Asian Journal of Water, Environment and Pollution. Vol. 22, No. 5 (2025), pp. 153-164.
doi: 10.36922/AJWEP025160117

ORIGINAL RESEARCH ARTICLE

Analyzing emission and carbon reduction support policies using latent Dirichlet allocation and a Sankey-bubble chart

Na Li* and **Xiaoming Wu**

School of Economics and Management, Southwest Petroleum University, Chengdu, Sichuan, China
***Corresponding author:** Na Li (1486096069@qq.com)

Received: April 14, 2025; Revised: June 17, 2025; Accepted: June 27, 2025; Published online: July 22, 2025

Abstract: This study presents an in-depth analysis of China's emission and carbon reduction support policies from 2016 to 2023 using text mining techniques. The main objective is to examine the evolution, thematic focus, and implementation outcomes of these policies across different stages, thereby providing insights into their development patterns and potential future direction. Based on the latent Dirichlet allocation model implemented in Python 3.7, the study identified and refined 14 initial topic terms spanning three policy phases, which were subsequently integrated and interpreted. Through topic clustering and visualization using the Sankey-bubble chart, the research simulated the evolution of policy themes over time. The results reveal a clear shift in policy focus – from market-driven mechanisms to green development and technology-led approaches. In the later stages, policies exhibit more comprehensive and systematic characteristics. In conclusion, the study contributes to a deeper understanding of the development trajectory, orientation, and implementation effectiveness of China's carbon reduction policies, offering valuable insights for future policy development.

Keywords: Emission and carbon reduction support policies; Evolutionary trends; Latent Dirichlet allocation; Sankey-bubble chart

1. Introduction

The growing global climate change has become a widespread concern. According to the United Nations Framework Convention on Climate Change (https://www.un.org/zh/node/181981) and China's national standard General Rules for Accounting and Reporting of Greenhouse Gas Emissions from Industrial Enterprises (GB/T 32150-2015) (https://openstd.samr.gov.cn/bzgk/gb/newGbInfo?hcno=29DE620206A268D0E27B8739E332D70E), the seven identified greenhouse gases – namely carbon dioxide (CO_2), methane, nitrous oxide, nitrogen trifluoride, hydrofluorocarbons, perfluorocarbons, and sulfur hexafluoride – have the strongest ability to cause the greenhouse effect. Among these gases, CO_2 accounts for the highest proportion and contributes the most to global warming. As a result, carbon reduction has become a central focus of international discourse. At present, more than 120 countries and regions around the world have committed to reducing carbon emissions, with most setting 2050 as the target year.[1] At the 75th United Nations General Assembly on September 22, 2020, Xi Jinping, General Secretary of the Communist Party of China, announced China's commitment to peak carbon emissions by 2030 and achieve carbon neutrality by 2060. Subsequently, on November 12, 2020, General Secretary Xi delivered a video address at the third Paris Peace Forum, further reaffirming China's determination to reduce carbon emissions. In China, government policies serve as guiding principles for state action, aimed at achieving designated goals through structured regulation.

The promulgation of China's emission and carbon reduction policies has played a crucial role. If policies are inconsistent with national development or contain conflicts and omissions, they may obstruct the smooth progress of the industry.[2] Therefore, this study aims to analyze China's emission and carbon reduction support by exploring policy patterns through a data-driven analysis of textual policy documents. Through a quantitative analysis of policy documents and government plans – specifically examining state-issued reports from 2016 to 2023 – a clearer picture of the framework underlying China's emission and carbon reduction support policies can be obtained. In-depth textual analysis further reveals their structural evolution and thematic focus.

This study identifies several challenges encountered by China during policy formulation and proposes targeted measures to improve the implementation and structural coherence of emission reduction policies.[3] In addition, textual analysis can quantitatively examine the national emission and carbon reduction policies and highlight the major research themes within China's policy landscape, with particular attention paid to prominent policy directions. This approach aims to offer meaningful guidance for optimizing China's support mechanisms related to emission and carbon reduction efforts in the future. Accordingly, this research initially utilizes latent Dirichlet allocation (LDA) to identify key topics and extract representative keywords from the text, subsequently leveraging word distributions within each theme to extract the most salient content.

The characteristics of the three stages of emission and carbon reduction policies are then outlined. Subsequently, the overall evolutionary trends and central elements of each phase are synthesized. Finally, the extracted themes are visualized and analyzed using the Sankey-bubble chart (SBC). Ultimately, the study offers practical recommendations and strategies to support the advancement of China's emission and carbon reduction efforts, grounded in the analytical findings.[4]

2. Literature review

At present, CO_2 constitutes the majority of greenhouse gas emissions resulting from human activities. Since the other greenhouse gases are diverse and more challenging to eliminate or control, CO_2 emission reduction is often used as a proxy for overall greenhouse gas mitigation in national emission reduction targets. To date, countries around the world have adopted various expressions to convey their commitment to reducing carbon emissions, including "energy saving and emission reduction," "net-zero emissions," "net zero," and "emission and carbon reduction."[5]

2.1. Current research status of carbon emission reduction

"Carbon emission" is commonly used as a shorthand for greenhouse gas emission, referring to the total amount of greenhouse gases released into the atmosphere. Since the 1960s, climate change has gradually attracted widespread attention from the international community. Greenhouse gas emissions resulting from human activities – such as fossil fuel combustion and industrial production – have been identified as the main cause of global warming.

In response, the United Nations Environment Program established the Intergovernmental Panel on Climate Change in 1988 to focus on carbon emission research – both scientific and political – with the goal of reducing carbon emissions.[6] Liu *et al*.[7] projected carbon emissions in China's electric power sector from 2011 to 2040, with particular emphasis on the effects of electricity consumption per unit of gross domestic product and changes in industrial structure. Based on observed climate change records in the Malaysian region from 2010 to 2015, researchers developed a downscaled, dynamic integrated model of climate and economy, designed to support long-term (95-year) mitigation planning. This model, which incorporated variables such as temperature, the carbon cycle, carbon emissions, and carbon control, was used to empirically assess vulnerability impacts and explore potential remediation strategies in response to climate change.[8]

Ahmed *et al*.[9] investigated CO_2 emission drivers in the regions of China and India – two of the world's largest coal consumers and most populous countries – using machine learning techniques, such as the long short-term memory method. Wang *et al*.[10] developed an integrated evaluation system focused on pollution reduction, carbon mitigation, and economic development to support regional policy formulation in China. They analyzed energy, emissions, and CO_2 output data from 30 provinces between 2016 and 2018, employing the grey correlation method to conduct a comparative analysis of regional performance. Zhang *et al*.[11] assessed trends in carbon emission intensity across 283 cities and strategic regions in China, highlighting regional disparities, dynamic shifts, and convergence patterns. The results indicated a substantial reduction in carbon intensity, with a consistent spatial pattern of higher values in the north and lower values in the south.

2.2. Research on the construction path of the carbon emission reduction system

Farahani *et al*.[12] employed panel cointegration analysis to analyze the relationship between economic growth and carbon emissions in seven emerging countries from 1995 to 2018. The results indicated that economic growth has a significant positive impact on CO_2 emissions in both the short and long term. In addition, Zhang *et al*.[13] analyzed China's carbon emission pathways under various scenarios by integrating a top-down carbon emission model with an existing bottom-up model. Their findings showed that by 2060, significant differences in emission levels will emerge across scenarios, highlighting the need for more stringent emission reduction policies across different industries. Song *et al*.[14] developed a localized Low Emissions Analysis Platform model, using Chongqing municipality in China as a case study. They explored key influencing factors and carbon reduction pathway characteristics under Chongqing's carbon peaking targets, applying methods such as the logarithmic mean Divisia index decomposition, Tapio decoupling elasticity coefficients, comparative analysis of individual mitigation measures, and cross-elasticity between pollution and carbon reduction. Neha *et al*.[15] emphasized that sustainable development can be advanced through multiple strategies, including the adoption of green building standards, the use of renewable energy, advanced building management systems, carbon capture and storage technologies, and life cycle analysis to monitor and reduce carbon emissions.

To achieve the Paris Agreement's target of limiting global temperature rise to 1.5°C or 2°C, mitigation actions must align with the Sustainable Development Goals. A review of selected studies from various regions drew conclusions related to trade-offs between mitigation strategies and Sustainable Development Goals, as well as their implications for costs and equity across different development contexts.[16] Xiao *et al*.[17] employed the Greenhouse Gas and Air Pollution Interactions and Synergies model for the Beijing-Tianjin-Hebei (JJJ) region (GAINS-JJJ) model to assess the synergistic benefits of pollution and carbon reduction under the 14th 5-Year Air Pollution and Control Policy in Taiyuan City. Their findings indicated that 13 air pollution control measures significantly reduced concentrations of particulate matter (PM) 2.5, PM10, sulfur dioxide, nitrogen oxides, volatile organic compounds, and ammonia under the 2025 policy scenario.

2.3. Study on the development path of emission reduction and carbon reduction policies

Ashina *et al*.[18] argued that achieving an 80% reduction in CO_2 emissions by 2050, relative to 1990 levels, is feasible. Their analysis, based on the Asia-Pacific Integrated Model/backtracking model and backcasting methodology, suggested that early action significantly increases the likelihood of realizing a low-carbon future. Marsden *et al*.[19] emphasized that the transport sector remains a major contributor to greenhouse gas emissions due to its reliance on fossil fuels and the uncertain transition toward ultra-low-carbon vehicles. They proposed that policy interventions can support emission reductions by encouraging shifts in individual travel behavior and decision-making. Dutta[20] emphasized that adopting low-carbon development strategies in the transport and energy sectors can effectively address India's significant infrastructure challenges. The study also noted that successful and timely implementation of an inclusive, low-carbon growth pathway requires robust technology transfer, strong institutional frameworks, and supportive policy mechanisms.

He *et al*.[21] stated that, in the long term, transitioning to a low-carbon energy system is essential to achieving peak carbon. While maintaining sustainable economic and social development, it is equally important to promote the early peaking of CO_2 emissions, thereby playing a constructive role in advancing global sustainable development. Jiang *et al*.[22] quantitatively assessed the drivers of synergistic reductions in PM2.5 and CO_2 emissions by integrating inter-provincial and global multi-regional input–output models in China. The findings highlighted the importance of adjusting production structures and exploring additional emission reduction opportunities tailored to local conditions. Xu[23] recommended that China adopt a systemic approach, promoting synergy among emission reduction, pollution control, green transformation, and economic development. The study emphasized the importance of refining high-level tax system design, improving and optimizing green tax policies, and more effectively promoting green development through synergistic tax policies.

Briefly, this study advances the field through innovations in the following areas:

(i) While most research on China's emission reduction policies focuses on the current status, system design, and development pathways, relatively few studies examine the policy texts themselves. The present study addresses this gap by analyzing policy documents and identifying temporal trends through text clustering and thematic analysis.

(ii) Although text mining techniques are widely applied in fields such as business analytics, fraud detection, and spam filtering, their application in the analysis of emission reduction and carbon support policies remains limited. This study employs LDA for keyword-based text clustering and integrates SBC for visual interpretation, thereby developing a combined LDA–SBC analytical framework.

In summary, this study adopts a combined LDA–SBC analytical approach for data analysis. First, the LDA topic clustering model is applied to extract thematic word distributions from policy texts. Next, the SBC is used to visualize and interpret these distributions, thereby revealing the scope and focus of the policies. Finally, this method supports a deeper understanding of enterprise development trends, industrial planning, and the strategic direction of national policies.

3. Research methodology

China adopts a 5-year cycle for its national development plans and is currently implementing the 14th 5-Year Plan. For comparative analysis, this study employed the 13th 5-Year Plan as a reference point. Accordingly, the period from 2016 to 2023 was examined to analyze both the similarities and differences between these two planning phases. The timeframe was segmented into three stages: 2016 – 2018 is defined as the first phase, representing the early stage of the 13th 5-Year Plan; 2019 – 2021 constitutes the second phase, covering the late of the 13th 5-Year Plan and the lead-up to the 14th; and the 2022 – 2023 represents the third phase, corresponding to the early stage of the 14th 5-Year Plan.

The LDA technique enables effective text classification by identifying and grouping related themes, while the K-means algorithm offers strong scalability and efficiency, operating within polynomial time.[24] Therefore, LDA was selected as the clustering method for the text data. In addition, SBCs can display flow paths and multiple data dimensions simultaneously within a single chart, making the visualization of complex data more intuitive and information-rich.[25] In this context, SBC diagrams can help decision-makers and analysts better understand relationships and trends within the data, thereby supporting more informed decision-making.

This study employed LDA topic clustering to extract more detailed information, thereby enriching the content and broadening the analytical perspective. The results were then visualized using SBC to more comprehensively illustrate how China's policies align with the actual demands of emission and carbon reduction.

3.1. LDA

a. Principles of LDA

The LDA algorithm, introduced by Blei *et al.*[26] in 2003, takes a document set $D = \{d_1, d_2, d_3, \dots d_n\}$ and a specified number of clusters m as input. The algorithm then calculates the probability p of each document d_i belonging to every topic.[27] Every document is thus represented by a probabilistic distribution over multiple topics, expressed as $d_i = (d_{p1}, d_{p2}, \dots, d_{pm})$. For every word within a document, the algorithm also calculates the probability values associated with each topic, denoted as $W_i = (W_{p1}, W_{p2}, \dots, W_{pm})$. The model ultimately produces two key matrices: one representing the distribution of topics across documents, and the other depicting the distribution of words across topics.

Therefore, the LDA algorithm maps documents and words into a series of topics and attempts to use these topics to discover hidden relationships between documents and comments, documents and other documents, and terms and phrases. LDA is an unsupervised learning method that does not require each target to meet certain constraints. Instead, it counts the word frequency distribution in each topic after clustering and identifies the highest-frequency words in each case, from which the topic meaning is inferred.

b. LDA process

Each document d in the document set D is regarded as a sequence of words, and let d have n observations in the series $<w_1, w_2, \dots w_n>$,where $w_i$denotes the *i-th* word.

All the unique words present in$D$ form a large set called the "vocabulary" (VOC). For each document d in$D$, the probability of belonging to different topics is represented as $\theta_d = <p_{t1}, \dots, p_{tk}>$, where p_{ti} denotes the probability that document d corresponds to the i-th topic in set T. This probability is calculated as:

$$p_{ti} = \frac{n_{ti}}{n} \tag{I}$$

Where n_{ti} denotes the number of words in d corresponding to the *i-th* topic, and n is the total number of all words in d.

For each topic t in the set T of topics, the probability of generating different words is represented by the vector $\varphi_t = < p_{w1}, ..., p_{wm} >$, where p_{wi} denotes the likelihood of topic t generating the i-th word in the VOC. This probability is calculated as:

$$p_{wi} = \frac{N_{wi}}{N} \quad \text{(II)}$$

Here, N_{wi} denotes the number of times the i-th word in VOC is assigned to topic t, and N represents the total number of words associated with topic t.

The core formula of LDA is given by:

$$p(w|d) = p(w|t) \times p(t|d) \quad \text{(III)}$$

Using the topic as an intermediate layer, the current values of θ_d and φt enable the calculation of the probability of occurrence of a word w in document d, where $p(t|d)$ is computed using θ_d, and $p(w|t)$ is derived from φt. In addition, with the current values of θ_d and φt, it is possible to calculate $p(w|d)$ by considering all possible topics that a word in a document might belong to. Based on this result, the topic assignment for the word is updated. If the topic assignment changes, it will, in turn, affect both θ_d and φt. LDA initializes θ_d and φt with random values and iteratively updates them until the model converges.[28]

3.2. SBC

a. Principle of Sankey-bubble diagram
   The Sankey diagram, also known as the Sankey energy flow diagram or Sankey energy balance diagram, was first introduced in 1898 by Matthew Henry Phineas Riall Sankey. He created a now-famous diagram called the “energy efficiency of the steam engine,” and since then, the visualization has been named after the “Sankey diagram.” Sankey-bubble diagrams combine the features of a bubble chart. Bubble charts can be used to show correlations between three measures or to present three-dimensional data simultaneously.
   The SBC integrates a Sankey diagram with a bubble chart to more effectively visualize complex, multi-dimensional data. These two charts are linked through the names of pathways. The Sankey diagram displays the genes associated with each pathway, while the bubble chart represents multiple attributes: the position of each bubble corresponds to the GeneRatio, the size of the bubble indicates the number of genes enriched in the pathway, and the color represents the pathway’s p-value.[29]

b. SBC process
   The SBC consists of a Sankey chart and a bubble chart, which together can more effectively and comprehensively display the multi-dimensional information within the data. The Sankey diagram shows the relationships between different pathways and genes – each pathway represents the gene distribution within that pathway. In the bubble chart, the position of each bubble indicates the GeneRatio, the size represents the number of genes, and the color reflects the p-value. The Sankey diagram and bubble chart are connected by multiple signaling pathways. In the Sankey diagram, different or identical genes contained within each pathway are shown, while in the bubble chart, the position of each bubble indicates the negative logarithmic of the p-value or Q-value of a pathway, the bubble size corresponds to the number of genes enriched in the pathway, and the color reflects the pathway’s Hit Ratio, among other attributes.[30]

4. Empirical analyses

4.1. Data acquisition and processing

The study employed custom-built web crawlers in a Python 3.7 (Python Software Foundation, USA) environment to collect emission and carbon reduction policy documents from the websites of the National Development and Reform Commission, the Ministry of Agriculture and Rural Affairs, the National Bureau of Statistics, the National Energy Administration, the Ministry of Science and Technology, and the Ministry of Ecology and Environment, covering the period from 2016 to 2023. These documents served as samples for analyzing relevant policies and regulations.

Considering the diverse and interconnected functions of Chinese government departments, this study collected carbon emission reduction data from 2016 to 2023 via official government websites, enabling a comprehensive analysis of emission reduction efforts at the departmental level.

Figure 1 illustrates the changing pattern in the number of policy documents issued between 2016 and 2023. Although there are fluctuations in issuance across various departments, the overall trajectory suggests an upward trend, with 70 documents released in 2023, reflecting China’s increasing focus on emission and carbon reduction as a key aspect of national development. In terms of the total number of documents, the peak occurred in 2021, with 115 documents published. During that year, the administrative arm of the State Council, the

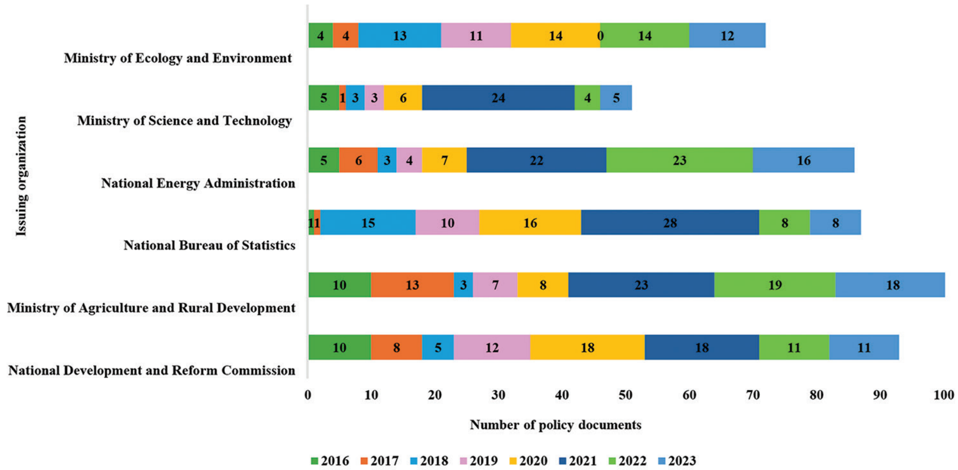

Figure 1. Trends in the issuance of support policies for emission and carbon reduction in China (2016 – 2023)

national transportation authority, the ministry overseeing agriculture and rural development, the national economic planning authority, and the commerce administration department all reached their historical highs.

From the end of 2019, owing to the influence of the general environment, the Chinese government began prioritizing the emission and carbon reduction sector, and various types of emission- and carbon reduction-related policies issued from 2020 onwards have guided the development of the emission reduction and carbon reduction industry across multiple levels. Thus, starting from 2021, the nation's focus on emission and carbon reduction intensified, which naturally increased the regulation of emission and carbon reduction in macro policy – marking the peak period of China's emission and carbon reduction support policy issuance.[31,32]

Unstructured policy texts form the core data of this research, making data preprocessing a crucial initial step. This involves removing invalid data – such as numbers, English characters, garbled text, intonation marks, punctuation, and irrelevant content like names of people, companies, and document descriptions. These unwanted elements can be processed using regular expressions to replace symbols or by adding a stopword dictionary to filter out unnecessary words.

4.2. Data analysis

4.2.1. LDA analysis

The number of clusters is selected based on the perplexity score of the topic model. Perplexity reflects the effectiveness of classification under varying topic counts – lower perplexity values generally indicate better topic separation and coherence. The optimal number of topics in the LDA model is determined based on this perplexity score. As shown in Figure 2, the lowest perplexity is observed when the number of topics is 14, indicating the best clustering performance. Therefore, the number of clusters for the LDA model is set to 14.

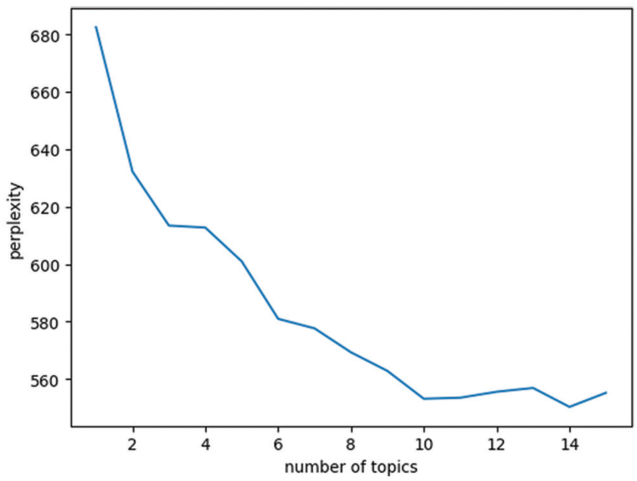

Figure 2. Line graph of theme perplexity

LDA topic modeling was performed on the text corpus, and the outcomes are displayed in Figures 3-5. In the topic clustering visualization, each circle on the left side corresponds to an individual topic. The greater the spatial separation between topics, the more distinct their content differences, indicating better classification. On the right, the distribution of words for each topic is shown, with words positioned further forward having higher frequencies within that topic.

As shown in Figures 3-5, based on the spatial layout of each category, it is observed that most categories are distributed at considerable distances from one another, with only two showing slight spatial overlap. Thus, the theme division across the three phases demonstrates relatively good performance. From the 14 thematic clusters identified through LDA, the most prominent

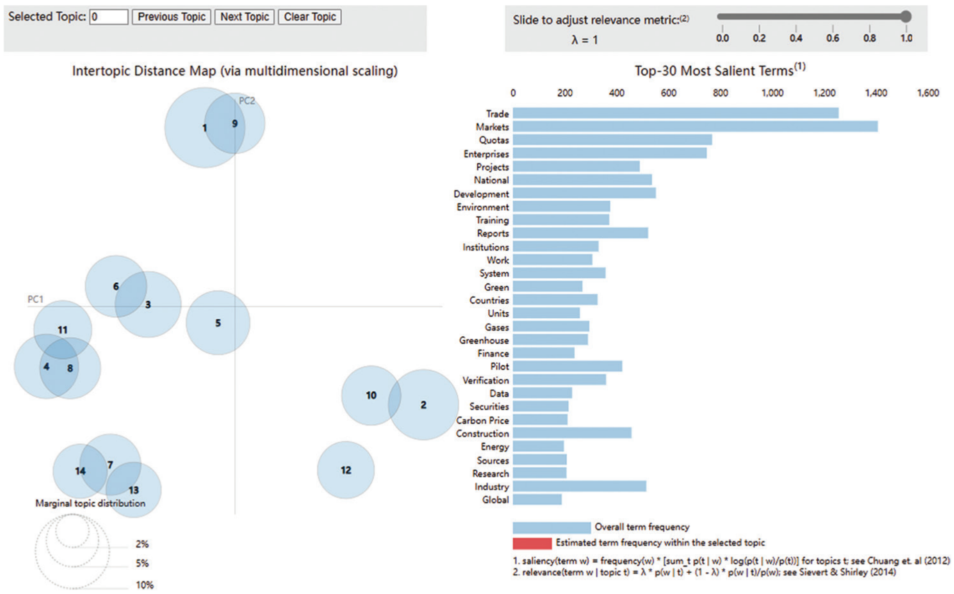

Figure 3. Clustering of latent Dirichlet allocation themes for emission and carbon reduction support policies (2016 – 2018)
Abbreviation: PC: Principal component.

and representative terms are extracted to characterize the unique thematic orientation of each category. The LDA topic clustering results are organized into three periods: 2016 – 2018, 2019 – 2021, and 2022 – 2023. The keywords are summarized according to the theme names, as shown in Table 1.

4.2.2. SBC analysis

Based on the theme names summarized from the content of each keyword, the data are visualized and analyzed using the SBC. The specific process of the SBC is as follows:

(i) Step 1: Data preparation (as shown in Table 2). This includes pathway names, genes, GeneRatio, gene count, and *p*-value.

Step 2: Tool selection. Python and Plotly are selected to create the Sankey-bubble plot.

(ii) Step 3: Data entry. The above data are prepared in a suitable format for use in Plotly.

(iii) Step 4: Sankey diagram creation: Generate the Sankey diagram using the formatted data.

As shown in Figure 6, based on the spatial distribution of the various categories in the Sankey-bubble map, the focus of China's emission and carbon reduction support policies at different stages – and their changing trends – is clearly illustrated. The first stage (2016 – 2018) emphasizes market and trading mechanisms; the middle stage (2019 – 2021) shifts the focus to the green development of enterprises; and the later stage (2022 – 2023) further reinforces the concept of green development while comprehensively promoting a green and low-carbon transformation. This evolutionary path reflects China's continued efforts and policy innovations in carbon reduction.[31]

Table 1. Ranking of latent Dirichlet allocation thematic clustering results across the three phases

Rank	2016 – 2018	2019 – 2021	2022 – 2023
1	Trade	Enterprises	Green
2	Markets	Green	Development
3	Quotas	Technologies	Energy
4	Enterprises	Energy	Technology
5	Projects	Development	Businesses
6	Development	Markets	Buildings
7	Environment	Projects	Markets
8	Training	Architecture	Focus
9	Report	Report	System
10	Institutions	Target	Standard
11	Work	Investment	Research
12	System	Sources	Peak carbon
13	Green	New energy	Work
14	National	Company	Targets

4.3. Results of the comprehensive analysis

This study applied sub-phrase segmentation and employed both LDA topic modeling and SBC

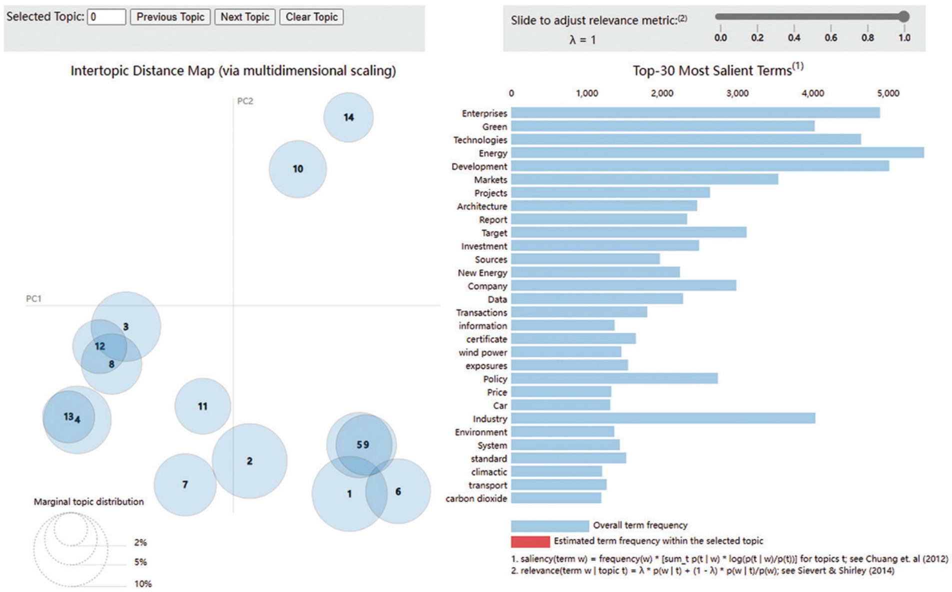

Figure 4. Clustering of latent Dirichlet allocation themes for emission and carbon reduction support policies (2019 – 2021)

Abbreviation: PC: Principal component.

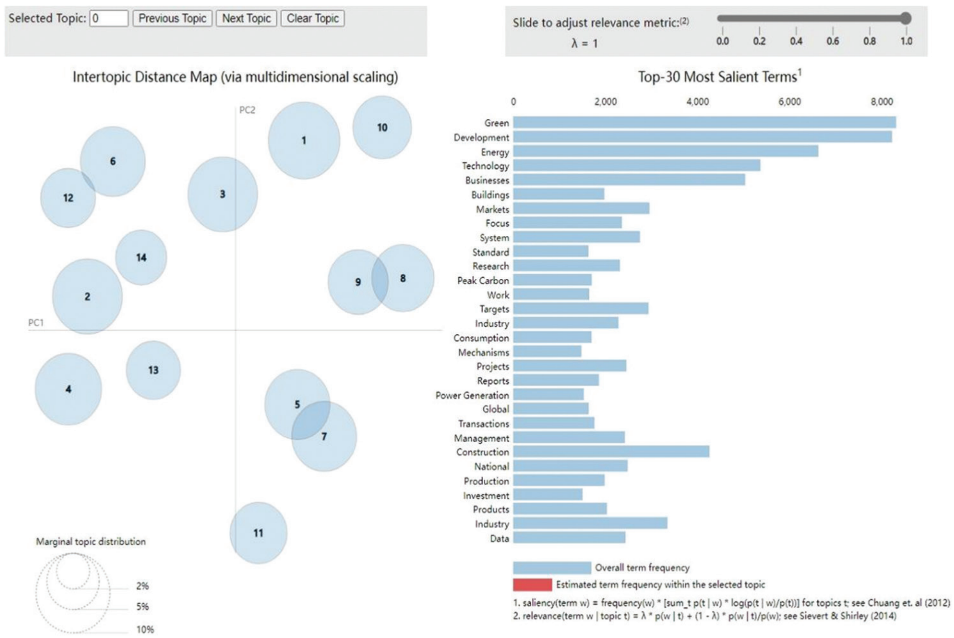

Figure 5. Clustering of latent Dirichlet allocation themes for emission and carbon reduction support policies (2022 – 2023)

Abbreviation: PC: Principal component.

visualization to extract and analyze core content related to China's emission and carbon reduction policies.[33] By clustering policy themes across three time phases and identifying representative keywords, it provides effective analytical support for policy assessment and industrial development.[34] The findings are summarized as follows:

(i) The development trend of key emission and carbon reduction themes demonstrates evident shifts and identifiable categorization characteristics. In the

Table 2. Sankey-bubble chart data

Pathway	GeneRatio	*p*-value	GeneID	Count
2016 – 2018	Trade/Markets/Quotas/Enterprises/Projects/Development/Environment/ Training/Report/Institutions/Work/System/Green/National	0.1 – 0.2	1 – 2	14
2019 – 2021	Enterprises/Green/Technologies/Energy/Development/Markets/Projects/ Architecture/Report/Target/Investment/Sources/New energy/Company	0.2 – 0.3	2 – 4	14
2022 – 2023	Green/Development/Energy/Technology/Businesses/Buildings/Markets/ Focus/System/Standard/Research/Peak Carbon/Work/Targets	0.4 – 0.5	4 – 6	14

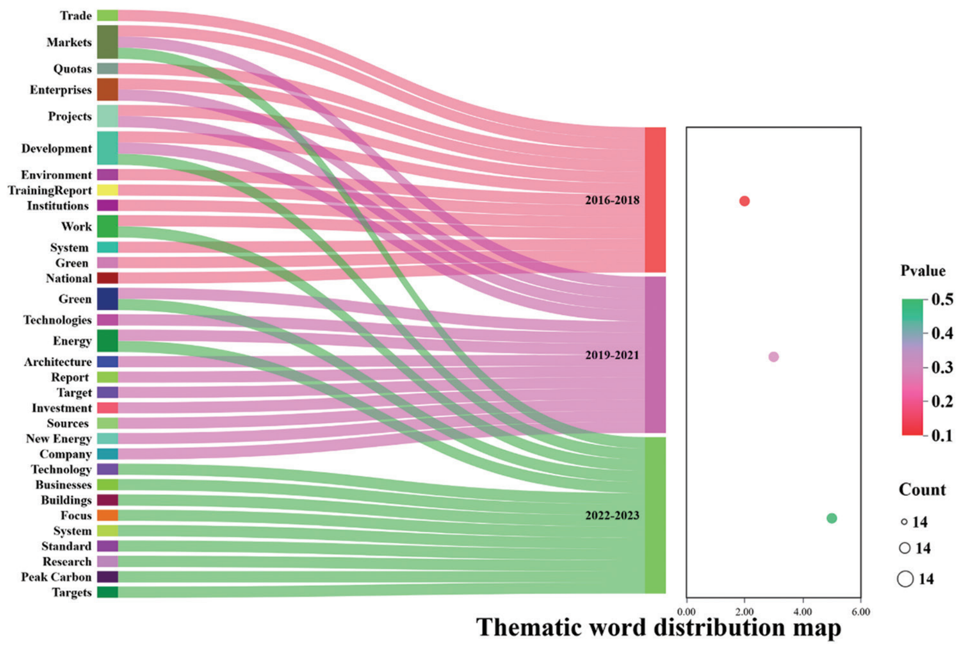

Figure 6. Distribution of thematic keywords across the three phases (2016 – 2023)

initial phase (2016 – 2018), the policies primarily focused on trading, including market mechanisms, quotas, enterprises, and related areas. In the middle phase (2019 – 2021), the focus shifted to enterprise development, particularly in green initiatives, technology, and energy. In the later phase (2022 – 2023), the policy emphasis was mainly on green development, encompassing areas such as energy and technology[35]

(ii) By conducting textual segmentation and analysis, the key areas of focus in China's emission and carbon reduction policies are identified. This approach enables the extraction of essential content, providing insights into policy emphasis across various dimensions in recent years. The results of lexical processing suggest that China's recent emission and carbon reduction support policies primarily emphasize green development, energy, and technology. The findings from sub-word segmentation not only align with but also enrich the results derived from clustering analysis, thereby providing a more comprehensive understanding

(iii) The Sankey-bubble diagram is linked by pathway names, where the Sankey diagram shows the genes associated with each pathway. The bubble diagram represents the GeneRatio value of each pathway by the position of the bubbles, the number of genes enriched in the pathway by the size of the bubbles, and the *p*-value of the pathway by the color of the bubbles. The power of the SBC lies in its ability to show both flow pathways and multiple data dimensions within the same chart, making the visualization of complex data more intuitive and informative.

5. Conclusion

Carbon reduction and abatement have already formed a more comprehensive policy dataset. In recent years, Chinese government departments at all levels have issued a series of policy documents on China's emission and carbon reduction support, which can be classified and organized to obtain a relatively complete dataset of China's emission and carbon reduction support policies.[36] These datasets serve as valuable resources for examining the trends and key features of carbon reduction policies, providing insights into strategic decision-making in the carbon abatement industry.

This study provides a new approach to analyzing China's emission and carbon reduction policies. By applying text mining techniques – such as LDA – to extract and analyze key elements from policy texts, combined with SBC visualization, it proposes a framework that includes text preprocessing, segmentation, frequency analysis, topic clustering, and visualization. This approach not only offers a novel perspective on China's emission reduction policies but also lays a foundation for extending these methods to other areas of policy research. Thematic analysis reveals that green development, scientific and technological innovation, and energy transition have become key directions in China's carbon reduction policies.[37] It is recommended that future policy optimization focus on enhancing synergy among these three themes. For example, green development policies should be better integrated with technological innovation initiatives to promote green technologies, while energy transition efforts should be supported by both regulatory and market-based tools to ensure efficiency and fairness.[38]

China's emission and carbon reduction policies have attracted significant attention, highlighting the need to improve their timeliness, coverage, and coherence. A combined approach involving green development, energy, and technology, along with the strategic use of various policy tools, is essential. Strengthening government support and leveraging financial, fiscal, and tax incentives can further advance these policies and promote national sustainability efforts.[39] Furthermore, technological advancements – such as high-performance air filtration materials – serve as critical enablers that complement policy efforts,[40] offering practical tools for energy conservation and carbon emission reduction under the broader framework of sustainable development.[41]

Acknowledgments

The authors thank all individuals and organizations who provided valuable assistance and support during this research.

Funding

This study was financially supported by National Social Science Fund of China Major Project Research on the high-quality development path of natural gasindustry driven by the energy revolution'(No.22&ZD105) and the Sichuan Key Research Base of Humanities and Social Sciences (Project Approval No. ZYZX-YB-2411).

Conflict of interest

The authors declare that they have no competing interests.

Author contributions

Conceptualization: Na Li
Data curation: Na Li
Formal analysis: Xiaoming Wu
Methodology: Xiaoming Wu
Writing–original draft: Na Li
Writing–review & editing: Xiaoming Wu

Availability of data

Data supporting the findings of this study are available from the corresponding author upon reasonable request.

References

1. Wang X. Analysis of the effectiveness of energy-use rights trading on emission reduction and carbon reduction. *Econ Issues*. 2023;(4):79-85. doi:10.16011/j.cnki.jjwt.2023.04.013
2. Binjie G, Haixia Z, Xinliao L, Tianyuan Z, Jinding F. Research progress and prospects of collaborative reduction of pollution and carbon dioxide based on bibliometrics. *J Environ Eng Technol*. 2023;13(1):85-95. doi: 10.12153/j.issn.1674-991X.20210780
3. Gong X, Sun T. Study on the impact mechanism of energy-saving and emission reduction policy tools to reduce carbon intensity: A qualitative comparative analysis of fuzzy sets based on data from 30 provinces (cities and autonomous regions). *Urban Issues*. 2021;(7):23-32. doi: 10.13239/j.bjsshkxy.cswt.210703
4. Li Y, Zhao Q, Xue Z. Development path of carbon dioxide capture, utilisation and storage technology and industrialisation under the 'double carbon' target. *Pet Drilling Tech*. 2023;45(6):655-660. doi: 10.13639/j.odpt.202201052
5. Sun Y, Fan J, Jia WG. A study on the stages and effective paths to achieve carbon peaking - based on the perspective of green gdp accounting. *Ind Technol Econ*.

2023;42(12):95-104.
6. Kan L. Study on China's 2060 carbon neutrality target and its implementation path. *Ecol Econ*. 2021;37(11):6.
7. Liu Z, Zhu K, Yan J, et al. Analysis of carbon emission reduction potential of the power sector under industrial structure optimisation. *J Ind Eng Eng Manage*. 2014;28(2):87-92+86.
doi: 10.13587/j.cnki.jieem.2014.02.004
8. Al-Amin AQ, Rasiah R, Chenayah S. Prioritizing climate change mitigation: An assessment using Malaysia to reduce carbon emissions in future. *Environ Sci Policy*. 2015;50:24-33.
doi: 10.1016/j.envsci.2015.02.002
9. Ahmed M, Shuai C, Ahmed M. Influencing factors of carbon emissions and their trends in china and india: A machine learning method. *Environ Sci Pollut Res*. 2022;29(32):48424-48437.
doi: 10.1007/s11356-022-18711-3
10. Han W, Hui L, Han W, Shulan W, Wenjie Z. Study on the difference between pollution reduction and carbon reduction and regional economic development level in China. *J Environ Eng Technol*. 2022;12(5):1584-1592.
doi: 10.12153/j.issn.1674-991X.20210268
11. Zhang Z, Zhang T, Feng D. Regional differences, dynamic evolution and convergence of carbon emission intensity in China. *J Quant Technol Econ*. 2022;39(4):67-87.
doi: 10.13653/j.cnki.jqte.2022.04.001
12. Farahani HR, Rassafi AA, Zhang K, Nie YM. A multi-hop control scheme for traffic management. *Transp Res Part C Emerg Technol*. 2021;130:103278.
doi: 10.1016/j.trc.2021.103278
13. Zhang S, Li M, Wang C. Provincial carbon emission trends and differentiated peak pathways in China. *China Popul Resour Environ.* 2021;31(9):45-54.
14. Song P, Zhang HM, Mao XQ. Carbon emission reduction pathways in Chongqing towards peak carbon targets. *China Environ Sci.* 2022;42(3):1446-1455.
doi: 10.19674/j.cnki.issn1000-6923.20210923.006
15. Neha S, Sharma RL, Kundan Y. Sustainable development by carbon emission reduction and its quantification: An overview of current methods and best practices. *Asian J Civil Eng Build Hous*. 2023;24(8):3797-3822.
doi: 10.1007/s42107-023-00732-z
16. Halsns K, Some S, Pathak M. Beyond synergies: Understanding sdg trade-offs, equity and implementation challenges of sectoral climate change mitigation options. *Sustain Sci*. 2024;19(1):35-49.
doi: 10.1007/s11625-023-01322-3
17. Xiao T, Shu Y, Li H, *et al*. Assessment of CO_2 synergistic benefits of air pollution control policies in Taiyuan City's 14th five-year plan. *Environ Sci*. 2024;45(3):1265-1273.
doi: 10.13227/j.hjkx.202304046
18. Ashina S, Fujino J, Masui T, Ehara T, Hibino G. A roadmap towards a low-carbon society in japan using backcasting methodology: Feasible pathways for achieving an 80% reduction in Co_2 emissions by 2050. *Energy Policy*. 2012;41:584-598.
doi: 10.1016/j.enpol.2011.11.020
19. Marsden G, Mullen C, Bache I, Bartle I, Flinders M. Carbon reduction and travel behaviour: Discourses, disputes and contradictions in governance. *Transp Policy*. 2014;35:71-78.
doi: 10.1016/j.tranpol.2014.05.012
20. Dutta V, Dasgupta P, Hultman N, Gadag G. Evaluating expert opinion on India's climate policy: Opportunities and barriers to low-carbon inclusive growth. *Clim Dev*. 2016;8(4):336-350.
doi: 10.1080/17565529.2015.1067181
21. He J, Lu L, Wang HL. Win-win path analysis of economic growth and CO_2 emission reduction. *China Popul Resour Environ*. 2018;28(10):9.
22. Jiang Y, Tang X, Ren KP, Ding I. A study on the drivers of pollution and carbon reduction in China based on double nested sda. *Syst Eng Theor Pract*. 2022;42(12):11.
23. Xu W. Improving tax policies in support of green development: Top-level design and policy synergy. *Int Taxation China.* 2023;(4):9-14.
doi: 10.19376/j.cnki.cn10-1142/f.2023.04.007
24. Ma H, Zhang B. A characterisation of China's open public data policy supply based on the lda model. *Mod Intell*. 2023;43(8):35-44.
25. Jiang T, Xiao W, Zhang C, Ge B. A time series text visualisation method based on Sankey diagram. *Comput Appl Res*. 2016;33(9):6.
26. Blei DM, Ng AY, Jordan MI. Latent dirichlet allo-cation. *J Mach Learn Res*. 2003;3:993-1022.
27. Cheng L, Lu XY, Xiao LM, Bai Y, Li XY. Water supply, use, consumption, discharge and return processes and their Sankey diagrams. *Hydropower Energy Sci*. 2021;39(6):5.
28. Zhang H, Zhang Y, Liu J, Chen T. Quantitative evaluation of China's industrial chain policy under the objective of industrial chain modernisation. *Stat Inform Forum*. 2023;38(9):32-46.
29. Sakshi, Kukreja V. Recent trends in mathematical expressions recognition: An lda-based analysis. *Expert Syst Appl*. 2023;213:119028.
doi: 10.1016/j.eswa.2022.119028
30. Song WL, Xiao RL. Evolutionary paths of policy instruments in China's photovoltaic industry and their implications. *J Inf.* 2022;41(1):177-184.
doi: 10.39674/j.cnki.issn1000-7490.2022.01.020
31. Xue F, Zhou M, Liu J. Can industrial transformation and upgrading reduce carbon emissions? Evidence from national industrial transformation and upgrading demonstration zones. *Ind Econ Res.* 2023;(2):1-13.
doi: 10.13269/j.cnki.ier.2023.02.002
32. Bakary K, Mouhamadou D, Seydou T. Contribution based on neurons networks for the prediction of greenhouse gas emissions in a handling port. *Asian J Water Environ Pollut*. 2024;21(6):261-269.

doi: 10.3233/AJW240094

33. Li N, Liu YT, Chen Z. Unlocking insights: Integrated text mining and interpretive structural modeling for enhanced user review analysis. *PeerJ Comput Sci*. 2024;10:e2541. doi: 10.7717/peerj-cs.2541
34. Huang W, Zhu S, Yao X. Destination image recognition and emotion analysis: Evidence from user-generated content of online travel communities. *Comput J*. 2020;64(3):296-304. doi: 10.1093/comjnl/bxaa064
35. Li N, Li S. Research on the LDA-ECD based support policy for China's agricultural cold chain logistics. *Sustain Futur*. 2025;9:100460. doi: 10.1016/j.sftr.2025.100460
36. Tu Z, Mao J, Xiao X, Ma G. From conflict to coordination: A new framework to measure the synergy level of pollution-carbon-growth. *J Asian Econ*. 2025;99.
37. Fetisov V, Gonopolsky AM, Davardoost H, Ghanbari AR, Mohammadi AH. Regulation and impact of VOC and CO_2 emissions on low-carbon energy systems resilient to climate change: A case study on an environmental issue in the oil and gas industry. *Energy Sci Eng*. 2023;11:1516-1535. doi: 10.1002/ese3.1383
38. Fetisov V, Gonopolsky AM, Zemenkova MY, *et al*. On the integration of CO_2 capture technologies for an oil refinery. *Energies*. 2023;16;865. doi: 10.3390/en16020865
39. Mork D, Delaney S, Dominici F. Policy-induced air pollution health disparities: Statistical and data science considerations. *Science*. 2024;385(6707):391-396. doi: 10.1126/science.adp1870
40. Karimi Kisomi M, Seddighi S, Mohammadpour R, Rezaniakolaei A H. Enhancing air filtration efficiency with triboelectric nanogenerators in face masks and industrial filters. *Nano Energy*. 2023;112:108514. doi: 10.1016/j.nanoen.2023.108514
41. Zhang S, Liu H, Tang N, Zhou S, Yu J, Ding B. Spider-web-inspired $PM_{0.3}$ filters based on self-sustained electrostatic nanostructured networks. *Adv Mater*. 2020;32(29):e2002361. doi: 10.1002/adma.202002361