Topics Mining of Domestic Green Development Research Based on System Clustering

Yan Wu[1], Changlu Zhang[2]
e-mail: 704660852@qq.com, 756865684@qq.com

[1]Qingdao Metro Operation Co., Ltd Qingdao, CHN

[2]School of Economics and Management, Beijing Information Science and Technology University Beijing, CHN

Abstract—In order to systematically sort out the relevant research results in the field of green development in China and to promote further research, the literature related to the green development topic in CNKI was selected. Cluster analysis and topic mining methods were adopted to identify the main research topics. The chi-square statistics method was used to extract keywords with a high correlation degree to condense the theme connotation. The results showed that domestic research on green development can be divided into 11 categories, concluding the construction of ecological civilization, ecological governance, sustainable development, innovation-driven development, environmental protection, green economy, two-mountains theory, etc. Among them, literature related to ecological civilization construction and ecological governance has the largest number, which belongs to the traditional research field. Although there was not much literature on the topic of two-mountains theory, innovation-driven development, green technology innovation, green economy research, etc., the research heat has increased significantly in recent years, which represents the research trend in the future.

Keywords- green development; topic mining; trend analysis; system clustering; Chi-square statistics

1 INTRODUCTION

In China's "12th Five-Year Plan", "green development, building a resource-saving and environment-friendly society" was listed as an important chapter for the first time. In the recommendations of the "13th Five-Year Plan" issued in November 2015, "green development" has become a key issue. General Secretary Xi Jinping praised the development thought of "lucid waters and lush mountains are invaluable assets". During the same period, China's practical and academic circles focused on issues such as: what is green development, how to achieve green development, and how to evaluate the level of green development. The explorations in terms of green development increase day by day.

In order to systematically sort out the relevant research results in the field of green development in China and to promote further research, we selected literature related to the green development topic in CNKI. Based on the literature data, 11 research topics were extracted. Furthermore, the research trend in the future was analyzed.

ICEMME 2022, November 18-20, Nanjing, People's Republic of China

DOI 10.4108/eai.18-11-2022.2326771

2 LITERATURE REVIEW

Regarding the formation process of green development, Wang Y. J. and et al.[1] showed that the period from 1949 to 1977 was the embryonic period, the period from 1978 to 2011 was the period of formation, and the period from 2012 to 2019 was the mature period. Based on the historical perspective, Wu X. X. et al.[2] systematically sorted out the relevant content from the "10th Five-Year Plan" to the "13th Five-Year Plan". The above research only analyzed the evolution of the concept of green development. However, the systematic research on green development was inadequate.

Regarding the connotation of green development, Zhu D. B.[3] carried out the concept of green development in terms of overall connotation, values, realization goals, institutional basis, practice paths, and fundamental guarantees. The value of green development was analyzed from multiple angles. Mao H. B.[4] clarified the concept of green development from the dimension of accelerating the transformation of economic development mode, promoting green technological innovation, creating a green economic and social development system, advocating low-carbon lifestyles and consumption patterns. Xu M. H.[5] systematically combed the five dimensions of the concept of green development including theoretical, practical, institutional, people-oriented, and universal. Wu J.[6] discussed green development from the three dimensions of the logical starting point of harmonious coexistence between man and nature, the ultimate goal of improving people's livelihood, and the practical measures of system construction. Liu J. [7] pointed out the specific path of green development. From the perspective of system characteristics, green development required the improvement of organic integrity, three-dimensional coordination and intergenerational sustainability. From the perspective of value orientation, green development should be followed by nature first, people first and harmonious symbiosis. From the perspective of the practical path, green development should be implemented into land green planning, social green transformation. Liu D. H.[8] expounded the rich connotation, theoretical value and logical relationship of the concept of green development, and put forward the path of integrating the concept of green development into the whole process of economic construction, political construction, cultural construction and social construction. Zhang J. W.[9] pointed out that Marxist scientific thought on the relationship of "man and nature" was an important philosophical basis for the concept of green development. Correctly grasping and handling the relationship between socialism and capital, implementing economic development with priority to ecological and environmental protection, developing green science and technology and promoting a green lifestyle in the whole society were the realization paths of green development. Zeng F. Y.[10] pointed out that the realization of green development required the effective support of green policy system and green supervision system, as well as the effective operation of the green social environment composed of government, enterprises and the public. Zhang Q. Y.[11] pointed out that in essence, green development was a new choice made by people, which adhered to the unity of material scale and human scale, and followed the two basic principles of value and regularity. The above research mainly focused on the scientific connotation and construction path of green development with qualitative analysis. The scientific combing using the quantitative method was in lack. At the same time, the above studies mostly focused on a certain dimension of green development depending on the experience and views of the researchers themselves, which was not systematic and comprehensive.

As for the research content of green development, Qin X. L.[12] combed the relevant research on

green development at home and abroad and found that foreign empirical research on green development focused on the discussion of the green energy economy, and there were still disputes on many aspects of green development and more generally sustainable development while the domestic related research mainly focused on connotation, strategy, path, policy and strategy. Si L. J.[13] selected research literature related to green development and used CiteSpace software to analyze the visual knowledge map of the core authors, institutions and their cooperation in green development related research, as well as the research hot spots in Chinese and foreign languages.

To sum up, the existing research was devoted to the evolution of green development, as well as the qualitative analysis of its connotation and construction path. The research on domestic green development from the perspective of the quantitative and systematic review was in lack. So, it was difficult to provide an effective reference for further research and scientific construction practice. Therefore, in our research, the research literature related to green development was collected, and the cluster analysis was conducted on literature data, identifying the main research topics. Then we quantitatively described the connotation characteristics and development trends of each research topic to provide further reference on green development.

3 RESEARCH TOPIC OF DOMESTIC GREEN DEVELOPMENT BASED ON CLUSTER ANALYSIS

3.1 Data collection and preprocessing

Keywords were very important to the research paper, which embodied the essence of the article and represented the article content. Therefore, keywords of literature were an important basis for studying the hot spots in a certain field.

Taking the CNKI database as the data source, the retrieval condition was ‘keywords=green development’. The source categories of journals include SCI, EI, CSSCI, CSCD and the core journals of Peking University. The retrieval time was April 7, 2021. The data selection period was up to 2020.

In the process of data preprocessing, the special preface, album introduction, special preface, special introduction, short message report, scholar introduction, conference notices, draft brief and other non-research literature were deleted, and the number of valid literature remaining was 1553, constituting the research data for our study.

Based on the above data sets, the word frequency matrix of literature keywords was constructed, some of which were shown in Table 1.

Table 1 Keyword frequency matrix of green development-related literature (partial)

Keywords / **Article**	green development	sustainable development	ecological civilization	…
Literature 1	1	1	0	…
Literature 2	1	0	1	…
Literature 3	1	0	0	…
…	…	…	…	…

In Table 1, each line represented a paper while each column represented a keyword. In the matrix, "1" represented that the literature contained the keyword, and "0" represented that the corresponding article did not contain the corresponding keyword.

3.2 Cluster Analysis of Domestic Green Development Research

The word frequency matrix was imported into SPSS software for system clustering. The intra-group connection method was selected for system clustering. Since the word frequency matrix has the characteristic of 0-1 binary matrix, the simple matching coefficient was used to measure the similarity between articles.

For the word frequency matrix, the simple matching coefficient was the proportion of the keywords which simultaneous and non-simultaneous existing in the two articles. The larger the simple matching coefficient was, the more similar the two articles were.

The simple matching coefficient can be expressed as Equation (1).

$$S=\frac{a+d}{a+b+c+d} \tag{1}$$

In Equation (1), "a" represents the number of keywords simultaneous occurrence in the two articles, "d" represents the number of keywords non-simultaneous occurrence in the two articles, "b" and "c" represent the number of keywords occurring in one article but not in the other article. The sum of a, b, c and d represents the number of all keywords.

It can be seen from the equation that the simple matching coefficient not only has a simple calculation formula but also can effectively use the sparsity of the word frequency matrix to measure the similarity of the literature. Through the system clustering, the current research related to green development in China can be divided into 11 categories.

The clustering results were shown in Table 2. We can see that the number of articles in each category varied at a considerable level. The number of articles in the first category is the largest, while the number of articles in the eleventh category is the least, which are 858 and 6 respectively. The number of articles in each category represented its research hot to some extent.

Table 2 Literature clustering results

Category	Number of articles	Category	Number of articles
1st category	858	7th category	39
2nd category	303	8th category	44
3rd category	43	9th category	31
4th category	60	10th category	37
5th category	36	11th category	6
6th category	96	Total	1553

4 SUBJECT MINING AND ANALYSIS BASED ON CHI-SQUARE STATISTICS

The chi-square statistical method was usually used for feature selection in text classification. In our study, the chi-square statistical method was adopted to measure the relevance of each keyword to each topic. Based on the relevance, the keywords with higher relevance can be selected to describe the connotation of each topic in the clustering results.

The core of the chi-square statistical method was to calculate the CHI value, and the calculation formula was as shown in Equation 2.

$$\chi^2(K_j, H_i) = \frac{N(AD - BC)^2}{(A+C)(A+B)(B+D)(C+D)} \tag{2}$$

In Equation (2), N represents the total number of articles; K_j represents the *j-th* keyword, and H_i represents the *i-th* category. "A" represents the number of articles belonging to the *i-th* category and contains the *j-th* keyword. "B" represents the number of articles that do not belong to the *i-th* category and contains the *j-th* keyword. "C" represents the number of articles that belong to the *i-th* category but do not contain *j-th* keyword; "D" represents the number of articles that do not belong to the *i-th* category and do not contain the *j-th* keyword.

Based on the calculation of the CHI value between each keyword and each category, the keywords were sorted in descending order according to the CHI value to the eleven the category. The higher the correlation degree of the keyword, the stronger the ability of the corresponding keyword to distinguish the topic category. So, the keyword with a high correlation degree can reveal the connotation of the topic. In our study, we selected the top 10 keywords of each category to define the topic. The specific results were shown in Table 3.

Table 3 High correlation keywords and Topic definition

Category	Keywords	Topic definition
1st category	ecology, ecological civilization, economy, construction, ecological civilization construction, Yangtze River Economic Belt, innovation, ecological environment, Marx, protection	Construction of ecological civilization
2nd category	ecology, ecological civilization, beautiful China, ecological governance, ecological compensation, Xi Jinping, ecological efficiency, ecological priority, ecological agriculture, contemporary value	Ecological governance
3rd category	sustainable development, sustainability, the fifth Plenary Session of the 18th CPC Central Committee, sustainability, Fifth Plenary Session, spatial autocorrelation, human-earth relationship, sustainable development goals, complex adaptation systems, water resources carrying capacity	Sustainable development
4th category	innovation, innovation drive, innovation development, open development, scientific and technological innovation, shared development, institutional innovation, green innovation, coordinated development, technological innovation	Innovation-driven development

5th category	environmental protection, protection, ecological environmental protection, environment, environmental protection tax, ecological environmental protection planning, pollution control battle, ecological environment, climate change economics, negative externalities of production	Environmental protection
6th category	economy, Yangtze River Economic Belt, circular economy, regional economy, economics, China's regional economy, county economy, political economy, development quality, energy consumption	Green economy
7th category	ecological civilization construction, construction, ecological civilization, green technology innovation, ecology, green technology, technology innovation, the Nineteenth National Congress, public participation, coal consumption	Green technology innovation
8th category	ecological environment, environment, wetland resources, greenways, green governance, ecology, wetlands, multiple governances, ecological advantages, environmental issues	Green environmental governance
9th category	index system, index, evaluation index system, evaluation index, comprehensive evaluation, evaluation, low carbon and green, Jiangxi Province, cloud model, highway transportation	Green development evaluation
10th category	Marx, Marxism, ecological view, Marxist Ecological View, ecology, ecological Marxism, ecological Marxism, Subject-object Relationship, Marxist View of Nature, Marxist Ecological Thought	Marxist Ecological View
11th category	Jinshan Yinshan, Yinshan, lucid waters and lush mountains, Jinshan, Green Water, Lucid waters and lush mountains are invaluable assets, lush mountains, ecological environment quality, ecological resources, ecological view of nature	Two-mountain theory

As shown in Table 3, the third column showed the topic definition of each category according to the corresponding keywords with high correlation. As explained above, there was a strong logical correlation between each topic and the representative keywords, therefore the research topic could be clearly defined accordingly.

Take the first three categories as examples. The high correlation keywords belonging to the first category include ecological civilization, ecological civilization construction, ecological environment and et al., therefore we should define the first category as “ecological civilization construction”. The high correlation keywords belonging to the second category include ecological compensation, ecological efficiency, ecological priority, ecological agriculture and so on, therefore we should define the second category as “ecological governance”. The high correlation keywords belonging to the third category include sustainable development, sustainability, human-earth relationship, sustainable development goals and so on, therefore we

should define the third category as “sustainable development”.

The remaining categories can be defined by analogy. The topics of the 4th to the 11th category are innovation-driven development, environmental protection, green economy, green technology innovation, green environmental governance, green development evaluation, Marxist ecological view and two mountains theory in order.

5 CONCLUSION

To effectively sort out the hot spots and development trends in the field of green development research in China, so as to provide a useful reference for further deepening research and application, we took the dissertations and journal articles related to the topics of green development in CNKI as the data source. Then all the related literature was clustered into 11 categories. To define the 11 categories, we adopted the chi-square statistical method to extract high correlation keywords of each category, by which we can describe the topic of each category reasonably. Finally, we defined the 11 research topics as ecological civilization construction, ecological governance, sustainable development, innovation-driven development, environmental protection, green economy, green technology innovation, green environmental governance, evaluation index system, Marxist ecological view, Two-mountain theory. Among them, papers belonging to the topic of ecological civilization construction research were most, signifying that the topic of ecological civilization construction was the research hotspot currently. Although there was little literature belonging to the topics such as two-mountain theoretical research, innovation-driven development research, green technology innovation and green economy research. However, the research heat has been increasing significantly in recent years which indicates the research trend of green development research in the future.

Through quantitative analysis of the literature data, the paper drew some valuable conclusions which can provide guidance for future research.

Acknowledgments. This research is funded by Research Program of Beijing Municipal Education Commission under grant KM201911232008 and National Key Research &Development Program under grant 2018YFB1003705.

REFERENCES

[1] W. J. Wang, D. Liu, The 70-year evolution and practice of green development thought in China[J]. Journal of Shaanxi Normal University (Philosophy and Social Sciences Edition), 2019, 48(6): 5-14.

[2] X. X. Wu, S. Y. Zhang, The formation and future trend of the concept of “green development” [J]. Economic Issues, 2017, (2): 30-34.

[3] D. B. Zhu, Xi Jinping’s Green Development Concept: Ideological Foundation, Connotation System and Time Value[J]. Economist, 2020, (3): 5-15.

[4] H. B. Mao, S. Y. Liu, The four dimensions of the green development concept[J]. Theoretical Vision, 2020, (1): 51-55.

[5] M. H. Xu, X. W. Li, The five dimensions and practical value of the green development concept[J]. Chongqing Social Sciences, 2018, (8): 74-82.

[6] J. Wu, On the three dimensions of green development[J]. Ningxia Social Sciences, 2018, (6): 17-21.
[7] J. Liu, K. K. Cao, Z. B. Huang, The system characteristics, value orientation and practical path of green development [J]. Academia, 2018, (5): 206-213.
[8] D. H. Liu, The scientific connotation and value orientation of the green development concept[J]. Jiangsu Social Sciences, 2017, (3):1-7.
[9] J. W. Zhang, Q. Wu, The philosophical foundation, realization path and great significance of the concept of green development [J]. Ecological Economy, 2017, 33(2): 172-175.
[10] F. Y. Zeng, An analysis of the basic framework of green development[J]. Environmental Protection, 2017, 45(12): 43-47.
[11] Q. Y. Zhang, L. H. Su, The value choice of green development and its realization path[J]. Journal of Xinjiang Normal University (Philosophy and Social Sciences Edition), 2017, 38(2): 25-32.
[12] X. L. Qin, Y. P. Liu, A review of research on green development[J]. Social Scientist, 2018, (4): 70-74, 81.
[13] L. J. Si, Changes in the hot topics of green development research——Knowledge graph analysis based on CiteSpace [J]. Lanzhou Academic Journal, 2019, (10): 130-145.