A Cross-source Topic Fusion and Multi-dimensional Synergistic Indicator Approach for Emerging Technology Identification

Xueli Yu
Business School
Shandong University of Technology
Zibo, China
e-mail:15063009091@163.com

Haiyun Xu
Business School
Shandong University of Technology; Max Planck Institute for Solid State Research
Zibo, China; Stuttgart, Germany
e-mail: xuhaiyunnemo@gmail.com

Robin Haunschild
Max Planck Institute for Solid State Research
Stuttgart, Germany
e-mail: R.Haunschild@fkf.mpg.de

Zhengyin Hu
National Science Library
Chinese Academy of science
Chengdu, China
e-mail: huzy@clas.ac.cn

Zenghui Yue
School of Medical Information Engineering
Jining Medical University
Rizhao, China
e-mail: yzh66123@126.com

Chunjiang Liu
National Science Library
Chinese Academy of science
Chengdu, China
e-mail: liucj@clas.ac.cn

***Abstract*—Emerging Technologies (ETs) play a crucial role in scientific revolutions and industrial transformation. Accurate identification of ETs contributes to effective national policymaking and the rapid advancement of science and technology. However, existing studies primarily rely on single data sources, leading to limitations in timeliness and comprehensiveness of identification results. To address these shortcomings, this study proposes a novel approach for emerging technology identification that integrates multi-source data fusion with coordinated multi-dimensional indicators. First, heterogeneous multi-source data are collected, and candidate technology topics are extracted using BERTopic-based topic modeling. Second, semantic similarity analysis is employed to fuse technology topics across different data sources, generating a comprehensive set of candidate ETs and constructing an indicator system for their identification. Finally, ET topics are screened and identified. The results are subsequently validated. An empirical analysis in the smart grid domain identifies nine ETs. The findings provide important references for technology forecasting, policy formulation, and industrial strategic planning.**

Keywords-Emerging Technologies; Multi-source data; BERTopic; Smart grid.

I. INTRODUCTION

Against the backdrop of rapid technological iteration, Emerging Technologies (ETs) serve as the core force driving the development of new-quality productive forces and industrial transformation. Conducting forward-looking identification of ETs is an important prerequisite for decision-makers to seize strategic opportunities [1]. With advances in computing and digital technologies, ETs evolve at an exponential rate and in a combinatorial manner. While characterized by radical novelty, relatively rapid growth, and potential impact [2], technological systems also demonstrate an evolutionary feature of coexisting high uncertainty and path dependence [3].

During the life cycle of ETs—from basic research to applied development, and then to commercialization and scaling-up—not every technology can successfully cross the "valley of death" [4]. Therefore, identifying ETs with substantial developmental potential amid a highly uncertain technological landscape is a core issue in current research.

In addition, most existing technology identification studies focus on papers and patents as research objects [5], leading to relatively one-sided analytical results. Moreover, the publication of scientific papers and patents entails a certain time lag, which tends to compromise the timeliness of prediction outcomes [6], while research on multi-source heterogeneous data remains insufficient [7][8]. In contrast, data from funding, industry, policies, and reports can reflect the dynamic changes of ETs at different levels, including early-stage layout, application-driven development, and policy orientation.

Accordingly, to comprehensively capture the trajectory characteristics of technological leapfrogging, the joint analysis of multi-source heterogeneous data helps to construct a more real-time, comprehensive, and refined picture of technological evolution. It is particularly critical to conduct an integrated analysis using multi-source data containing multi-level technical information to identify ETs [9].

Against this background, we propose an ET identification method that integrates multi-source data and multi-dimensional features. The main contributions are as follows: integrating multi-source heterogeneous data including papers, patents, funding data, industry reports, industrial market data and policy documents; depicting the comprehensive characteristics of ETs from multiple dimensions; and promoting more accurate identification of ETs that reflects their scientific foundation and application potential.

Nevertheless, this study still has several limitations. First, the proposed framework mainly relies on static cross-sectional data and lacks dynamic analysis of technology evolution over time. Second, although multi-source heterogeneous data are integrated, differences in data quality and coverage may affect the completeness of topic extraction results. In addition, the uncertainty dimension mainly focuses on structural entropy within topic networks, while external factors, such as market and policy changes are not fully considered. These limitations can be further addressed through richer data integration and dynamic evolutionary analysis in future research.

This paper is organized as follows. Section I introduces the research background, research motivation, and main contributions of this study. Section II reviews the concepts, characteristics, and existing identification methods of emerging technologies. Section III presents the research design, including multi-source data acquisition, BERTopic-based topic extraction and fusion, and the construction of the ET identification indicator system. Section IV conducts an empirical analysis in the smart grid domain and reports the identification and validation results of emerging technology topics. Finally, Section V concludes the paper and discusses the limitations and future research directions.

II. RELATED RESEARCH

This section systematically reviews the existing literature on emerging technologies, focusing on their conceptual definitions, characteristic frameworks, and mainstream identification approaches, so as to clarify the research gaps that the present study aims to address.

A. *Concepts and Connotations of Emerging Technologies*

At present, the academic community has not yet given a unified definition of "ET". Initially proposed by the Wharton School of the University of Pennsylvania, it is believed that ET is rooted in scientific discoveries, which can not only spawn new industries but also reshape existing ones [10]. On this basis, scholars have further supplemented and defined it from the dimensions of technical characteristics and technical effects combined with the theories of G.S. Day et al. [11]-[13]. In addition, compared with cutting-edge technologies and core technologies, ETs emphasize more on the early scientific foundation and potential impact, and show significant growth potential in the prototype stage of its life cycle.

B. *Feature Framework and Identification Dimensions of Emerging Technologies*

ETs occupy an important position in knowledge breakthroughs, industrial transformation, and future competition, and have long been the focus of research in the fields of technology forecasting and scientometrics. Existing studies can be roughly divided into two categories: One is the identification path based on "method tools", and the other is the identification path based on "feature framework".

In the research from the perspective of method tools, traditional scientometric methods mainly rely on the citation relationship between documents or the co-occurrence relationship of keywords, including direct citation analysis, co-citation analysis, citation coupling, co-word analysis, and overlay mapping [14]. These methods can effectively depict the research hotspots, knowledge structure, and development context of technologies, and are the basic tools for identifying technological frontiers with the advancement of technology and the expansion of identification tools. Different from the research path based on method tools, another group of scholars starts from "feature attributes" and understands ETs as a set of identifiable and measurable features. A representative one is the five-feature model proposed by Rotolo et al. [2], which holds that ETs have radical novelty, relatively rapid growth, coherence, significant impact, and uncertainty and ambiguity. This model provides a systematic conceptual framework for identifying and understanding ETs, clarifying five key dimensions, but the model itself focuses on conceptual elaboration. In recent years, methods for identifying ETs using text mining have been widely applied due to their high efficiency and accuracy [15]. Among them, the ET identification method based on topic modeling has gradually become the mainstream. It forms topics by clustering semantically related or similar technical keywords, and identifies topics that meet the attributes of ETs by analyzing their evolutionary characteristics, growth rate, and structure [16]-[18].

In addition, most existing studies adopt a single data source for technology identification, which has certain limitations. Methods for identifying ETs by fusing multi-source data have attracted more and more attention. By integrating data, such as academic papers, patents, and news reports, the development trend and market potential of ETs can be captured more comprehensively and accurately [19].

In summary, a relatively complete conceptual recognition and indicator identification system has been formed for ET research, laying a solid theoretical foundation for the identification of ETs. However, existing studies still have limitations in the identification and analysis of ETs: The data sources are relatively single, most studies only rely on papers and patents for identification, and lack the integration of multi-source data, such as industrial market data, reports, and funding, leading to insufficient representativeness of identification results and inadequate guidance for policy formulation and enterprise development. In view of this, we propose an ET identification method based on the fusion of multi-source data and multi-dimensional features. It obtains technical topics through topic modeling, constructs an ET identification indicator system, and systematically identifies ETs, in order to provide support for policy implementation and enterprises to gain a leading position in development.

III. RESEARCH DESIGN

This section elaborates the overall research framework and technical routes of this study, including multi-source data processing, topic extraction and fusion, and the construction of a multi-dimensional indicator system for emerging technology identification.

A. Research Design

We construct a research framework that mainly consists of three core modules, as shown in Figure 1.

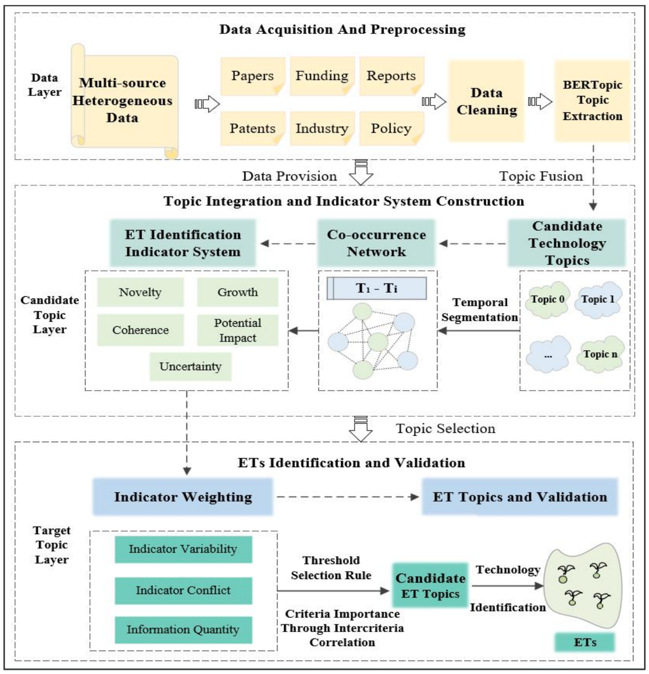

Figure 1. Research Framework Diagram.

First, the acquisition and preprocessing of multi-source heterogeneous data are carried out. Through the construction of a search query system, papers, patents, funding data, industry reports, industrial market data and policy documents are collected. After manual cleaning and screening, as well as standardized preprocessing, such as word segmentation and vectorization, standardized data is formed.

Second, based on the preprocessed data, the BERTopic model is used to extract initial topics, and cross-data source topic fusion is completed according to semantic similarity to construct a global candidate technical topic and topic co-occurrence network, providing support for indicator calculation.

Finally, the calculation of candidate ET topics is completed according to the ET identification indicator system, and the screening and verification of ETs are carried out accordingly.

B. Acquisition and Identification of Technical Topics

1) Acquisition and Fusion of BERTopic-based Topics

As a topic modeling method integrating embedding models and clustering algorithms, BERTopic can realize deep semantic mining, efficient dimensionality reduction, clustering, and visual analysis, making it suitable for identifying potential topics in complex corpora. We conducted technical topic modeling with BERTopic as the core. First, we converted multi-source texts into semantic vectors through a pre-trained model, then automatically identified semantically similar documents through Uniform Manifold Approximation and Projection (UMAP) dimensionality reduction and Hierarchical Density-Based Spatial Clustering of Applications with Noise (HDBSCAN) density clustering to avoid deviations caused by manually setting the number of topics. Finally, we generated topic representations by combining the weights of topic words, laying a foundation for topic fusion and indicator calculation.

On this basis, we used cosine similarity for topic fusion. Since the topic boundaries of papers and patents are clear, while the topics of data, such as funding data, reports, industrial market data and policy documents are cross-domain and ambiguous, making direct alignment difficult, we designed a two-stage fusion strategy to unify the semantic space and improve topic directions. In the first stage, we calculated the cosine similarity between paper and patent topics, and pair and merge them according to the principle of second-highest similarity to eliminate redundancy within sub-directions. In the second stage, we fused the obtained topics with public topics again, and finally formed a global candidate technical topic that covers multi-source data, has a unified semantics, and complete directions.

2) Identification of Target Technical Topics

We used the TF-IDF algorithm to extract keywords from topic texts and construct a co-occurrence network of technical topics. Then, we performed calculation and standardization processing in accordance with the constructed indicator system. To reduce the impact of subjectivity, we adopted the CRITIC objective weighting method to determine the objective weight of each indicator, and screen out representative emerging technology topics by setting thresholds.

C. Emerging Technology Indicator System

By sorting out existing research on ET identification, we constructed a multi-dimensional identification indicator system from five dimensions—novelty, growth, coherence, potential impact, and uncertainty—to systematically depict the characteristics of ETs.

1) Novelty

To comprehensively depict the novelty of topics, we constructed topic novelty indicators from both temporal and semantic perspectives. The former reflects the recency of a topic's emergence in the temporal dimension, while the latter measures the difference of a topic in the semantic space.

We use the average publication time of all documents in a topic as the temporal novelty indicator [20], and the formula is as follows:

$$N_i^{(t)}=\frac{1}{N_i}\sum_{j=1}^{N_i} T_{ij} \tag{1}$$

Here, $N_i^{(t)}$ denotes the temporal novelty of the i-th topic; T_{ij} represents the publication year of the j-th document within the i-th topic; and N_i indicates the total number of documents contained in the i-th topic.

Semantic novelty is operationalized as the inverse of the average inter-topic semantic similarity calculated based on TF–IDF vectors, thereby measuring the distinctiveness of a topic within the semantic space. Specifically, the procedure is as follows: first, the keywords of each topic are transformed into semantic vectors using the TF–IDF representation. Second, cosine similarity between topics is computed to characterize the degree of proximity in their semantic content. Finally, the inverse of a topic's average similarity to all other topics is taken as its semantic novelty indicator. The corresponding formulation is as follows:

$$\overline{Sim}(t_i)=\frac{1}{n\text{-}1}\sum_{j=1,j\neq i}^{n}\frac{v_i\cdot v_j}{\|v_i\|\|v_j\|} \tag{2}$$

$$N_i^{(s)}=1\text{-}\overline{Sim}(t_i) \tag{3}$$

Here, $N_i^{(s)}$ denotes the semantic novelty of the i-th topic. The keyword set of the i-th topic is represented as a vector v_i, and n denotes the total number of topics. The semantic similarity between any two topics t_i and t_j is defined using cosine similarity. $\overline{Sim}(t_i)$ represents the average semantic similarity between topic t_i and all other topics.

2) Growth

Topic growth is measured by topic intensity, defined as the number of supporting documents under a given topic. This value can be directly obtained from the document–topic mapping generated by the BERTopic model. Specifically, we calculated the proportion of documents associated with a given topic relative to the total number of documents across all topics in each year and then computed the average of these proportions to reflect the sustained growth level of the topic within the overall research landscape [21]. The formulation is as follows:

$$G_i=\frac{1}{Y}\sum_{y=1}^{Y}\frac{C_{ij}}{C_{total,\,y}} \tag{4}$$

In (4), G_i denotes the growth indicator of topic i; C_{ij} represents the number of documents under topic i in year y; and $C_{total,\,y}$ denotes the total number of documents across all topics in year y. In this study, the time window is fixed at $Y=5$. The average growth rate is calculated over the most recent five years.

$$N_i^{(s)}=1\text{-}\overline{Sim}(t_i) \tag{5}$$

3) Coherence

Topic coherence is used to characterize the semantic stability and conceptual consistency of a topic over time, namely, the degree of similarity in its keywords and semantic representations across consecutive years. This study constructs a dual-measurement framework at both the lexical and semantic levels. The two measures are integrated to form a comprehensive coherence indicator, reflecting the stability and consistency of a topic during its semantic evolution.

The similarity between the same topic across adjacent time slices is measured based on the documents it contains. A higher Jaccard coefficient indicates stronger coherence [16]. Specifically, topic texts for each year are vectorized, and the Jaccard coefficient between the Top-K keyword sets of two consecutive years is used to quantify lexical stability. This coefficient is compared longitudinally with the topic's own historical values to trace its evolutionary trajectory, rather than being compared horizontally with other topics in the same period. The formulation is:

$$J_{t,\,t+1}^{(\mathrm{k})}=\frac{|\mathrm{K_t}\cap \mathrm{K_{t+1}}|}{|\mathrm{K_t}\cup \mathrm{K_{t+1}}|} \tag{6}$$

Here, $\mathrm{K_t}$ denotes the Top-K keyword set extracted via TF–IDF from the topic-related texts in year t; $J_{t,\,t+1}^{(\mathrm{k})}$ denotes the Jaccard similarity between the two consecutive annual keyword sets.

An autocorrelation coefficient is employed to measure the similarity of a topic to itself at different time points. In the context of topic coherence, it captures semantic self-similarity over time. We utilized SBERT sentence embeddings to compute the cosine similarity between the average textual vectors of adjacent years:

$$R_{t,\,t+1}^{(s)}=\cos(\overrightarrow{\mathrm{v_t}},\overrightarrow{\mathrm{v_{t+1}}})=\frac{\overrightarrow{\mathrm{v_t}}\cdot\overrightarrow{\mathrm{v_{t+1}}}}{\|\overrightarrow{\mathrm{v_t}}\|\|\overrightarrow{\mathrm{v_{t+1}}}\|} \tag{7}$$

In (7), v_t denotes the mean embedding vector of all topic-related texts in year t encoded by SBERT, v_{t+1} represents the corresponding vector in year t+1, and $R_{t,\,t+1}^{(s)}$ is the semantic cosine similarity.

4) Potential Impact

In the context of multi-source data integration, potential impact reflects the cross-domain diffusion degree of a topic across heterogeneous data sources. The more often a topic appears in data sources, the stronger its cross-domain influence. Accordingly, we operationalized potential impact as the number of distinct data source categories covered by a given topic.

5) Uncertainty

Uncertainty is assessed through changes in structural entropy. A higher structural entropy indicates greater complexity and uncertainty within the technological topic network. Extracted topic keywords are treated as nodes, and a network is constructed based on associations among technological topics. Structural entropy theory is then applied to analyze the network structure [22]. The formulations are:

$$p(x_i)=\frac{d_i}{2E} \tag{8}$$

$$H(X)=-\sum_{\mathrm{i=1}}^{\mathrm{n}}p(x_i)\log_2 p(x_i) \tag{9}$$

Here, $H(X)$ denotes structural entropy; $p(x_i)$ represents the probability associated with node i, calculated as d_i (the degree of node i, i.e., the number of edges connected to it) divided by twice the total number of edges in the network ($2E$).) The term 2E is used because each edge is counted twice when computing node degrees (once for each endpoint).

IV. Empirical Analysis

This section carries out an empirical study in the smart grid domain to validate the proposed method, including data collection, topic modeling, indicator calculation, result screening, and comparison with authoritative forecasts.

A. Data Acquisition and Preprocessing

We selected the smart grid sector as the research object for ET identification. On the one hand, this field features a high degree of technological intersection, and ETs often emerge rapidly in the processes of engineering application, system coupling, and cross-scenario deployment. On the other hand, the smart grid undertakes multiple national strategic tasks, and its technological development path is highly concerned by policies, industries, and scientific research institutions, ETs play a critical role in the overall performance of the smart grid system.

We systematically constructed search strategies to obtain data on papers, patents, funding data, industry reports, industrial market data and policy documents in eight sub-directions of the smart grid field. The data preprocessing link mainly includes deduplication, word segmentation, and stop-word removal, etc. The sources and descriptions of data acquisition are shown in Table 1.

TABLE I. Sources and Descriptions of Data Acquisition

Data Type	Source	Preprocessed Data Volume
Publications	Web of Science	44,642
Patents	Incopat Patent Database	11871
Funding data	NSF Database	375
Industry reports	Conference Reports	126
Industry market data	Kehui Network, Cigre Database	496
Policy documents	Peking University Law Database	725

B. Acquisition of Multi-source Technology Topics

We conducted topic modeling on multi-source text data based on the BERTopic model. On the one hand, we evaluated the effect of topic modeling by relying on the Intertopic Distance Map provided by BERTopic. We determined topics according to the following criteria: The clearer the topic boundaries, the closer the semantic clusters, and the less overlap, the more interpretable and structural the topics extracted by the model [23]. On the other hand, in order to adapt to the differences in text scale and semantic density among different data sources, we adopted the "auto" default parameter of BERTopic to determine the appropriate number of topics in the selection of topic quantity. The advantages of this strategy are as follows: (1) It avoids overfitting or excessive topic splitting that may be caused by artificially setting the number of topics. (2) It can identify the optimal number of clusters according to the actual density distribution of data. When the corpus of data expands or shrinks, the optimal number of topics will increase or decrease accordingly, thus ensuring the reproducibility of results and the applicability of domain migration.

C. Identification Results of Emerging Technology Topics

After completing the topic extraction of multi-source data, the 309 identified initial technology topics were cross-source fused according to the two-stage cosine similarity fusion strategy to obtain candidate technology topics. On this basis, according to the indicator system constructed in the theoretical part, the indicator values were calculated for the fused candidate topics. After determining the weights by the CRITIC objective weighting method, the comprehensive indicator values were calculated and normalized, and finally the target technology topics were screened out. The process and results of technology topic identification in the smart grid research are shown in Table 2.

TABLE II. The Process and Results of Technology Topic Identification in Smart Grid Research

Step	Method / Strategy	Number of documents and topics
Multi-source data acquisition	Data retrieval and collection	40,248 documents
Candidate technical topics	BERTopic topic modeling	309 topics
Integrated technical topics	Cosine similarity-based fusion	233 topics
Target technical topics	Weight determination via CRITIC entropy method	117 topics

After indicator calculation, we assigned weights according to the CRITIC entropy method (Weight determination via CRITIC entropy method) to calculate the comprehensive indicator values. We then selected the top 50% of the comprehensive values as ET topics, and a total of 117 target technology topics were screened out. The threshold setting is based on analyzing the characteristics of the indicator values of candidate ET topics, and considering the conversion rate of the technology life cycle [2] from "emergence" to "take-off", so 50% is taken to ensure that the subsequent incubation resources are of the same order of magnitude as the conversion rate. The standardized characteristic indicator values of some ET topics in each sub-direction are shown in Table 3.

TABLE III. CHARACTERISTIC INDICATOR VALUES OF SOME ET TOPICS IN SUB-DIRECTIONS

Subfield	Topic ID	Novelty	Growth	Coherence	Potential Impact	Uncertainty	Comprehensive Value
Power System and Planning	Topic 0	0.3787	0.9584	0.8316	0.3333	0.8699	0.5862
	Topic 7	0.2737	1.0000	0.6890	0.6666	0.5370	0.5470
	Topic 1	0.2801	0.4617	0.8511	0.0000	1.0000	0.4740
Power Supply Side	Topic 0	0.2254	1.0000	1.0000	1.0000	0.9537	0.7493
	Topic 2	0.3108	0.1649	0.4861	0.5000	1.0000	0.5065
	Topic 1	0.2524	0.3746	0.7844	0.2500	0.7567	0.4646
Power Grid Side	Topic 3	0.1030	0.1469	0.1280	0.1820	0.1077	0.6677
	Topic 1	0.1201	0.0725	0.1537	0.1213	0.1408	0.6086
	Topic 0	0.1062	0.0858	0.1240	0.0000	0.2273	0.5436
Load Side	Topic 1	0.3096	0.7109	0.9425	1.0000	1.0000	0.8019
	Topic 11	0.2427	0.9997	0.8697	1.0000	0.5740	0.7328
	Topic 3	0.2900	0.4081	0.7584	1.0000	0.9204	0.6977
......							

Next, we selected the names of some ET topics, the number of included documents, and the topic words from Table 3 to display them in Table 4.

TABLE IV. NAMES, NUMBER OF INCLUDED DOCUMENT ITEMS AND TOPIC WORDS OF SOME EMERGING TECHNOLOGY TOPICS

Subfield	Topic ID	Topic name	Document count	Representative keywords
Power System and Planning	Topic 1	wind_power_energy_model	717	wind, power, energy, model, generation, storage, speed, uncertainty, systems, farms
Power Supply Side	Topic 2	heat_efficiency_temperature_thermal	340	heat, efficiency, temperature, thermal, fuel, cell, solar, performance, cycle, exergy
Power Grid Side	Topic 0	energy_microgrid_model_optimal	413	energy, microgrid, model, optimal, storage, microgrids, optimization, proposed, operation, cost
Power materialsSide	Topic 11	gas_natural_co_electricity	21	gas, natural, co-power , electricity, model, power, integrated, fired, systems, technologies
......				

D. Verification

After obtaining the ET topics in the smart grid field through the above process, in order to verify the rationality and reliability of the technology identification indicator system constructed in this paper, the technology topics identified in this paper are compared with relevant technology forecast reports. As shown in Table 5.

TABLE V. COMPARISON OF ETS IDENTIFIED IN THIS PAPER AND THOSE MENTIONED IN RELEVANT TECHNOLOGY FORECAST REPORTS

Emerging Technology Content	Supporting Reports and Policies
Core Processes of Solid-State Batteries, Sodium-Ion Battery Technology, New Battery Chemistry Technology	IEEE CS Authoritative Forecast: One of the breakthrough technologies in 2025: New battery chemistry technologies, including solid-state and sodium-ion batteries, need to break through mass production processes and supply chain management to further improve their energy density and safety [24].
Structural Battery Composite Materials, High-Performance Electrode Material Technology, "Solid-Solid Interface Regulation Preparation Process"	The "Frontier Situation Analysis of Key Scientific and Technological Fields 2025" jointly researched by the Institute of Scientific and Technical Information of China and the Shanghai Institute of Science of Science focuses on cutting-edge directions, such as the innovation of solid-state electrolyte material systems, the research and development of high-performance positive/negative electrode materials, and the breakthrough of solid-solid interface regulation preparation processes, conducting annual dynamic tracking [25].
Structural Battery Composite Materials, Osmotic Energy Power Generation System Technology, Green Carbon Sequestration Technology	The World Economic Forum released the Top 10 ETs of 2025 at the 16th Summer Davos Forum, where green carbon sequestration, structural battery composite materials and osmotic energy power generation system technology were selected into the annual "list" [26].

E. Discussion

Taking the smart grid as the research object, we integrated six types of multi-source heterogeneous data, completed technology topic extraction and cross-source fusion based on the BERTopic model, and combined the five-dimensional indicator system and CRITIC weighting method to finally identify 117 ET topics, covering eight sub-directions. The core technologies are highly consistent with the 2025

 ISBN: 978-1-68558-397-2

technology forecasts of authoritative institutions, such as IEEE CS and the World Economic Forum, verifying the applicability of the method.

From the perspective of identification results, the development of ETs in the smart grid field highlights the trends of low carbonization, energy storage, and interdisciplinarity. Cross-field technologies, such as solid-state batteries, green carbon sequestration, and structural battery composite materials have become important innovation tracks. Compared with traditional identification methods based on single data sources, the proposed framework integrates scientific research, industrial application, and policy orientation, enabling a more comprehensive depiction of ET evolution characteristics and improving the timeliness and completeness of technology identification.

The research results can provide accurate guidance for policy formulation and industrial layout in the smart grid field. For high-potential technology directions, policy support, and R&D investment can be increased, while paying attention to the trend of technological cross-integration to promote cross-field innovation. The identification method based on multi-source data fusion and multi-dimensional indicator coordination proposed in this research also provides a transferable research framework for ET identification in other fields.

V. CONCLUSION AND FUTURE WORK

This study proposes an ET identification method integrating multi-source data fusion and coordinated multi-dimensional indicators. By combining BERTopic-based topic extraction, semantic similarity-based topic fusion, and a five-dimensional indicator system, the method enables systematic identification of ETs from heterogeneous data sources. The empirical analysis in the smart grid field demonstrates the applicability of the proposed framework.

The main contributions of this study are as follows: multi-source data integration compensates for the limitations of single-source identification; scientific research, industrial application, and policy orientation are incorporated into a unified analytical framework; and a multi-dimensional indicator system is constructed to quantitatively characterize ETs.

This research has certain limitations: It lacks dynamic evolution analysis of technology topics, conducts topic extraction and identification based on static cross-sectional data, and does not carry out tracking analysis from a time series dimension. The uncertainty indicator does not consider external factors, such as market and policies. The data sources focus on specific providers. In the future, data can be further enriched, time series analysis methods can be introduced, and cross-field empirical research can be carried out to further improve the versatility and adaptability of the method.

Funding: This research was funded by the National Natural Science Foundation of China (No. 72274113), Shandong Provincial Social Science Foundation (No. 23CTQJ07), Shandong Provincial Natural Science Foundation (No. ZR2022MG052), Beijing Natural Science Foundation (No. 9242006) and the Taishan Scholar Foundation of Shandong province of China (tsqn202103069).

Use of Generative Artificial Intelligence for Writing: We used ChatGPT for translation, proofreading, and grammar checking. We evaluated the output by cross-referencing the translated and revised content with the original text to ensure accuracy, consistency, and alignment with the intended meaning. Additionally, we reviewed the final version to confirm that all technical terms and concepts were appropriately conveyed.

REFERENCES

[1] Y. Zhou, F. Dong, Y. Liu, and L. Ran, “A deep learning framework to early identify emerging technologies in large-scale outlier patents: an empirical study of CNC machine tool,” *Scientometrics*, vol. 126, no. 2, pp. 969–994, Feb. 2021.

[2] D. Rotolo, D. Hicks, and B. R. Martin, “What is an emerging technology?,” Research Policy, vol. 44, no. 10, pp. 1827–1843, Dec. 2015.

[3] "The Nature of Technology: What It Is and How It Evolves, W.B. Arthur. Free Press, New York (2009), 246 pp.," ResearchGate, Aug. 2025.

[4] X. Wu, Y. Shao, and F. Lin, "The Path Mechanism for Breaking Through the "Valley of Death" of Key Core Technology Innovation," Sci. Res., vol., no., Dec. 2024, Advance online publication.

[5] Z. Wang, Y. Chen, Z. Jiang, and Z. Liu, "An input output analytical method of identiHcation of the advanced technology and core technology; The case study of hybrid electric vehicle technology," Sci. Res., vol. 33, no. 11, pp. 1612–1620, 2015.

[6] X. Wang, J. Dong, T. Yu, Y. Chen, and T. Chen, "Research on Technology Fronts Forecasting Method Based on Informetrics Analysis Using Multi-Type Information," J. Inf., vol. 37, no. 10, pp. 70–75, 89, 2018.

[7] G. Li, X. Wu, and B. Ning, “Research Frontier Detection Method Based on Topic Time Series Diffusion Network between Multi-source Data,” Data Anal. Knowl. Discov., Apr. 2025, Advance online publication, Accessed: Nov. 15, 2025.

[8] X. Zhang, Z. Zhang, L. Cao, W. Ruan, X. Ren, and Z. Feng, "Research Progress of Research Front Recognition Methods in Subject Fields," Lib. Inf. Serv., vol. 66, no. 12, pp. 139–151, 2022.

[9] J. Shi, X. Zhao, and Q. Liu, "National Security-Oriented Intelligence Support Path," J. China Soc. Sci. Inf., vol. 39, no. 7, pp. 675–686, 2020.

[10] G. S. Day and P. J. H. Schoemaker, Wharton on Managing Emerging Technologies. John Wiley & Sons, 2004.

[11] G. S. Day and P. J. H. Schoemaker, “Avoiding the Pitfalls of Emerging Technologies,” California Management Review, vol. 42, no. 2, pp. 8–33, Jan. 2000.

[12] A. Breitzman and P. Thomas, “The Emerging Clusters Model: A tool for identifying emerging technologies across multiple patent systems,” Research Policy, vol. 44, no. 1, pp. 195–205, Feb. 2015.

[13] L. Wang, S. Fang, and P. Ji, "Using Patent Documents to Study the Technology Framework of Detecting Emerging Technology Topics," Lib. Inf. Serv., vol. 55, no. 18, pp. 74–78, 23, 2011.

[14] G. Li, F. Xie, X. Tang, B. Wang, and Y. Xu, "Research on the Identification of Emerging Technologies Integrating Multidimensional Perspectives," J. China Soc. Sci. Inf., vol. 44, no. 8, pp. 962–976, 2025.

[15] T. Zhang, J. Chen, X. Zhou, S. Sun, and Y. Zhang, "Emerging Technology Identification Method Based on Potential Impact Prediction and Multi-Source Information Fusion," J. Inf., vol. 44, no. 9, pp. 134–142, 2025.

[16] H. Xu, J. Winnink, Z. Yue, H. Zhang, and H. Pang, “Multidimensional Scientometric indicators for the detection of emerging research topics,” Technological Forecasting and Social Change, vol. 163, p. 120490, Feb. 2021.

[17] Y. Zhou, H. Lin, Y. Liu, and W. Ding, “A novel method to identify emerging technologies using a semi-supervised topic clustering model: a case of 3D printing industry,” Scientometrics, vol. 120, no. 1, pp. 167–185, Jul. 2019.

[18] M. Jiang, S. Yang, and J. Wei, "Emerging-Technology Identification Based on Technology Structure-Function Semantic Association,"J. China Soc. Sci. Tech. Inf., vol. 44, no. 5, pp. 522–534, 2025.

[19] S. Yang and M. Jiang, "Research Progress on Connotation Characteristics and Identification Methods of Emerging Technologies," Inf. Sci., vol. 41, no. 5, pp. 181–190, 2023.

[20] X. Zhu, J. Zhang, W. Li, X. Liu, and G. Geng, "Detecting Emerging Topics of Artificial Intelligence Based on Perspective of Top International Conferences," Inf. Theory Pract., vol. 47, no. 9, pp. 147–155, 2024.

[21] P. Liu, "Identifying emerging technology topic based on multi-source data and co-occurrence network," M.S. thesis, Univ. Chinese Acad. Sci. (Nat. Lib. China), 2024.

[22] D. Wang, X. Zhou, P. Zhao, J. Pang, and Q. Ren, “Early identification of breakthrough technologies: Insights from science-driven innovations,” Journal of Informetrics, vol. 19, no. 1, p. 101606, Feb. 2025.

[23] M. Grootendorst, “BERTopic: Neural topic modeling with a class-based TF-IDF procedure,” Mar. 11, 2022.

[24] ″IEEE CS authoritative forecast: 22 breakthrough technologies lead the future in 2025,″ Accessed: Feb. 16, 2026. [Online]. Available: http://www.nscc-gz.cn/newsdetail.html?8213

[25] ″Analysis of Frontier Trends in Key Science and Technology Fields 2025 released at the 18th Pujiang Innovation Forum,″ Accessed: Feb. 16, 2026. [Online]. Available: https://www.siss.sh.cn/c/2025-09-24/665730.shtml

[26] ″World Economic Forum releases top 10 emerging technologies of 2025," Accessed: Feb. 16, 2026. [Online]. Available: https://www.ncsti.gov.cn/kjdt/kjrd/qtrd_kjrd/202506/t20250625_208621.html

 ISBN: 978-1-68558-397-2