

Article

Research Frontiers in the Field of Agricultural Resources and the Environment

Limin Chuan ^{1,†}, Jingjuan Zhao ^{1,†}, Shijie Qi ¹, Qian Jia ¹, Hui Zhang ^{1,*} and Sa Ye ^{2,*} 

¹ Institute of Data Science and Agricultural Economics, Beijing Academy of Agriculture and Forestry Sciences, Beijing 100097, China; chuanlm@agri.ac.cn (L.C.); zhaojj@agri.ac.cn (J.Z.); qishijie18@163.com (S.Q.); jq19880701@163.com (Q.J.)

² Key Laboratory of Knowledge Mining and Knowledge Services in Agricultural Converging Publishing, National Press and Publication Administration, Agricultural Information Institute, Chinese Academy of Agricultural Sciences, Beijing 100081, China

* Correspondence: zhanghui@agri.ac.cn (H.Z.); yesa@caas.cn (S.Y.); Tel.: +86-10-51502787 (H.Z.); +86-10-82109652 (S.Y.)

[†] These authors contributed equally to this work.

Abstract: From the perspective of project and paper datasets, research frontier recognition in the field of agricultural resources and the environment using the Latent Dirichlet Allocation (LDA) topic extraction model was studied. By combining the wisdom of domain experts to judge the similarities and differences of clustering topics between the two data sources, multidimensional indicators, such as the emerging degree, attention degree, innovation degree, and intersection degree, were comprehensively constructed for frontier identification. The methods for hot research frontiers, emerging research frontiers, extinction research frontiers, and potential research frontiers were proposed. The empirical research in the field of agricultural resources and the environment showed that the “interaction mechanism of plant–rhizosphere–microbial diversity” was a hot research frontier in the years 2016–2021. The themes of “wastewater treatment technology and efficient utilization of water resources”, the “value-added utilization of agricultural wastes and sustainable development”, the “soil ecological response mechanism under agronomic management measures”, and the “mechanism of soil landslide, erosion, degradation and prediction evaluation” were judged as potential research frontiers. The theme of “ecosystems management and pollution control of agricultural and animal husbandry” was recognized as an emerging research frontier. The results confirm that the fusion method of extracting topics from project and paper data, combined with expert intelligence and frontier indicators for fine classification of frontiers, is an optional approach. This study provides strong support for accurately identifying the forefront of scientific research, grasping the latest research progress, efficiently allocating scientific and technological resources, and promoting technological innovation.

Keywords: LDA model; research frontier; agricultural resources and the environment; frontier indicators; theme fusion



Citation: Chuan, L.; Zhao, J.; Qi, S.; Jia, Q.; Zhang, H.; Ye, S. Research Frontiers in the Field of Agricultural Resources and the Environment. *Appl. Sci.* **2024**, *14*, 4996. <https://doi.org/10.3390/app14124996>

Academic Editor: Chilukuri K. Mohan

Received: 7 May 2024

Revised: 5 June 2024

Accepted: 6 June 2024

Published: 7 June 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Global scientific and technological innovation is flourishing and is showcasing new development trends and characteristics. A deep understanding and accurate grasp of research frontier plays a vital role in the forward-looking deployment of innovation strategies and the optimal allocation of innovation resources. Frontiers represent the most novel, potential, forward-looking, and leading research directions in the process of scientific and technological innovation. Accurately identifying a frontier can effectively predict future research trends, providing strong support for scientific and technical personnel to accurately grasp the latest research progress. It also provides strong support for scientific research managers and decision-makers to accurately invest manpower, materials, and financial

resources in the direction with the most strategic value, to promote the effective allocation of resources and promote continuous progress. Therefore, research frontier identification is of great significance for gaining insight into research trends, targeting frontier fields, developing knowledge services, and promoting technological innovation.

Currently, many scholars have carried out fruitful research on frontier recognition, which can be summarized as methods based on experts' subjective judgment, methods based on objective data analysis, and methods based on content mining (Table 1).

Table 1. Summary of frontier recognition methods.

Method Classification	Methodology Used	References
Based on experts' subjective judgment	Delphi; expert knowledge and experience	[1–5]
Based on objective data analysis	Citation analysis; co-citation analysis; coupling analysis; word frequency statistics and co-word analysis; co-authorship network analysis and co-citation of authors; knowledge graphs; combining multiple methods	[6–15]
Based on content mining	Latent Dirichlet Allocation; PhraseLDA, CitationLDA, FW-LDA; LDA++	[16–22]

1.1. Recognition Method Based on Experts' Subjective Judgment

Such methods mainly include frontier identification methods coupled with a Delphi survey, scenario analysis, content analysis, science and technology policy analysis, comparative analysis, social survey, expert consultation, and technology roadmap, etc. Many related institutions have carried out research on frontiers to grasp the development direction [1]. The Delphi method fully relies on the knowledge and experience of experts, who make judgments, evaluations, and predictions about problems through investigations and research. It is an effective means of fully exploring expert perspectives [2,3]. The method, based on the subjective judgment of experts, has a long history and fully utilizes the collective wisdom and experience of domain experts. It comprehensively organizes, summarizes, and analyzes various ideas and viewpoints to make cutting-edge predictions, and the method is already quite mature. However, this type of method is time consuming, with strong subjectivity among experts, making it easier to give more attention to the field of interest. It is also susceptible to limitations in the breadth and depth of technical expert knowledge, resulting in deviations in expectations or assumptions [4]. In addition, Liu et al. put forward two new methods to identify disruptive technologies with the help of experts. One method was to obtain a list of potential disruptive technologies from experts and then evaluate the technology's disruptive potential by using a multidimensional index system. Another method was to generate a list of potential disruptive technologies by mining multi-source data and then evaluate the technology's disruptive potential through the knowledge of experts [5]. Currently, with the emergence and gradual deepening of the data-intensive knowledge innovation paradigm, the role of data analysis has become increasingly prominent and has become an important auxiliary basis for expert judgment.

1.2. Recognition Method Based on Objective Data Analysis

Scientific papers and patents record the vast majority of scientific research and technological development achievements in the world, respectively, with strong reliability and standardization, making them the preferred data sources for frontier identification. The establishment of a scientific citation index database has promoted the development of bibliometrics and has led to a gradual rise in frontier recognition. Some schools of thought have already been formed: schools of citation analysis represented by citation analysis, co-citation analysis, and literature coupling analysis [6]; schools of vocabulary analysis represented by word frequency statistics and co-word analysis [7,8]; schools of the scientific community represented by co-authorship network analysis and the co-citation

of authors [9]; schools of scientific knowledge graphs [10]; and schools of topic extraction methods, coupled with multiple methods [11,12].

Significant progress has been made in obtaining research frontiers based on citation analysis, which can be carried out from multiple dimensions such as document coupling analysis, author co-citation analysis, and mechanism co-citation analysis. A research frontier in a particular field can be identified by analyzing core documents, authors and institutions. China's frontier identification mainly focuses on using the Essential Science Indicators (ESI) database to continuously track the co-citation of highly cited papers in clusters around the world to discover research frontiers [13]. This method believes that when a cluster of highly cited papers is cited together and reaches a certain level of activity and coherence, a research frontier is formed. This cluster of highly cited papers is the core cluster of papers that constitute the research frontier, while this group of highly cited core papers together with the cited documents that cite the core papers constitute the research frontier. Using this method, the National Science Library of the China Academy of Sciences has published a research frontier report every year since 2014, and has continuously improved its methodology [14,15].

1.3. Recognition Method Based on Content Mining

The recognition method based on content mining uses models to extract topics, and researchers identify frontiers by analyzing such topics. The topic model, represented by the Latent Dirichlet Allocation (LDA) algorithm, extracts more valuable potential topic distribution through full-text semantic analysis, which makes up for the shortcomings of the citation analysis method and the vocabulary analysis method, to a great extent [16]. In recent years, with the deepening of research, the LDA model has been gradually updated and improved. PhraseLDA, CitationLDA, FW-LDA, and other models have been proposed to extract different text content. Deng and Ke [17] first used the LDA model to detect interdisciplinary topics from the text of interdisciplinary projects. Then, based on the topic's supporting documents, project funding time and funding amount, they selected the emerging interdisciplinary topics using topic novelty and topic intensity indicators. Huang et al. [18] analyzed hot topics concerning technology and content evolution in patents related to carbon capture, utilization, and storage technology using the LDA model. Ye et al. [19] used the LDA model to identify research topics and built mapping relationships on topics to both help fund projects and identify cross-domain categories; also, the evolution of research topics was analyzed. Zhang and Sun [20] analyzed the subjects' interdisciplinary characteristics and constructed a measurement index system, and then identified interdisciplinary subjects with the help of the PhraseLDA model. Nguyen et al. [21] believed that the LDA algorithm could not work effectively in regard to short text topic recognition and put forward the Citation LDA++ model. Research has shown that this model can improve the performance of the LDA algorithm by inferring topics from research papers based on the title, abstract, and citation information, and citation information can improve the performance of the LDA algorithm in regard to research paper-based topic recognition when the full text of the paper cannot be obtained. Liu et al. [22] improved the results of LDA probability topic extraction by defining the topic identifiers and introducing negative sampling models. They proposed the FW-LDA method and conducted adjacent time slice topic keyword evolution analysis in the hydrogen production field, verifying the effectiveness of the FW-LDA method. With the development of machine learning technology, scholars have developed more targeted topic mining methods, such as the dynamic topic model, related topic model, author topic model, supervised topic model, and Bayesian nonparametric model, etc. [23]. The current research on text mining based on topic models has produced many achievements for reference, and the research on topic recognition needs to explore and apply new cutting-edge and practical models in depth to improve the effect of topic recognition.

1.4. Problems

At present, one of the main problems of research frontier recognition methods is the lack of identification indicators. The existing methods for identifying research frontiers mainly rely on citation relationships, vocabulary accumulation levels, and vocabulary relationships. However, many studies lack unified discriminant indicators. Another problem is the neglect of semantic information between texts in the process of research frontier recognition, in which citation analysis and vocabulary analysis overlook the semantic correlation between the literature and the words, and lacks exploration of machine learning or topic modeling methods. The third problem is that many research frontiers are extracted from one data source, which leads to one-sided and limited results, and lacks mutual verification from multiple data sources. More importantly, the existing research on frontier identification lacks in-depth research, application, and practice in the field of agricultural technology, and its support for forward-looking research and layouts in regard to agricultural technology is not prominent enough.

1.5. Research Objectives

Projects are scientific and technological tasks organized and implemented around priority areas of relevant scientific and technological planning, which reflect the specific deployment at the current stage. Scientific papers are effective carriers of scientific research achievements. The above datasets, with rich semantics and strong effectiveness, have become important sources for some scholars to carry out frontier exploration.

Therefore, this study firstly selected project data and SCI paper data to construct a target domain dataset and adopted a topic extraction model for text semantic recognition and the extraction of research topics. Secondly, a frontier identification index system of projects and research papers was constructed. Thirdly, the agricultural field was selected to carry out research frontier discrimination. By extracting frontiers in the target field, the accuracy of frontier identification and the forward-looking value of frontier detection can be improved, which provides important support for knowledge services and technological innovation.

2. Materials and Methods

2.1. Research Ideas

The specific research ideas are as follows: project data and paper data are used as data sources, among which project data comes from the National Institute of Food and Agriculture (NIFA) and paper data comes from the Web of Science (WOS) core collection database. In terms of coverage, the research topics from the NIFA project data are all in the United States, while the WOS paper data cover the whole world. Therefore, the comparison between project data and paper data topics is mainly carried out from two dimensions: a comparison between the NIFA project data and the WOS global paper data; and a comparison of the NIFA project data and the WOS domestic paper data.

The fusion of various data sources can be carried out at an early stage, during the intermediate process, or after the completion of data processing. Considering the large difference in the number of papers and project samples, data fusion at an early stage may lead to difficulty in extracting and presenting project topics. Therefore, this study adopts the post-fusion method, that is (Figure 1), based on obtaining research topics from the project data and paper data, the indicators are comprehensively analyzed such as the emerging topic degree and the topic attention degree of the project data and paper data, and puts forward identification and judgment methods, namely hot research frontiers, emerging research frontiers, extinction research frontiers, and potential research frontiers.

The flowchart of the research process is as follows:

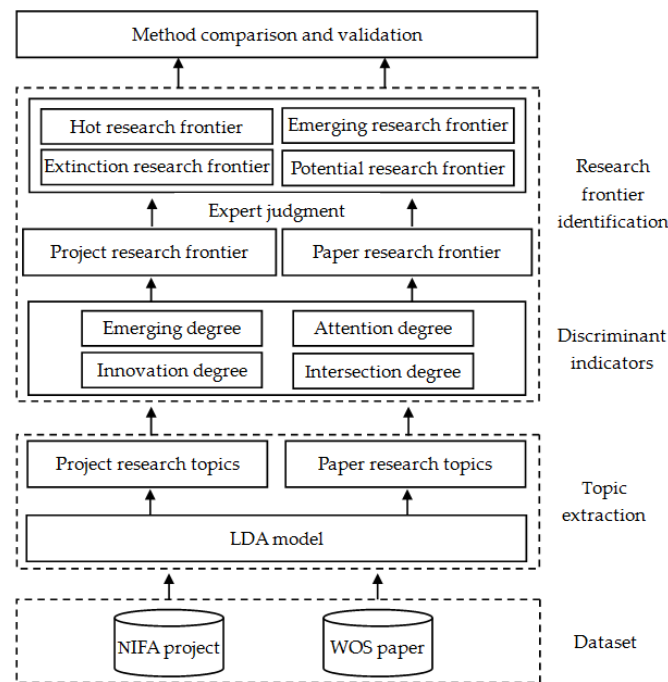


Figure 1. The flowchart of the research process.

2.2. Topic Extraction Method

The topic clustering model is a statistical model for clustering the implicit semantic structure of text using unsupervised learning, which is widely used in mining the latent semantic relationship and topic information of text. The LDA model is an unsupervised model, which was first proposed by Blei et al. in 2003 [24]. It can automatically extract semantic information from text and explore the underlying semantic associations. This model has the advantage of modeling massive heterogeneous text data and has been widely applied in scientific literature knowledge mining, scientific research hot spot discovery, topic evolution analysis, emerging frontier topic detection, and other research directions. The LDA model considers each dataset as a mixture of a group of potential topics. In the process of text modeling, each document is summarized by using topic probability distribution. In addition, the LDA model is based on the “word bag” model, which assumes that the document is a combination of word frequency and does not consider the order relationship between the words in the document, thus simplifying the complexity of the model (Figure 2).

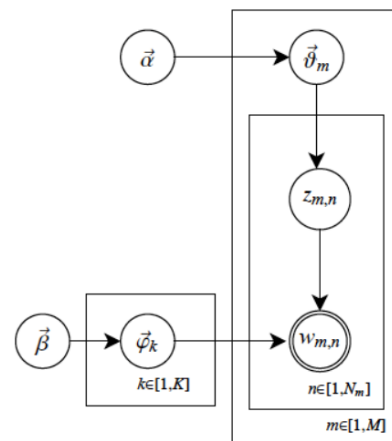


Figure 2. Inference network of the LDA model.

Firstly, a document d_i is selected according to the probability $P(d_i)$, the topic distribution θ_m of the document d_i is generated by sampling from α of the Dirichlet distribution, and the topic $Z_{m,n}$ of the j -th word of the document d_i is extracted from θ_m . From the Dirichlet distribution β , the word distribution φ_k corresponding to the topic $Z_{m,n}$ is generated by sampling, and finally the word $w_{m,n}$ is generated by sampling from the word distribution φ_k . The corpus generation probability of the LDA model is shown in Formula (1), where α is the prior distribution parameter of the topic distribution θ ; β is the prior distribution parameter of the subject word distribution φ ; w and z , respectively, represent the topic generated by the LDA model extraction; the final subject word, K , is the number of topics; and M is the total number of documents.

$$p\left(\vec{w}, \vec{z} \mid \vec{\alpha}, \vec{\beta}\right) = \prod_{k=1}^K \frac{\Delta\left(\vec{\varphi}_k + \vec{\beta}\right)}{\Delta \vec{\beta}} \prod_{m=1}^M \frac{\Delta\left(\vec{\theta}_m + \vec{\alpha}\right)}{\Delta \vec{\alpha}} \quad (1)$$

In this study, the LDA topic analysis tool is used to cluster and analyze the research topics in the above datasets. Combined with perplexity parameters, the number of research topics is determined. Perplexity is an evaluation index derived from the concept of entropy in information theory, which indicates the degree of certainty of mapping a topic to each word when a topic is given. The smaller the perplexity value, the better the classification effect of the model on new samples. The calculation formula is shown in Formula (2).

$$Perplexity(D) = \exp\left(-\frac{\sum_{d=1}^M \log p(w_d)}{\sum_{d=1}^M N_d}\right) \quad (2)$$

Among them, D represents the test dataset, M represents the number of documents, w_d represents the word set that constitutes the document set D , $p(w_d)$ refers to the probability of each word appearing in the test dataset, and N_d represents the total number of words appearing in the d -th document.

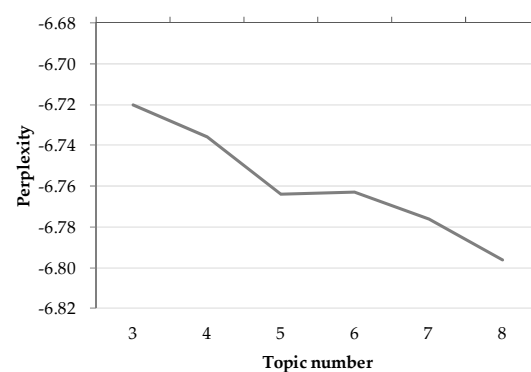
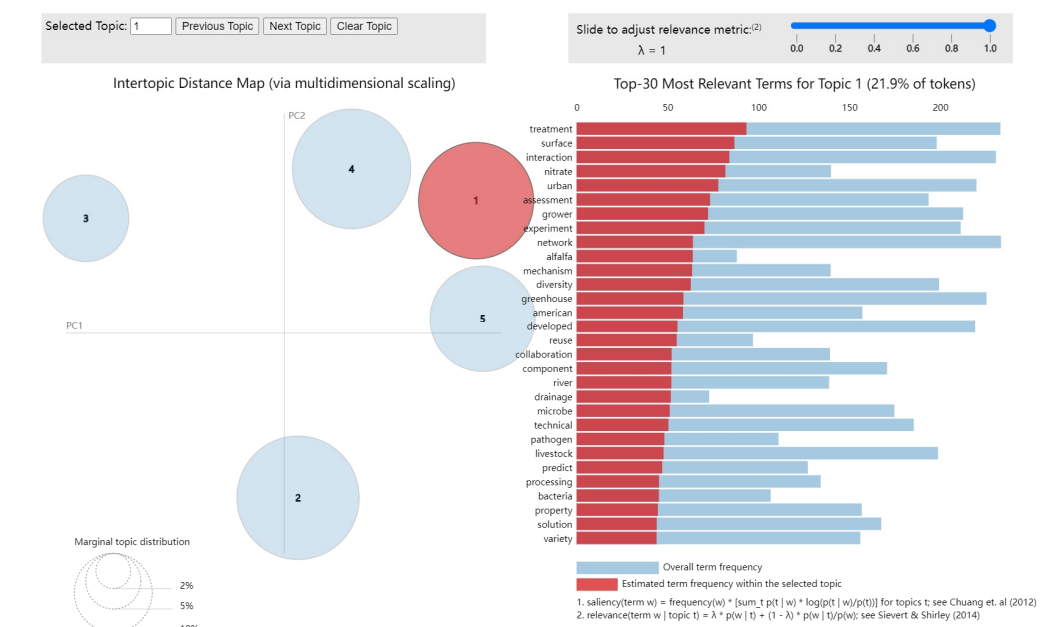
This article is based on the Gensim tool in Python, used for the LDA topic analysis. The determination of the number of topics is mainly based on two parameters: the number of topics (K) and the number of iterations (passes). The optimal number of topics is determined by combining the perplexity values and the topic model visualization. Among them, K is the initial number of topics set based on the expert's understanding of the field. The passes parameter controls the frequency of the training models across the entire corpus.

The decision-making process for the number of topics is as follows: the LDA model conducts training analysis based on the K and passes submitted each time and uses perplexity algorithms and visualization conducted by pyLDAvis to evaluate the training results [25,26]. Through expert analysis of the results, the K and passes parameters are continuously adjusted until the optimal number of topics is determined.

Taking the NIFA data as an example, through experimental analysis, the K and passes parameters were set as shown in Table 2 and Figure 3, respectively. When $K = 5$ and passes ≥ 150 , the perplexity value tended to flatten (or be the lowest), and each topic was clearly divided (Figure 4). Therefore, the optimal number of topics for the NIFA dataset in this study was ultimately determined to be 5. Using the same method, the optimal number of topics for the SCI paper dataset was determined to be 5.

Table 2. The number of topics (K), the number of iterations (passes), and the perplexity of the NIFA data.

K	Passes	Perplexity	K	Passes	Perplexity
3	50	−6.734	3	150	−6.720
4	50	−6.756	4	150	−6.736
5	50	−6.780	5	150	−6.764
6	50	−6.795	6	150	−6.763
7	50	−6.813	7	150	−6.776
8	50	−6.839	8	150	−6.796
3	100	−6.722	3	200	−6.720
4	100	−6.740	4	200	−6.735
5	100	−6.760	5	200	−6.752
6	100	−6.770	6	200	−6.760
7	100	−6.784	7	200	−6.774
8	100	−6.805	8	200	−6.792

**Figure 3.** The perplexity of the different topic numbers for the NIFA project data.**Figure 4.** The results of the best topic distribution map for the NIFA project data. Note: ^a: 1, 2, 3, 4, 5 refer to the clustering themes, ^b: Chuang et al. (2012) and Sievert & Shirley (2014) are the references [25] and [26].

It should be noted that in terms of topic model selection, we first compared the applicability of the LDA model with the LSA (Latent Semantic Analysis) model. LSA is

another important probabilistic latent semantic analysis model. Compared with LSA, the LDA model introduces Dirichlet in regard to both the document topic distribution and the topic word distribution, thereby enhancing the model's generalization ability. At the same time, using the NIFA data as test samples, the topic perplexity index was used to evaluate and determine the advantages and disadvantages of the two models. In order to better observe the comparison, the number of topics was set from 3 to 8, and the trend in the perplexity values for the two models under different numbers of topics were analyzed (Figure 5). The results showed that the perplexity value of the LDA model was lower than that of the LSA model, indicating that the LDA model was superior to the LSA model. Therefore, in this study, the LDA model was selected for topic extraction.

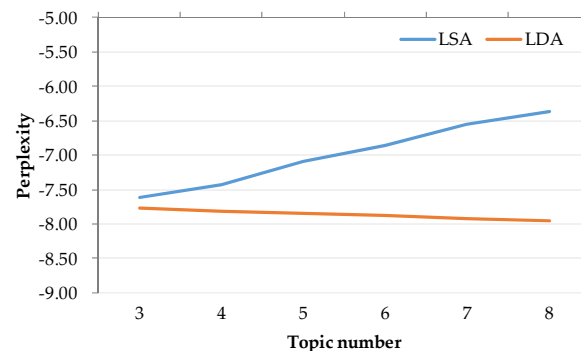


Figure 5. A comparison of the perplexity calculated by the LSA and LDA models.

2.3. Method for Frontier Categories Judgment

After using the LDA topic clustering model to identify the research topics in the project and paper data, the first step is to combine emerging, innovative, attention, and intersection indicators for scoring, and to normalize the indicators according to equal weights. Finally, the comprehensive scores for each topic are obtained. A theme with a comprehensive score above 0.50 is considered to have performed well in regard to various indicators and is identified as a research frontier topic. Then, combined with the wisdom of domain experts, it is determined whether it is a coexisting research frontier. Finally, based on the results of the four indicators mentioned above, multidimensional analysis is conducted on the identified research frontier topics, which are divided into four types: hot research frontiers, emerging research frontiers, extinction research frontiers, and potential research frontiers.

The emerging degree refers to the novelty of the research topic at the time. The more new topics that appear, the easier it is to discover the latest research content, and it is more likely to become a research frontier. The emerging degree is characterized by the average year of the project/paper in regard to the topic.

Innovation emphasizes the breakthrough and leadership in the topic content. Using Kleinberg's burst detection algorithm and infinite state automata to model time-series data, the state transition of time-series data marks the emergence of unexpected events, which is represented by the sum of the burst probability values in subject words.

The content with a high attention level can represent the development level of the field at the current stage or can influence the future development trend of the field. The attention of a project is reflected by the duration and the amount of funding granted, the number of projects, and the annual changes, as well as the proportion of projects on the theme. Papers are characterized by the average number of citations, the change in the number of papers per year, and the proportion of papers on the topic.

The intersection represents the breadth of the interdisciplinary aspect of the topic. The intersection of multiple disciplines increases the opportunities for the cross-disciplinary application of scientific research results and the probability of having an innovative impact. It is represented by the average number of research fields corresponding to all the projects/papers related to a topic.

Due to the fact that the projects have a forward-looking layout regarding scientific research by management institutions, and the papers are an embodiment of the staged achievements of scientific research, the projects are more forward-looking than the papers. Combined with the research ideas of existing scholars [27–29], in this study, the emerging degree of “coexisting topics” is calculated based on the project data, while the emerging degree of “non-coexisting topics” is calculated separately. The degree of attention of “coexisting topics” is the sum of the attention of the projects and papers, while the attention degree, innovation degree, and crossover degree of “non-coexisting topics” are calculated, respectively. The height and strength of each indicator refer to the overall average value, and those above the average value are judged as high or strong, while those less than the average value are judged as low or weak. The specific discrimination method is shown in Figure 6.

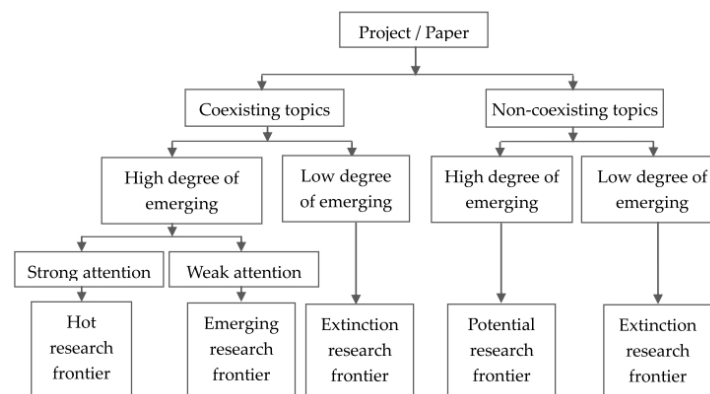


Figure 6. Research frontier discrimination method based on projects and papers.

The field of agricultural resources and the environment was selected to carry out empirical research. Due to the similarity between the clustering themes of the global SCI papers in this study and those of SCI papers in the United States, the themes of the global SCI papers could be used to represent them. Therefore, in the process of data fusion, only the NIFA fund project and the WOS global papers were fused and compared. The following situations may occur:

(1) A topic appears in the NIFA project and a WOS paper, which can be divided into the following three situations, according to the emerging degree and attention degree:

① The theme has a high degree of emerging and strong attention. It means that the theme layout is new, and the current research is hot. This kind of topic has a high degree of participation, and the research is in a rapid development stage, so this type of topic is judged as a hot research frontier;

② The theme has a high degree of emerging but weak attention. It is indicated that the theme layout is new in the particular year, but has not attracted wide attention. A layout concerning this kind of topic is of strategic significance, and the research will attract more attention and will develop further in a certain period. This type of topic will be judged as an emerging research frontier;

③ A low level of emerging themes. It is indicated that the theme layout is outdated. This type of topic is generally mature in regard to its development or has low research value, and the research direction will gradually weaken or shift in a short period, so this type of topic is judged as an extinction research frontier.

(2) A topic only appears in the NIFA fund project data or WOS paper data. Combined with the emerging degree of the topic, it can be divided into the following two situations:

① A high degree of emerging themes, representing a new theme layout in that year. Some topics have not yet achieved phased results in scientific research, but the layout of fund projects reflects its strategic significance and research value. Another part of these topics has not been laid out in the fund project, but some researchers have taken part in forward-looking research and have achieved initial results. These topics will have a

greater possibility to develop well, and this type of topic will be judged as a potential research frontier;

② A low degree of emerging themes, indicating outdated theme layout years. This kind of topic is generally mature in regard to its development or has low research value, and its research direction will gradually weaken or shift in a short period, so this type of topic is judged as an extinction research frontier.

3. Results

3.1. Frontier Extraction Results from a Single Data Source

The LDA topic analysis tool was used to cluster the NIFA project data and SCI paper data in the field of agricultural resources and the environment from 2016 to 2021, and five research topics were obtained, respectively. Then, combining emerging, innovative, attention, and intersection indicators for scoring, and normalizing the indicators according to equal weights, comprehensive scores of each topic were finally obtained. Subjects with scores above 0.50 were considered to have performed well in regard to various indicators and were identified as research frontier topics. So, the research frontiers from the project data and SCI paper data in the field of agricultural resources and the environment from 2016 to 2021 are shown in Tables 3 and 4. After interpretation by field experts, it was proposed that the research based on the project data mainly focused on four directions: wastewater treatment technology and the efficient utilization of water resources, the value-added utilization of agricultural wastes and the sustainable development of agriculture, the plant–rhizosphere–microorganism interaction and diversified farmland management, and ecosystems management and pollution control related to agricultural and animal husbandry. The research based on the SCI paper data mainly focused on three directions: the soil ecological response mechanism under agronomic management measures; the interaction mechanism of plant–rhizosphere–microbial diversity; and the mechanism of soil landslide, erosion, degradation, and prediction evaluation.

Table 3. Research frontiers in the NIFA project in the field of agricultural resources and the environment from 2016 to 2021.

Research Frontiers	Subject Words
Theme 1 Wastewater treatment technology and the efficient utilization of water resources	treatment; surface; interaction; nitrate; urban; grower; assess; experiment; network; alfalfa; mechanism; divergence; greenhouse; American; developed; reuse; component; river; collaboration; livestock
Theme 2 Value-added utilization of agricultural waste and the sustainable development of agriculture	grain; biomass; sensor; groundwater; plastic; regional; resilience; emission; adaptation; wheat; business; policy; resistance; network; fiber; scholar; urban; developed; renewable; variety
Theme 3 Plant–rhizosphere–microorganism interaction and diversified farmland management	vegetable; plain; efficiency; soybean; business; cotton; bioenergy; grower; measurement; canola; marketing; stress; quantify; transfer; ecologic; microbiome; Navajo; southern; assess; livestock; watershed; antibiotic; biochar
Theme 4 Ecosystems management and pollution control related to agricultural and animal husbandry	temperature; microbiome; efficiency; biomass; conservation; stress; microbe; sense; divergence; dairy; forage; weather; building; policy; habitat; developed

In this study, the clustering topics from the global SCI papers in the field of agricultural resources and the environment were similar to those in the United States, which could be characterized by the topics in the global SCI papers.

Table 4. Research frontiers in the SCI papers on global agricultural resources and the environment from 2016 to 2021.

Research Frontiers	Subject Words
Theme 1 Soil ecological response mechanism under agronomic management measures	manure, diversity, straw, residue, fertilization, stock, phosphorus, enzyme, respiration, amendment, grassland, rotation, sequestration, compost, mineralization, stability, mulch, fertility, availability, labile
Theme 2 Interaction mechanism of plant–rhizosphere–microbial diversity	rhizosphere, decomposition, bacteria, fungi, diversity, fungal, grazing, mycorrhizal, nitrification, grassland, interaction, nematode, inoculation, trait, denitrification, micro-organization, functional, availability, strain, arbuscular
Theme 3 Mechanism of soil landslide, erosion, degradation, and prediction evaluation	slope, estimate, rainfall, runoff, prediction, river, parameter, index, prediction, density, movement, character, variation, profile, physical, variability, measurement, variable, environmental, landscape

3.2. Research and Judgment on Different Frontier Types

This study conducted artificial professional interpretations for the frontiers obtained from the NIFA project and the SCI papers based on the LDA model. The topic on the “interaction mechanism of plant–rhizosphere–microbial diversity” was determined as the co-existent research frontier in the project and the paper. The “wastewater treatment technology and efficient utilization of water resources”, the “value-added utilization and sustainable development of agricultural waste”, the “ecosystems management and pollution control related to agricultural and animal husbandry”, the “soil ecological response mechanism under agronomic management measures”, and the “mechanism of soil landslide, erosion, degradation, and prediction evaluation” were identified as non-coexisting research frontiers. Indicators such as the emerging and attention degree of each frontier are shown in Table 5.

By calculating the average value of each indicator, the average emerging degree was 2018.50. The average innovation level was 0.20. The average attention value of the project fund was 0.946 and for the papers was 9.778. The cumulative average of the project fund/paper attention value was 10.724. The average project crossover degree was 6.479 and for the papers was 1.974.

By comparing the indicators of the abovementioned coexisting themes and non-coexisting themes with the average value, the coexisting theme of the “interaction mechanism of plant–rhizosphere–microbial diversity” had a high degree of emergence, attention, innovation, and cross-cutting, which indicated that this theme was a hot research frontier in recent years. The themes of “wastewater treatment technology and the efficient utilization of water resources”, the “value-added utilization of agricultural wastes and sustainable development” in the NIFA project layout, as well as the themes of the “soil ecological response mechanism under agronomic management measures” and the “mechanism of soil landslide, erosion, degradation, and prediction evaluation” extracted from the SCI papers, had higher levels of novelty, innovation, and attention than the average value, indicating that the layouts concerning such research topics were relatively new. The layout of fund projects reflects their strategic significance and research value. However, there is no high-output stage achievement within scientific research, or alternatively some topics have not been laid out in fund projects; however, some researchers have carried out forward-looking research and achieved phased results, which means that some topics will have a greater possibility of further development. Therefore, the above four topics were judged as potential research frontiers.

Table 5. Comparative analysis of frontiers between the NIFA project and global SCI papers.

Category	Research Frontiers	Emerging	Innovation	Attention	Cross	Comprehensive Score	Frontier Type
1 ^a	Interaction mechanism of plant–rhizosphere–microbial diversity	2018.64	0.287	11.098	6.518	0.971	Hot frontier
	Wastewater treatment technology and the efficient utilization of water resources, NIFA project	2018.68	0.311	0.95	6.389	0.863	Potential frontier
	Value-added utilization and sustainable development of agricultural waste, NIFA Project	2018.74	0.251	0.95	7.090	0.930	Potential frontier
2 ^b	Ecosystems management and pollution control related to agricultural and animal husbandry, NIFA project	2018.74	0.189	0.94	6.401	0.672	Emerging frontier
	Soil ecological response mechanism under agronomic management measures, SCI paper	2018.73	0.196	12.362	1.65	0.980	Potential frontier
	Mechanism of soil landslide, erosion, degradation, and prediction evaluation, SCI Paper	2018.67	0.254	10.423	1.92	0.855	Potential frontier

Note: ^a: 1 refers to a coexisting theme, ^b: 2 refers to a non-coexisting theme.

The theme of “ecosystems management and pollution control related to agricultural and animal husbandry” laid out by the NIFA project, had the highest emerging degree, but its attention was slightly lower than the average value, which indicated that although the layout of this research theme by US project management departments was relatively new in the particular year, it had not attracted high attention due to the short time since its publication, and had been recognized as an emerging research frontier. In the future, with the continuous layout of projects and the output of related results, it will attract high attention and will have the opportunity to become a potential research frontier or hot research frontier.

3.3. Verification of Frontier Identification Method

The obtained research frontier was compared with the global agricultural research hot spot frontier results in 2022 developed by the Institute of Agricultural Information, within the Chinese Academy of Agricultural Sciences, to verify the scientific nature of this study, which was based on the LDA theme model and frontier indicator system. Based on the data in the SCI papers and CSCD papers on the Web of Science platform (2017–2021), the Chinese Academy of Agricultural Sciences extracted highly co-cited papers (i.e., papers with the top 1% citation frequency in the same year) from the ESI database, and the core topic cluster composed of highly co-cited papers and cited papers was regarded as the hot spot frontier. Through paper classification mapping, combined with co-citation theory, the semantic clustering method, index screening, expert interviews, and qualitative and quantitative analysis, 71 agricultural research hot spots in 9 major agricultural disciplines in 2022 were selected from the initial 989 agricultural topic data. Among them, the hot frontiers in the field of agricultural resources and the environment mainly focus on the rhizosphere ecological environment, soil carbon and nitrogen fixation mechanism, fertilizer

research and development, plant beneficial bacteria resistance, and soil-crop management. The results are shown in Table 6.

Table 6. Research hot spots and frontiers in the field of agricultural resources and the environment proposed by the Chinese Academy of Agricultural Sciences.

Number	Category	Research Hot Spot or Frontier
1	Hot spot	Influence of soil-crop management on greenhouse gas emissions and its regulation
2	Hot spot	Response and mechanism of soil and rhizosphere microorganisms compared to land use patterns
3	Hot spot	Research and efficient utilization of new fertilizers
4	Hot spot	Mechanism and regulation of nitrogen fixation in plants
5	Hot spot	Effect and mechanism of growth-promoting and stress-resistant plant beneficial bacteria
6	Hot spot	Assembly patterns and control mechanism of the root microbiome
7	Key hot spot	Soil carbon sequestration potential and its regulation mechanism
8	Frontier	Microbial mechanism of soil nitrogen transformation and its agronomic environmental effects

The results showed that the frontier concerning the “interaction mechanism of plant–rhizosphere–microbial diversity” proposed in this study covers many research hotspots, such as the “response and mechanism of soil and rhizosphere microorganisms compared to land use patterns”, the “effect and mechanism of growth-promoting and stress-resistant plant beneficial bacteria”, the “assembly patterns and control mechanism of the root microbiome”, and the “microbial mechanism of soil nitrogen transformation and its agronomic environmental effects”, which were similar to the research hot spots obtained by Chinese Academy of Agricultural Sciences based on the data from the ESI highly co-cited papers. The emerging research frontier “ecosystems management and pollution control related to agricultural and animal husbandry” proposed by this study included the abovementioned hot frontier the “influence of soil-crop management on greenhouse gas emissions and its regulation”, The potential research frontier on the “soil ecological response mechanism under agronomic management measures” proposed by this study also covered the above-mentioned hot frontiers concerning the “mechanism and regulation of nitrogen fixation in plants” and “soil carbon sequestration potential and its regulation mechanism”. The potential frontier concerning the “value-added utilization and sustainable development of agricultural wastes” proposed in this study also involved some of the above-mentioned content, namely the “research and efficient utilization of new fertilizers”. Therefore, it can be inferred that the research frontier mining methods based on the LDA model and frontiers recognition system using the SCI papers and project integration in this study are feasible. In addition, combined with specific discriminant indicators, this study also conducted multidimensional analysis of the obtained research frontiers, and further divided them into four types: hot research frontiers, emerging research frontiers, extinction research frontiers, and potential research frontiers.

4. Discussion

4.1. About the Theoretical Implications

This study has deepened and expanded the knowledge on agricultural resources and the environment by applying new methods and technologies, such as text mining tools involving LDA. By systematically analyzing a large amount of textual data, it revealed the research frontiers in this field, and provided new insights and supplements to the existing knowledge system. This study put forward the hot research frontiers, emerging research frontiers, extinction research frontiers, and potential research frontiers in the field of agricultural resources and the environment. It can provide strong support for the

research direction and topic selection of frontline researchers in the field of agricultural resources and the environment. Researchers still need to focus on in-depth research in the areas of plant–soil–microbial diversity, water treatment and utilization, agricultural wastes utilization, soil ecology and quality conservation, and ecological pollution and remediation.

The use of methods such as LDA and text mining tools is of great significance in supporting researchers to conduct research in a target field. These tools can provide an effective data analysis method to help researchers extract useful information from a large amount of text. By applying these methods, researchers can gain a deeper understanding of the research direction, providing strong support for future research topics. This study proposed a method for topic clustering based on the LDA model and verified the applicability of the LDA model in topic extraction. This model can automatically extract implicit topic structures from large-scale text sets through unsupervised learning, providing powerful tools for organizing, understanding, and applying text data.

By using LDA and other machine learning tools, existing knowledge about agricultural resource utilization and environmental maintenance can be further improved. These tools can help researchers more accurately identify and classify themes in text data, providing more accurate analysis results. In addition, machine learning can also be used to predict and optimize the utilization of agricultural resources, as well as formulate corresponding environmental protection strategies. This study provides new ideas and methods for sustainable development by delving into the relationship between agricultural resource utilization and the environment. By optimizing the utilization of agricultural resources and reducing their impact on the environment, the sustainable development of agriculture can be promoted while protecting ecosystems and natural resources.

This study provides a new approach for identifying frontier topics and theoretical and technical support for the selection of scientific research directions and technological innovation. With the acceleration of globalization, there is an increasing amount of cross-domain and cross-language textual data. How to effectively integrate these data and use the LDA model or other tools associated with text mining for topic extraction will be a research direction with practical significance. By exploring methods such as domain adaptation and cross-language learning, the scope of application of the LDA model or other machine learning tools can be expanded to better serve text analysis tasks in multiple domains and languages.

4.2. About the Practical Implications

Using LDA and other artificial intelligence tools to assist governments and regulatory agencies in developing better strategies for agricultural resource utilization, while considering environmental protection, will help improve the efficiency and sustainability of agricultural production. By accurately analyzing and predicting the utilization of agricultural resources and environmental impacts, governments and regulatory agencies can formulate more scientific and reasonable policies to promote the sustainable development of agriculture.

Managers in the agricultural sector can utilize the research directions proposed to better improve the methods and measures of agricultural production processes. By applying text mining and topic mining techniques, managers can systematically analyze and evaluate existing agricultural production technologies, identify problems and bottlenecks, and propose targeted improvement measures.

The use of topic mining (as well as general text mining) can help to ensure that management personnel related to agricultural resource extraction and utilization can observe the impact of their activities on the environment and establish practical and feasible measures, such as establishing monitoring and feedback mechanisms. By regularly collecting and analyzing relevant textual data, such as agricultural activity reports, and environmental monitoring data, etc., the impact of agricultural resource extraction on the environment can be discovered and evaluated in a timely manner, and feedback

and suggestions can be provided to management personnel to promote environmentally friendly agricultural practices.

4.3. About the Limitations and Difficulties

However, there are still some shortcomings related to the methods of topic extraction, discriminant indicators, and data source selection. The inherent limitations and difficulties in this study include the available time selected, differences in geographic space, the insufficient number of texts that make up the corpus, and limitations to the technology used. These factors may lead to limitations and inaccuracies in the research results, which need to be overcome and improved in subsequent studies.

In the process of applying topic extraction models, problems such as low data quality, and the overlapping or blurring of topics may be encountered. To overcome these difficulties, we took a series of measures, such as modifying the data extraction methods, optimizing data processing, improving model parameter settings, and introducing more external knowledge, etc., to improve the accuracy and reliability of the topic extraction. In this study, the number of topics determined was the first and most important step, and it mainly related to the number of iterations. We continuously adjusted the two above parameters and used perplexity algorithms and visualization conducted by pyLDAvis to evaluate the training results, until the optimal number of topics was determined. In addition, we also screened for keywords in papers in the field of agricultural resources and the environment and added them to the corpus to enrich its content, in order to achieve better thematic effects. At the same time, relevant experiences and research results from other fields were also borrowed to provide more ideas and methods for overcoming these difficulties.

In this study, topic extraction was based on the LDA model. Although the topic extracted by the LDA algorithm largely compensated for the shortcomings of citation analysis and vocabulary analysis methods, the current topic recognition results still have more single words and fewer keywords in the form of phrases, and the expression of the topic content was not rich enough. In addition, the current data sources only involve paper data and project data, which represent the theme of basic research, but do not consider and integrate data sources such as planning texts, policy texts, and patents representing technology research and development.

In the future, it is necessary to increase the types of data sources and conduct extensive and in-depth analysis. Combining the tools and methods with the semantic analysis function of the text content, it should strengthen the preprocessing of datasets, improve the custom dictionary and the accuracy of the subject phrase recognition, in order to put forward more precise frontiers. The selection of indicators for each dimension still needs to be further optimized, and the weight of the indicators needs to be scientifically improved in combination with the characteristics and contributions of data sources. By increasing the multidimensional data sources, optimizing the topic extraction models, and improving the judgment indicators, comprehensive support can be provided for scientific research and management layout decisions.

5. Conclusions

Based on the project and paper datasets, this study investigated the method of topic extraction by the LDA model. By combining the wisdom of domain experts to judge the similarities and differences in topics, the frontier types were identified comprehensively using indicators such as the emerging degree, attention degree, innovation degree, and intersection degree of the topic. It put forward the identification and judgment methods for hot research frontiers, emerging research frontiers, extinction research frontiers, and potential research frontiers, and also carried out empirical research in the field of agricultural resources and the environment.

The results showed that the “interaction mechanism of plant–rhizosphere–microbial diversity” was a hot research frontier in the years 2016–2021. The themes of “wastewater treatment technology and the efficient utilization of water resources”, the “value-added

utilization of agricultural wastes and sustainable development”, the “soil ecological response mechanism under agronomic management measures”, and the “mechanism of soil landslide, erosion, degradation, and prediction evaluation” were judged as potential research frontiers. The theme of “ecosystems management and pollution control related to agricultural and animal husbandry” was recognized as an emerging research frontier. The research proved that the method of extracting topics from project and paper data, combined with expert wisdom and frontier indicators for frontier identification, is an alternative method. It helps to improve the accuracy of frontier identification and the forward-looking value of frontier detection, which provides important support for knowledge services and technological innovation.

However, there are still some shortcomings in the methods of topic extraction, discriminant indicators, and data source selection, which may lead to limitations and inaccuracies in the research results. Although the topics extracted by the LDA model largely compensate for the shortcomings of citation analysis and vocabulary analysis methods, the current topic recognition results still need to be improved. In addition, the current data sources only involve paper data and project data, and they do not consider and integrate data sources such as planning texts, policy texts, and patents representing technology research and development. In the future, it is necessary to increase the multidimensional data sources available, optimize the topic extraction models, and improve the judgment indicators, so that comprehensive support can be provided for scientific research and management layout decisions.

Author Contributions: Conceptualization, L.C., H.Z. and S.Y.; methodology, L.C., H.Z. and S.Q.; validation, L.C., J.Z. and Q.J.; formal analysis, L.C. and J.Z.; investigation, L.C. and H.Z.; data curation, L.C. and J.Z.; writing—original draft preparation, L.C. and J.Z.; writing—review and editing, Q.J. and S.Q.; supervision, S.Y. and H.Z.; project administration, L.C., J.Z. and H.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This study was funded by the “Science and Technology Innovation Project in Beijing Academy of Agriculture and Forestry Sciences”, grant numbers KJCX20230208, KJCX20240313, and KJCX20240311, and the “Open Research Fund Program of Key Laboratory of Knowledge Mining and Knowledge Services in Agricultural Converging Publishing, National Press and Publication Administration”, and grant number 2023KMKS01.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The raw data supporting the conclusions of this article will be made available by the author (chuanlm@agri.ac.cn) on request.

Acknowledgments: We thank the National Institute of Food and Agriculture (NIFA) and the Web of Science (WOS) core collection database for providing the data.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Li, X.; Ma, X.D.; Feng, Y. Early identification of breakthrough research from sleeping beauties using machine learning. *J. Informetr.* **2024**, *18*, 101517. [\[CrossRef\]](#)
2. Savic, L.C.; Smith, A.F. How to conduct a Delphi consensus process. *Anaesthesia* **2023**, *78*, 247–250. [\[CrossRef\]](#)
3. Zeng, W.; Yan, T.T.; Liu, X.L. A detection method for science and technology frontiers based knowledge tree. *Inf. Stud. Theory Appl.* **2024**, *47*, 158–162.
4. Khodyakov, D.; Grant, S.; Kroger, J.; Gadwah-Meaden, C.; Motala, A.; Larkin, J. Disciplinary trends in the use of the Delphi method: A bibliometric analysis. *PLoS ONE* **2023**, *18*, 0289009. [\[CrossRef\]](#)
5. Liu, X.W.; Wang, X.Z.; Lyu, L.C.; Wang, Y.P. Identifying disruptive technologies by integrating multi-source data. *Scientometrics* **2022**, *127*, 5325–5351. [\[CrossRef\]](#)
6. Castanha, R.G.; Grácio, M.C.C.; Perianes-Rodríguez, A. Co-citation analysis between coupler authors of a scientific domain’s citation identity: A case study in scientometrics. *Scientometrics* **2024**, *129*, 1545–1566. [\[CrossRef\]](#)
7. Wen, C.; Liu, W.; He, Z.H.; Liu, C.Y. Research on emergency management of global public health emergencies driven by digital technology: A bibliometric analysis. *Front. Public Health* **2023**, *10*, 1100401. [\[CrossRef\]](#)

8. Zhang, T.; Chen, J.; Lu, Y.; Yang, X.Y.; Ouyang, Z.L. Identification of technology frontiers of artificial intelligence-assisted pathology based on patent citation network. *PLoS ONE* **2022**, *17*, 0273355. [[CrossRef](#)]
9. Xu, S.S.; Liu, J.B.; Li, S.N.; Yang, S.; Li, F.N. Exploring and visualizing research progress and emerging trends of event prediction: A survey. *Appl. Sci.* **2023**, *13*, 13346. [[CrossRef](#)]
10. Xu, S.S.; Liu, S.R.; Jing, C.F.; Li, S.N. Event knowledge graph: A review based on scientometric analysis. *Appl. Sci.* **2023**, *13*, 12338. [[CrossRef](#)]
11. da Silva, P.B.V.; Brenelli, L.B.; Mariutti, L.R.B. Waste and by-products as sources of lycopene, phytoene, and phytofluene—Integrative review with bibliometric analysis. *Food Res. Int.* **2023**, *169*, 112838. [[CrossRef](#)]
12. Gao, N.; Zhou, Q.S. Research fronts identification and evolution trend analysis of information science based on co-citation method. *J. Mod. Inf.* **2024**, *44*, 3–19.
13. Xie, X.F.; Wang, Q.; Chen, T.; Huang, F. Study on frontier development trends in Neuroscience: An analysis based on deep mining of ESI research fronts. *World Sci.-Technol. Res. Dev.* **2023**, *45*, 63–76.
14. Institutes of Science and Development; National Science Library; Chinese Academy of Sciences; Clarivate Analytics. *Research Fronts 2022*; Clarivate Analytics: Beijing, China, 2022.
15. Institutes of Science and Development; National Science Library; Chinese Academy of Sciences; Clarivate Analytics. *Research Fronts 2023*; Clarivate Analytics: Beijing, China, 2023.
16. Ma, R.; Kim, Y.J. Tracing the evolution of green logistics: A latent dirichlet allocation based topic modeling technology and roadmapping. *PLoS ONE* **2023**, *18*, 0290074. [[CrossRef](#)]
17. Deng, Q.P.; Ke, J.X. Identifying emerging interdisciplinary topics based on the fund project data: A case study of quantum technology. *Libr. Inf. Serv.* **2023**, *67*, 130–141.
18. Huang, L.C.; Hou, Z.M.; Fang, Y.L.; Liu, J.H.; Shi, T.L. Evolution of CCUS Technologies Using LDA Topic Model and Derwent Patent Data. *Energies* **2023**, *16*, 2556. [[CrossRef](#)]
19. Ye, G.H.; Wang, C.C.; Wu, C.; Peng, Z.; Wei, J.Y.; Song, X.Y.; Tan, Q.T.; Wu, L.Q. Research frontier detection and analysis based on research grants information: A case study on health informatics in the US. *J. Informetr.* **2023**, *17*, 101421. [[CrossRef](#)]
20. Zhang, Z.Q.; Sun, W. Interdisciplinary subject recognition based on feature measurement and PhraseLDA model—Case study of nanotechnology in agricultural environment. *Data Anal. Knowl. Discov.* **2023**, *7*, 32–45.
21. Neuyen, T.; Do, P. Citation LDA plus plus: An extension of LDA for discovering topics in document network. In Proceedings of the 9th International Symposium on Information and Communication Technology (SoICT), Da Nang, Vietnam, 6–7 December 2018; pp. 31–37.
22. Liu, J.X.; Zhang, Z.Y.; Wang, F. FW-LDA combination improvement for patent topics words and keywords evolution analysis. *J. Intell.* **2022**, *41*, 57–64.
23. Liang, S.; Liu, X.P. Research progress on topic evolution of scientific and technical literatures based on text mining. *Libr. Inf. Serv.* **2022**, *66*, 138–149.
24. Blei, D.M.; Ng, A.Y.; Jordan, M.I. Latent Dirichlet allocation. *J. Mach. Learn. Res.* **2003**, *3*, 993–1022.
25. Chuang, J.; Manning, C.D.; Heer, J. Termite: Visualization techniques for assessing textual topic models. In Proceedings of the International Working Conference on Advanced Visual Interfaces (AVI '12). Association for Computing Machinery, Capri Island, Italy, 21–25 May 2012; pp. 74–77.
26. Sievert, C.; Shirley, K.E. LDAvis: A method for visualizing and interpreting topics. In Proceedings of the Workshop on Inter-active Language Learning, Visualization, and Interfaces, Baltimore, MD, USA, 27 June 2014; pp. 63–70.
27. Liu, B.W.; Bai, R.J.; Zhou, Y.T.; Wang, X.Y. Identifying frontier topics from funding and paper—Case study of carbon nanotube. *Data Anal. Knowl. Disc.* **2019**, *3*, 114–122.
28. Bai, R.J.; Liu, B.W.; Leng, F.H. Frontier identification of emerging scientific research based on multi-indicators. *J. China Soc. Sci. Tech. Inf.* **2020**, *39*, 747–760.
29. Niu, X.R. *Research on Identification of Research Fronts Based on Policy Texts and Funded Projects*; The University of the Chinese Academy of Sciences (Nature Science Library): Beijing, China, 2022.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.