the corpus. Same as below. 2. Frequency: The ratio of frequency to the number of texts in the corpus. Same as below.
development (3.64), and innovation-driven development (1.12) were put forward, promoting the ecological civilization construction that focuses on improving environmental quality. The report of the 19th CPC National Congress proposed that “accelerate the development of a green and low-carbon circular economic development concept system”, and it guides the high-quality development (9.33) of the “Fourteenth Five-Year Plan”. The development of “high-level opening up (0.74)” and “dual circulation of domestic economic cycle and international economic cycle (1.03)” deepen and expand win-win cooperation with the outside world. The “14th FYP” and the “13th FYP” follow the same development line. However, the development model has changed from a single industry development (real economy) to a comprehensive “green and high-quality development”, and more esteem for the “sharing economy” and “balanced development” model. “Domestic and international dual cycle” is a positive exploration of the current international situation. So as to realize the economic circulation among various regions in China, China and other countries.

4.3.4 Pollution Prevention and Control

Pollutants are caused by the social and economic development concept. Even if the resources and energy are used efficiently, the pollutants cannot be avoided. Therefore, the control and prevention of

<table>
<thead>
<tr>
<th>Table 4. Resources and energy</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Fourteenth Five-Year Plan</strong></td>
</tr>
<tr>
<td><strong>Phrases</strong></td>
</tr>
<tr>
<td>Energy</td>
</tr>
<tr>
<td>New Energy</td>
</tr>
<tr>
<td>New Energy Cars</td>
</tr>
<tr>
<td>Clean Energy</td>
</tr>
<tr>
<td>Renewable Energy</td>
</tr>
<tr>
<td>Total Energy Consumption</td>
</tr>
<tr>
<td>New Energy Automotive Industry</td>
</tr>
<tr>
<td>New Energy Industries</td>
</tr>
<tr>
<td>Reduced Energy Consumption Per Unit Of Gross Regional Product</td>
</tr>
<tr>
<td>Lower Energy Consumption Per Unit Of GDP</td>
</tr>
<tr>
<td>Energy Consumption</td>
</tr>
<tr>
<td>Energy Internet</td>
</tr>
<tr>
<td>Primary Energy Consumption Share</td>
</tr>
<tr>
<td>Smart Energy</td>
</tr>
<tr>
<td>Wind Energy</td>
</tr>
<tr>
<td>Photovoltaic Power Generation</td>
</tr>
<tr>
<td>Wind Power Generation</td>
</tr>
<tr>
<td>Biomass Power Generation</td>
</tr>
<tr>
<td>Solar Power Generation</td>
</tr>
<tr>
<td>Mineral Resources</td>
</tr>
<tr>
<td>Saving Resources</td>
</tr>
<tr>
<td>Comprehensive Utilization Of Resources</td>
</tr>
<tr>
<td>Resource Exploitation</td>
</tr>
<tr>
<td>Resource Consumption</td>
</tr>
<tr>
<td>Resources Recycling</td>
</tr>
<tr>
<td>Resource Utilization Of Waste</td>
</tr>
<tr>
<td>Recycling Of Renewable Resources</td>
</tr>
</tbody>
</table>
pollutants is still an important part of the ecological civilization construction. The pollution problems faced during the “Fourteenth Five-Year” and “Thirteenth Five-Year” periods are mainly water pollution and air pollution. For disaster problems, there are natural disasters, geological disasters, and

Table 5. Economic development concept

<table>
<thead>
<tr>
<th>Fourteenth Five-Year Plan</th>
<th>Thirteenth Five-Year Plan</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Phrases</strong></td>
<td><strong>Counts</strong></td>
</tr>
<tr>
<td>Green Development</td>
<td>1162</td>
</tr>
<tr>
<td>Synergistic Development</td>
<td>1133</td>
</tr>
<tr>
<td>Innovation Development</td>
<td>1078</td>
</tr>
<tr>
<td>Coordinate Development</td>
<td>1044</td>
</tr>
<tr>
<td>Innovation-Driven Development</td>
<td>334</td>
</tr>
<tr>
<td>High-quality Development</td>
<td>3247</td>
</tr>
<tr>
<td>Promoting High Quality Development</td>
<td>728</td>
</tr>
<tr>
<td>High-quality Development Of The Economy</td>
<td>346</td>
</tr>
<tr>
<td>High-quality Development Of Industry</td>
<td>160</td>
</tr>
<tr>
<td>High-quality Development Of manufacturing Industry</td>
<td>158</td>
</tr>
<tr>
<td>High-quality Development Of Service Industry</td>
<td>49</td>
</tr>
<tr>
<td>Green High-Quality Development</td>
<td>63</td>
</tr>
<tr>
<td>Circular Economy</td>
<td>296</td>
</tr>
<tr>
<td>Dual Circulation Of domestic Economic Cycle And International Economic Cycle</td>
<td>360</td>
</tr>
<tr>
<td>Recycling Development</td>
<td>82</td>
</tr>
<tr>
<td>Vigorous Development Of Circular Economy</td>
<td>85</td>
</tr>
<tr>
<td>Opening Up Development</td>
<td>260</td>
</tr>
<tr>
<td>High-level Of Opening up</td>
<td>256</td>
</tr>
<tr>
<td>High-level Of Open economy</td>
<td>163</td>
</tr>
<tr>
<td>High-level Of Opening To The Outside World</td>
<td>150</td>
</tr>
<tr>
<td>Digital Economy</td>
<td>1221</td>
</tr>
<tr>
<td>Digital Transformation</td>
<td>539</td>
</tr>
<tr>
<td>Digitalization Of Industry</td>
<td>332</td>
</tr>
<tr>
<td>Real Economy</td>
<td>683</td>
</tr>
<tr>
<td>Supporting The Real Economy</td>
<td>104</td>
</tr>
<tr>
<td>The Focus Of Economic Development Is On The Real Economy</td>
<td>100</td>
</tr>
<tr>
<td>Serving the Real Economy</td>
<td>88</td>
</tr>
<tr>
<td>Financial Services For The Real Economy</td>
<td>59</td>
</tr>
<tr>
<td>Real Economy Integration Development</td>
<td>52</td>
</tr>
<tr>
<td>Balanced Development</td>
<td>363</td>
</tr>
<tr>
<td>Sharing Economy</td>
<td>246</td>
</tr>
<tr>
<td>Business Environment</td>
<td>1101</td>
</tr>
<tr>
<td>Optimizing The Business Environment</td>
<td>312</td>
</tr>
<tr>
<td>Green Finance</td>
<td>209</td>
</tr>
</tbody>
</table>
meteorological disasters. Thus, the pollution and disaster control is a long game. The problems faced in the two periods remain unchanged, and the so does the prevention and control policies adopted. During the “Fourteenth Five-Year Plan” period, “joint prevention and control” was emphasized,
indicating that the coordinated strategy adopted during the “Thirteenth Five-Year Plan” period has achieved good results. The joint prevention and control mechanism has strengthened the new path of ecological environment prevention and control of regional linkage, coordination and cooperation, and mutual benefit. The “carbon neutrality (0.75) and peak carbon dioxide emissions (0.51)” targets of the “Fourteenth Five-Year Plan” will become one of the important tasks in this period and the next stage, which puts forward new goals and requirements for atmospheric environmental governance in China. The “carbon” task of the development plan will play a positive gravitational role for ecological and environmental protection departments and local governments in formulating policies, decomposing targets, and supervising process. The carbon neutrality and peak carbon dioxide emission goals are closely related to economic growth and are important ways for high-quality and green development. They actively respond to ecological and environmental protection requirements, and promote the continuous optimization of energy and resource consumption structure and industrial structure.

4.4 Topic Classification

According to the constructed phrase topic model, the open source gensim package in python is used to default hyper-parameters $\alpha = 0.0001, \beta = 0.00002$. Through perplexity calculation value, topic interpretability and topic distribution (Figure 6), it is determined that the final topic number k is 7, the maximum number of iterations is 1000. The topic sets are shown in Table 7.

The top 30 key-phrases of topic are highly relevant to the topic, so the topic set in Table 7 only shows the top 10 phrases and their frequencies that are highly relevant to the topic. From the categories of the topic set, it can be found that the topic of the development plans in the two periods are relatively similar, covering overall development, industrial layout, green ecology, living standard, innovation and entrepreneurship, etc. However, there are great differences in topic feature words.

T1: the “Thirteenth Five-Year Plan” focuses on industrial development, including the layout of modern industries (1.49), clusters (4.57), and informatization (1.02). From the perspective of the overall high-quality development of the economy and society, the “Fourteenth Five-Year Plan” emphasizes the digital economy (3.43), the real economy (1.87) and structural reforms (2.78).

T2: the topic of feature words the “Thirteenth 13th Five-Year” is employment and entrepreneurship (1.16), which emphasizes basic employment stability, skills, occupational employment and services. The topic of the “Fourteenth Five-Year” feature words is innovation and entrepreneurship and employment, and the “Fourteenth Five-Year” plan focuses on higher education (0.98), high-level talents (1.39), innovation and entrepreneurship (2.59), which is in line with the high-quality development of human capital and employment and entrepreneurship.

T3: the “Thirteenth Five-Year Plan” focuses on ecological protection and restoration to achieve ecological civilization construction. In the “Fourteenth Five-Year Plan”, clean energy (2.09), carbon sinks (0.59) and carbon neutrality (0.75) are incorporated in the ecological civilization construction.

T4: The feature words of the development strategy in the “Thirteenth Five-Year” include economic and social development (8.02), building a moderately prosperous society (5.78), innovation, green, and coordinated development (2.85, 2.70, 4.05). The topic feature words of the “Fourteenth Five-Year Plan” cluster focus on high-quality, ecological priority development (9.32, 1.16) strategy. The relationship between man and nature and the sustainability of development are highly valued at the new stage.

T5: Internet develops rapidly during the “Thirteenth Five-Year Plan” period, so the development and integration of Internet (4.73) receives greater attention. The topic feature words of the “Fourteenth Five-Year Plan” are innovation and development. The strategic emerging industries based on informatization and Internet have become the mainstay of innovation and development.

T6 and T7 of the “Thirteenth Five-Year Plan” are social security and improving people’s livelihood. The overall development strategy in China always focuses on the “people-oriented”, which clarifies the purpose of development, and emphasizes the essence of development. The T6 of the “Fourteenth Five-Year Plan” is opening up. In the face of the complex and changing development environment and the achievements of the “Belt and Road” (5.65), it emphasizes the diversity of
Table 6. Pollution-disaster-prevention

<table>
<thead>
<tr>
<th>Fourteenth Five-Year Plan</th>
<th>Counts</th>
<th>Frequency</th>
<th>Thirteen Five-Year Plan</th>
<th>Counts</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pollution</td>
<td>839</td>
<td>2.41</td>
<td>Pollution</td>
<td>1581</td>
<td>4.01</td>
</tr>
<tr>
<td>Sewage Treatment Plant</td>
<td>597</td>
<td>1.72</td>
<td>Sewage Treatment Plant</td>
<td>1032</td>
<td>2.62</td>
</tr>
<tr>
<td>Sewage</td>
<td>574</td>
<td>1.65</td>
<td>Sewagez</td>
<td>789</td>
<td>2.00</td>
</tr>
<tr>
<td>Pollution Prevention And Control</td>
<td>555</td>
<td>1.59</td>
<td>Pollution Prevention And Control</td>
<td>759</td>
<td>1.93</td>
</tr>
<tr>
<td>Pollution Control</td>
<td>463</td>
<td>1.33</td>
<td>Pollution Control</td>
<td>591</td>
<td>1.50</td>
</tr>
<tr>
<td>Sewage Treatment</td>
<td>445</td>
<td>1.28</td>
<td>Sewage Treatment</td>
<td>827</td>
<td>2.10</td>
</tr>
<tr>
<td>Air Pollution Prevention And Control</td>
<td>216</td>
<td>0.62</td>
<td>Air Pollution Prevention And Control</td>
<td>474</td>
<td>1.20</td>
</tr>
<tr>
<td>Water Pollution Prevention And Control</td>
<td>192</td>
<td>0.55</td>
<td>Water Pollution Prevention And Control</td>
<td>415</td>
<td>1.05</td>
</tr>
<tr>
<td>Soil Pollution Prevention And Control</td>
<td>181</td>
<td>0.52</td>
<td>Soil Pollution Prevention And Control</td>
<td>156</td>
<td>0.40</td>
</tr>
<tr>
<td>Joint Defense And Control</td>
<td>249</td>
<td>0.72</td>
<td>Joint Defense And Control</td>
<td>137</td>
<td>0.35</td>
</tr>
<tr>
<td>Joint Prevention And Control Mechanism</td>
<td>96</td>
<td>0.28</td>
<td>Joint Prevention And Control Mechanism</td>
<td>57</td>
<td>0.14</td>
</tr>
<tr>
<td>Joint Prevention And Control Of Atmospheric Pollution</td>
<td>57</td>
<td>0.16</td>
<td>Joint Prevention And Control Of Atmospheric Pollution</td>
<td>60</td>
<td>0.15</td>
</tr>
<tr>
<td>Carbon Sink</td>
<td>206</td>
<td>0.59</td>
<td>Carbon Sink</td>
<td>354</td>
<td>0.90</td>
</tr>
<tr>
<td>Reduced CO2 Emissions Per Unit Of GDP</td>
<td>94</td>
<td>0.27</td>
<td>Reduced CO2 Emissions Per Unit Of GDP</td>
<td>177</td>
<td>0.45</td>
</tr>
<tr>
<td>Gross Regional Product Per Unit Reduced Carbon Dioxide Emissions</td>
<td>122</td>
<td>0.35</td>
<td>Gross Regional Product Per Unit Reduced Carbon Dioxide Emissions</td>
<td>124</td>
<td>0.31</td>
</tr>
<tr>
<td>Carbon Dioxide Emissions</td>
<td>78</td>
<td>0.22</td>
<td>Carbon Dioxide Emissions</td>
<td>118</td>
<td>0.30</td>
</tr>
<tr>
<td>Carbon Cycle Development</td>
<td>120</td>
<td>0.34</td>
<td>Carbon Cycle Development</td>
<td>218</td>
<td>0.55</td>
</tr>
<tr>
<td>Carbon Neutrality</td>
<td>262</td>
<td>0.75</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reach The CO2 Emission Peak</td>
<td>179</td>
<td>0.51</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Disaster</td>
<td>474</td>
<td>1.36</td>
<td>Disaster</td>
<td>537</td>
<td>1.36</td>
</tr>
<tr>
<td>Natural Disaster</td>
<td>282</td>
<td>0.81</td>
<td>Natural Disaster</td>
<td>399</td>
<td>1.01</td>
</tr>
<tr>
<td>Geologic Disaster</td>
<td>228</td>
<td>0.66</td>
<td>Geologic Disaster</td>
<td>348</td>
<td>0.88</td>
</tr>
<tr>
<td>Meteorological Disasters</td>
<td>96</td>
<td>0.28</td>
<td>Meteorological Disasters</td>
<td>203</td>
<td>0.52</td>
</tr>
<tr>
<td>Fire Disaster</td>
<td>129</td>
<td>0.37</td>
<td>Fire Disaster</td>
<td>133</td>
<td>0.34</td>
</tr>
<tr>
<td>Mountain Flood Prevention And Control</td>
<td>82</td>
<td>0.24</td>
<td>Mountain Flood Prevention And Control</td>
<td>122</td>
<td>0.31</td>
</tr>
<tr>
<td>Disaster Prevention And Mitigation</td>
<td>389</td>
<td>1.12</td>
<td>Disaster Prevention And Mitigation</td>
<td>588</td>
<td>1.49</td>
</tr>
<tr>
<td>Reserve Of Disaster Relief Materials</td>
<td>82</td>
<td>0.24</td>
<td>Reserve Of Disaster Relief Materials</td>
<td>150</td>
<td>0.38</td>
</tr>
<tr>
<td>Level Of Natural Disaster Preparedness</td>
<td>98</td>
<td>0.28</td>
<td>Disaster Prevention And Mitigation System</td>
<td>150</td>
<td>0.38</td>
</tr>
<tr>
<td>Disaster Monitoring And Early Warning</td>
<td>68</td>
<td>0.20</td>
<td>Disaster Monitoring And Early Warning</td>
<td>81</td>
<td>0.21</td>
</tr>
</tbody>
</table>

cooporative development modes. T7: agriculture and rural development. The poverty alleviation during the Thirteenth Five-Year Plan period has been transformed into the comprehensive rural vitalization (5.57), and the depth and direction of development have changed greatly.

5. CONCLUSIONS AND PROSPECT

Co-words and topic analysis play a very important role in interpreting text content and text classification. It is one of the most widely used methods in text mining. In this paper, combined the co-words and topic, 742 text data from the “Outline for National Economic and Social Development
Plan” are utilized to understand the changing trends of development strategy goals of China in two different periods. The following conclusions are obtained:

(1) In the “Thirteenth Five-Year Plan” text, the institutional ecological security barriers are important task in the ecological environment planning. In the Fourteenth Five-Year Plan, safety barrier, control system, and supervising related ecological environmental protection system are implemented. Among them, the feature words of the “Thirteenth Five-Year Plan” contain 13 binding phrases, such as restrictive indices (1.66), restrictive ecological civilization (1.19), restrictive resource environment (0.63), restrictive ecological environment (0.4) etc. In the “Fourteenth Five-Year Plan”, there are only 2 binding phrases, and the frequency of “binding index” is only 0.68. The red line (1.67) in the “Thirteenth Five-Year” means to delineate ecological red line, and formulate the ecological indicators. The red line (0.66) in the “Fourteenth Five-Year Plan” focus on ensuring that the red line in ecological protection in not crossed and implementing the ecological redline to implement and supervise the ecological protection indicators of the previous stage.

(2) Compared with the “Thirteenth Five-Year Plan”, the text of the “Fourteenth Five-Year Plan” emphasizes the integrity and systematization of ecological civilization construction. The “Fourteenth Five-Year Plan” emphasizes high-quality development in the overall social and economic development concept, adheres to the ecological priority, and holds that lucid waters and lush mountains are invaluable assets. In the topic model, from the overall social development, T4 of the “Fourteenth Five-Year” focuses on high-quality development with ecological protection and green development as the center, and integrates development and ecological environmental protection to promote the sustainable development. The T4 of the “Thirteenth Five-Year” development strategy still focuses on economic and social development and building a well-off society, and it only emphasizes the green concept in the process of partial development and construction. The construction of ecological civilization in the “Fourteenth Five-Year Plan” takes “governance system (4.43)”, “modernization of governance ability (2.34)” and “system governance (1.34)” as new missions. Compared with the higher-weighted “Environmental renovation (1.41)”, “Environmental quality (1.66)” and “comprehensive environmental renovation
Table 7. Topic set

<table>
<thead>
<tr>
<th>Thirteenth Five-Year Plan</th>
<th>Fourteenth Five-Year Plan</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>T1 Industrial Development</strong></td>
<td><strong>T1 Development And Construction</strong></td>
</tr>
<tr>
<td>Topic Feature Phrase: Industry Clusters(1802), Equipment Manufacture(1235), Demonstration Base(134), Strategic Emerging Industries(1176), Energy Conservation And Environmental Protection(1211), New Energy Cars(1040), Level Of Informatization(306), Industry Upgrade(465), Productive Service Industry(736), Industrial Layout(587)</td>
<td>Topic Feature Phrase: Development Pattern(2789), Industrial Cluster(2207), Digital Economy(1193), Integration Development(2041), Advanced Manufacturing Industry(859), Strategic Emerging Industries(785), Real Economy(652), Smart Manufacturing(967), High-quality Development(3245), Structural Reform(967)</td>
</tr>
<tr>
<td>Ecological Civilization Feature Phrase: Strategic Emerging Industries(1176), Energy Conservation And Environmental Protection(1211), New Energy Car(1040), Information Services(912), Level Of Informatization(506)</td>
<td>Ecological Civilization Feature Phrase: High-quality Development(3245), Strategic Emerging Industries(785), Tangible Cultural Heritage(437), Development Quality(440)</td>
</tr>
<tr>
<td><strong>T2 Employment and Entrepreneurship</strong></td>
<td><strong>T2 Innovation And Entrepreneurship</strong></td>
</tr>
<tr>
<td>Senior High School Education Universal(629), Vocational Skills Training(396), Practical Training Base(113), Employment And Entrepreneurship(457), Career Development(376), Ability To work(273), National Education(139), Entrepreneurship Promotes Employment(269), Professional Technical Personnel(291), Employment Services(256)</td>
<td>Eco-friendly(611), Energy Saving And Environmental Protection(527), Innovation And Entrepreneurship(903), Science And Technological Innovation(1655), College graduates(335), Fundamental Research(516), Employment And Entrepreneurship(515), High-level Talent(484), New Research And Development Organization(447), Entrepreneurship-led Employment(229)</td>
</tr>
<tr>
<td>Circular Economy(1188), Harmonious Development(186)</td>
<td>Eco-friendly(611), Energy Saving And Environmental Protection(527), Sewerage Pipe Network(310), Disaster prevention And Mitigation(389), Value Of Ecological Products(277)</td>
</tr>
<tr>
<td><strong>T3 Ecological Protection</strong></td>
<td><strong>T3 Ecological Environment</strong></td>
</tr>
<tr>
<td>Eco-environmental Protection(791), Ecological Construction(994), Energy Conservation And Emissions Reduction(1176), Key Ecological Function Areas(496), Forest Coverage(611), Tangible Cultural Heritage(641), Ecological Restoration(662), Wetland Protection(347), Ecological Space(510), Pollution Prevention And Control(526)</td>
<td>Clean Energy(728), Pollution Prevention And Control(487), Territorial Space Planning(446), Forest Coverage(315), Household Waste Classification(364), Eco-environmental Quality(278), Biodiversity Conservation(261), Carbon Sink(206), Pollution Treatment(384), Carbon Neutrality(261)</td>
</tr>
<tr>
<td>Ecological Protection(1638), Cultural Tourism(1430), Eco-environmental Protection(791), Energy Conservation And Emissions Reduction(1176), Key Ecological Function Areas(496), Forest Coverage(611), Tangible Cultural Heritage(641), Ecological Restoration(662), Wetland Protection(347), Ecological Space(510), Water Containment(335), Pollution Prevention And Control(759), Wetland Park(693)</td>
<td>Clean Energy(728), Pollution Prevention And Control(487), Household Waste Classification(364), Forest Coverage(315), Pollution Prevention And Control(484), Eco-environmental Governance(208), Biodiversity Conservation(261), Smart Tourism(315), Tourist Attractions(731), Rural Tourism(730), Territorial Spatial Planning(444), Carbon Sink(206), Carbon Neutrality(261)</td>
</tr>
<tr>
<td><strong>T4 Social Development Strategy</strong></td>
<td><strong>T4 Social Security</strong></td>
</tr>
<tr>
<td>Economic And Social Development(3159), Building A Moderately Prosperous Society(2276), Ecological Civilization Construction(299), Open Development(484), Scientific Development(675), Innovative Development(1124), Green Development(1065), Coordinated Development(1597), Scientific Development Of View(478), Structural Reform(774)</td>
<td>Support For Development(214), Greening Projects(186), New Employment(224), Balanced Population Development(255), Urban And Rural Low Income Insurance(152), Social Security Pocket(111), Medical Insurance System(102), Health Industry(346), Drug System(253), Cultural And Creative Industrial Park(262)</td>
</tr>
<tr>
<td>Ecological Civilization Construction(1299), Ecological And Environmental Protection(651), Green Development(1065), Ecological Civilization(1782)</td>
<td>Sewage Treatment Plant(1032), Disaster Prevention and Mitigation(588), Binding Indicators(653), Environmental Renovation(555), Scenic Spots(378)</td>
</tr>
<tr>
<td><strong>T5 Internet Integration Development</strong></td>
<td><strong>T5 Well-being Of The People’s</strong></td>
</tr>
<tr>
<td>Internet Plus(1865), Logistics Center(622), Entrepreneurship And Innovation(641), Adjustment of Industrial Restructuring(402), Service Platform(1240), School-Enterprise Cooperation(388), Enabling Strength(401), Service System Construction(371), Human Resource Market(336), Informatization Construction(679)</td>
<td>Improving People’s Livelihood(823), Shanty Town Rehabilitation(668), Sheltered Housing(587), Ecological Function Area(270), Equalization Of Public Services(950), High School Education(272), Rural Drinking Water(304), Public Rental Housing(238), Leisure Vacation(424)</td>
</tr>
<tr>
<td>Tourist Attractions(917), Travel Products(430), Clean Energy(1161)</td>
<td>Leisure Travel(692), Resource Advantages(668), Eco-tourism(789), Saving Resources(596), Photovoltaic Power Generation(44)</td>
</tr>
<tr>
<td><strong>T6 Social Security</strong></td>
<td><strong>T7 Well-being Of The People’s</strong></td>
</tr>
<tr>
<td>Support For Development(214), Greening Projects(186), New Employment(224), Balanced Population Development(255), Urban And Rural Low Income Insurance(152), Social Security Pocket(111), Medical Insurance System(102), Health Industry(346), Drug System(253), Cultural And Creative Industrial Park(262)</td>
<td>Improving People’s Livelihood(823), Shanty Town Rehabilitation(668), Sheltered Housing(587), Ecological Function Area(270), Equalization Of Public Services(950), High School Education(272), Rural Drinking Water(304), Public Rental Housing(238), Leisure Vacation(424)</td>
</tr>
<tr>
<td>Sewage Treatment Plant(1032), Disaster Prevention and Mitigation(588), Binding Indicators(653), Environmental Renovation(555), Scenic Spots(378)</td>
<td>Leisure Travel(692), Resource Advantages(668), Eco-tourism(789), Saving Resources(596), Photovoltaic Power Generation(44)</td>
</tr>
<tr>
<td><strong>T7 Well-being Of The People’s</strong></td>
<td><strong>Ecological Civilization Feature Phrase</strong></td>
</tr>
<tr>
<td><strong>T8 Ecological Environment</strong></td>
<td><strong>Economic And Environmental Protection</strong></td>
</tr>
<tr>
<td><strong>T9 Economic And Environmental Protection</strong></td>
<td><strong>Ecological Civilization</strong></td>
</tr>
<tr>
<td><strong>T10 Ecological Civilization</strong></td>
<td><strong>Circular Economy</strong></td>
</tr>
<tr>
<td><strong>T11 Circular Economy</strong></td>
<td><strong>Energy Saving And Environmental Protection</strong></td>
</tr>
<tr>
<td><strong>T12 Environment</strong></td>
<td><strong>New Energy Cars</strong></td>
</tr>
<tr>
<td><strong>T13 Ecological Environment</strong></td>
<td><strong>Industrial Cluster</strong></td>
</tr>
</tbody>
</table>
(1.44)” in the “Thirteenth Five-Year Plan” for improving the ecological environment, it is more systematic.

(3) The ecological civilization construction plan in the “Fourteenth Five-Year” has higher initiative and scientific rationality. “Carbon neutrality (0.75) and peak carbon dioxide emissions (0.51)” are climate commitments from China for a period of time in the future, and they have become important and clear ecological and environmental protection tasks in the “Fourteenth Five-Year” development plan, reflecting China’s initiative and responsibility in the ecological civilization construction. The targets of carbon reduction (0.27), carbon cycle (0.55), and reduction of carbon dioxide emissions per unit of GDP (0.45) in the “Thirteenth Five-Year Plan” promote the carbon target development. The utilization of energy resources in the “Thirteenth Five-Year Plan” aimed at energy conservation (5.06), energy conservation and emission reduction (2.98), and energy conservation and environmental protection (3.65) has been transformed into the green transformation (1.13), smart energy (0.64), and energy Internet (0.30) during the “Fourteenth Five-Year” period, improving the utilization of resources and energy more accurately and scientifically.

There are similar regional plans or development strategies in other countries and regions. For example, the development strategy of intelligent manufacturing, Germany’s “Industry 4.0”, Japan’s “Society 5.0” and the United States’ “Industrial Internet”. Mid- and long-term plans, such as the EU’s “Digital Compass 2030: The Road to a European Digital Decade”, the U.S. “Spatial Strategy 2050”, and the UK’s “Regional Economic Development Planning System”. With text mining, it is possible to explore the development priorities and changing trends of these plans, which not only provides a new way of thinking for government public management, but also expands the application of these methods. Through mining the themes with key words and clusters from the five-year plan texts or other strategic documents, combined with the theories of ecological civilization construction, spatial data analysis and industrial development, our research will be important in the following three aspects: (1) Studying the historical stage of ecological civilization construction according to

<table>
<thead>
<tr>
<th>Thirteenth Five-Year Plan</th>
</tr>
</thead>
<tbody>
<tr>
<td>T4 High-quality Development</td>
</tr>
<tr>
<td>T5 Innovative Development</td>
</tr>
<tr>
<td>T6 Open To The Public</td>
</tr>
<tr>
<td>T7 Rural Revitalization</td>
</tr>
</tbody>
</table>

Ecological function(224), Ecological Tourism(471), Ecological Industry(201)
the evolution of high-frequency words, and forecasting the development of the situation can provide suggestions for policy optimization and implementation. (2) Analyzing regional differences according to regional high-frequency word forms and clustering differences can provide theoretical basis for regional development optimization, regional characteristic policy designation and implementation. (3) According to the co-word association between the country and a region and between one region and another, we can investigate the role of a region in the national development and the coordinated development strategy among regions.
REFERENCES


ENDNOTES

1 Jieba is a Chinese word segmentation, which combines two types of methods: rules and statistics. It supports simplified and traditional Chinese word segmentation. This word segmentation tool comes with a dictionary containing more than 20,000 words. Jieba supports three-word segmentation methods: exact mode, full mode, and search engine mode.

2 The numbers in parentheses are phrase frequencies.