

Identification of core technological topics in the new energy vehicle industry

The SAO-BERTopic topic modeling approach
based on patent text mining

Author: Jianxin Zhu, Yutong Chuang, Zhinan Wang, Yunke Li
Presenter: Yutong Chuang



CONTENT

Part 1

Research Background

Part 2

Research Design

Part 3

Research Result

Part 4

Conclusions



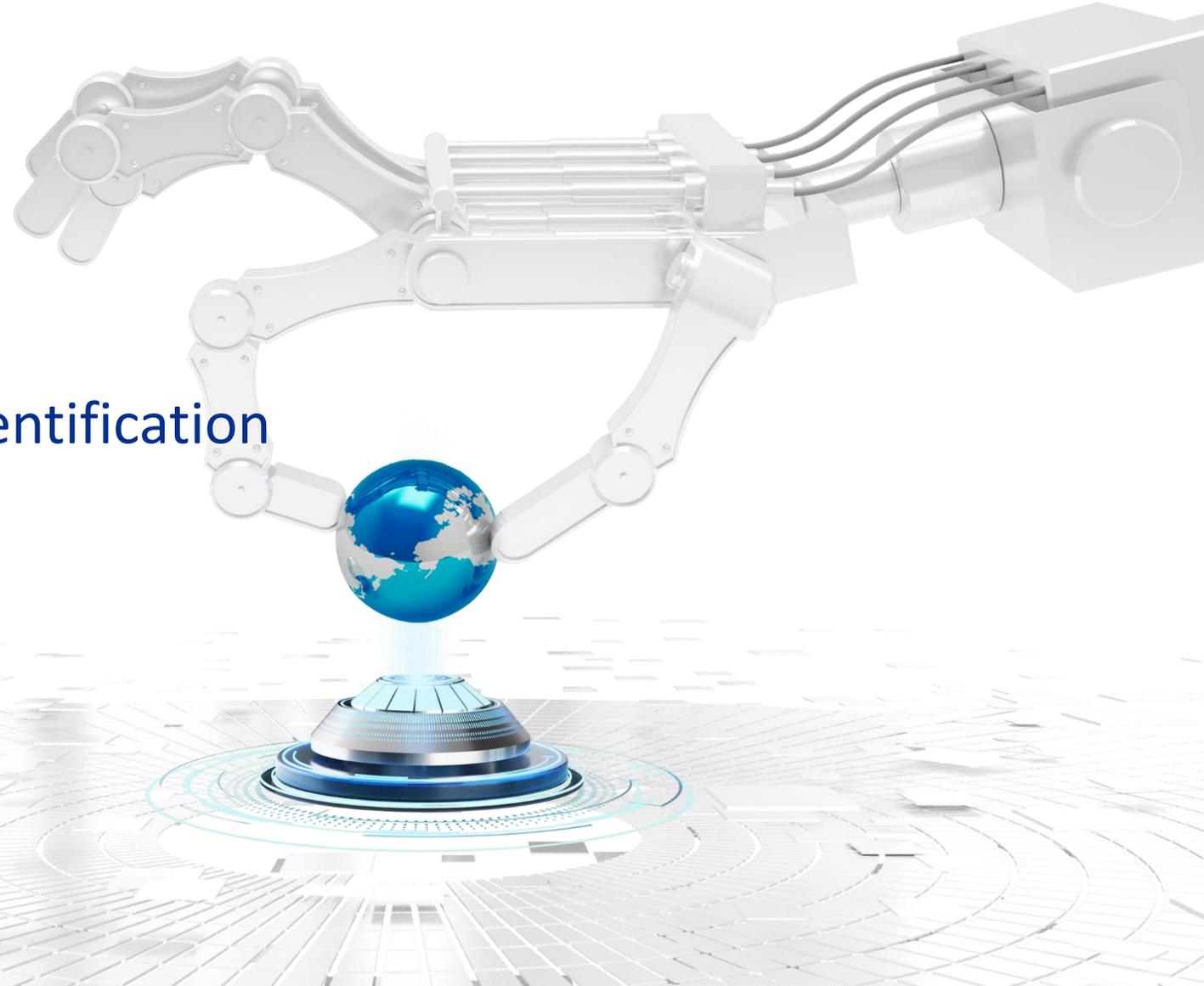
PART 1

Research

Background

WHY we do this research?

- Industry Significance
 - Technological Complexity
 - Challenges in Technology Identification
-



WHY new model?

Topic model	Disadvantage
LDA	Identifying topics in a document primarily based on word frequency can overlook semantic complexity and conceptual nuance in the document, resulting in missing some important but uncommon technical terms.
LSA	Based on linear algebra, it may be difficult to deal with nonlinear and complex semantic relations.
CTM	The high computational complexity limits its rapid application on large document collections. It mainly focuses on the statistical information of words in the document, and fails to fully understand the structural information and deep semantic information in the document.

Purpose

This study aims to precisely identify and analyze core technologies in the new energy vehicle industry by employing an innovative SAO-BERTopic model to extract structured information from extensive patent data.



PART 2
Research
Design

Research Process

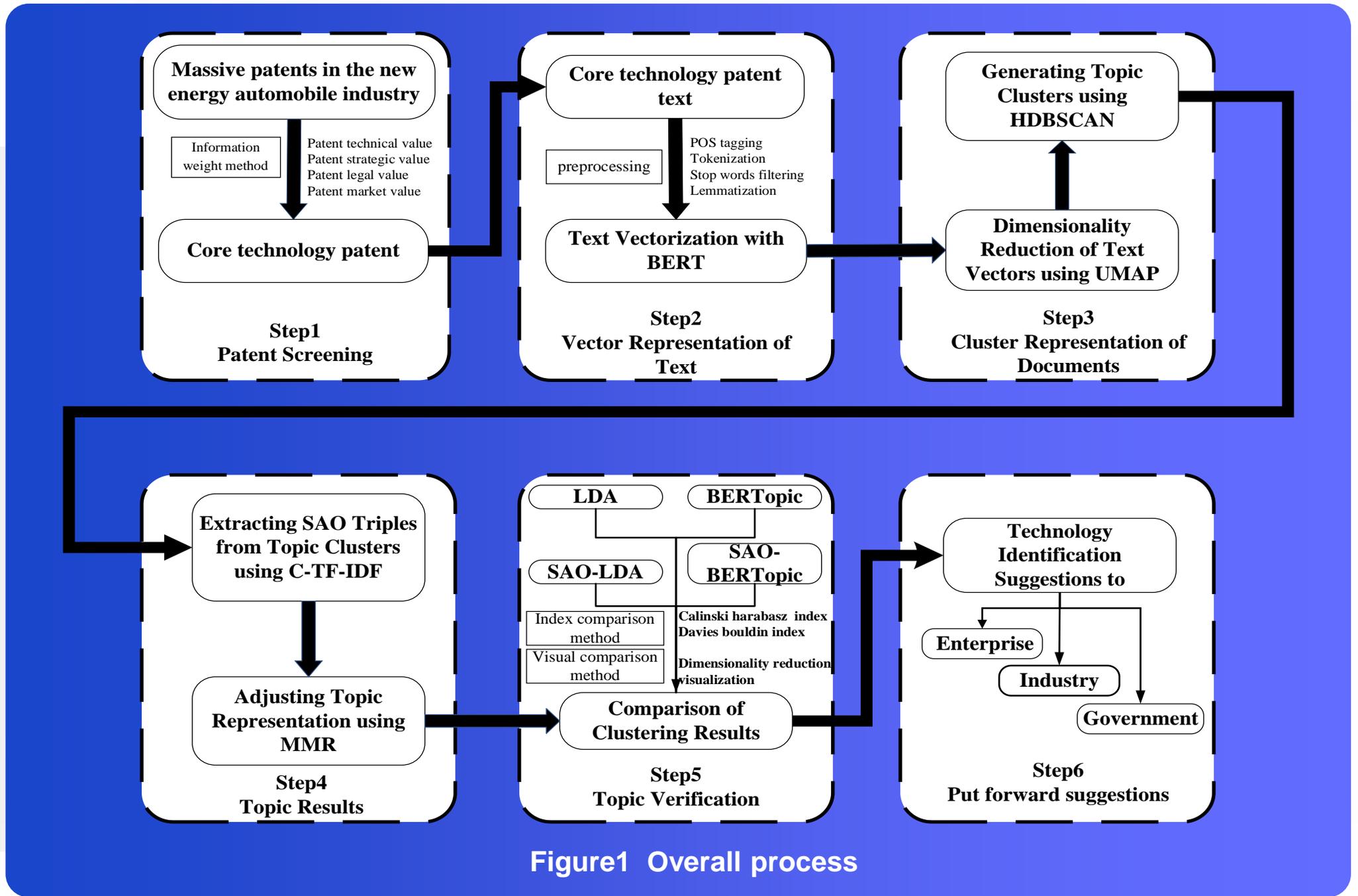


Figure1 Overall process

Information Weight Method

We employed the information weight method to meticulously screen and select high-quality core patents, ensuring our analysis focuses on the most impactful and innovative technologies in the new energy vehicle sector.

Identification Dimensions	Indicators
Patent Technological Value	Citation Frequency PCT Application
Patent Strategic Value	IPC Classification Number of Inventors
Patent Legal Value	Legal Validity Claims Count
Patent Market Value	Number of Family Patents Number of Licenses and Transfers

Table 2 Screening dimension and index

SAO-BERTopic model

We construct the SAO-BERTopic model, a new approach that combines subject action object (SAO) analysis with the BERTopic topic model.

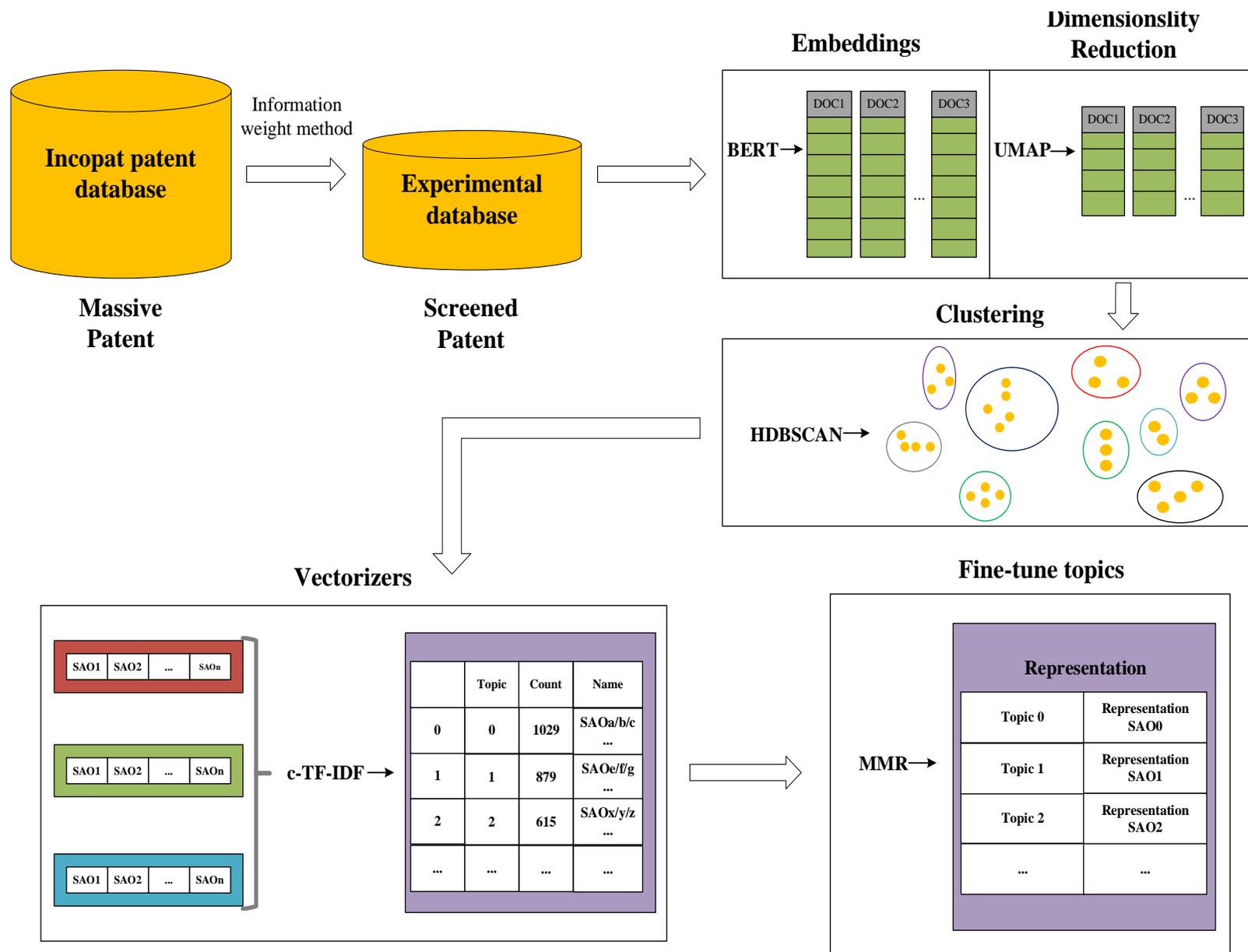
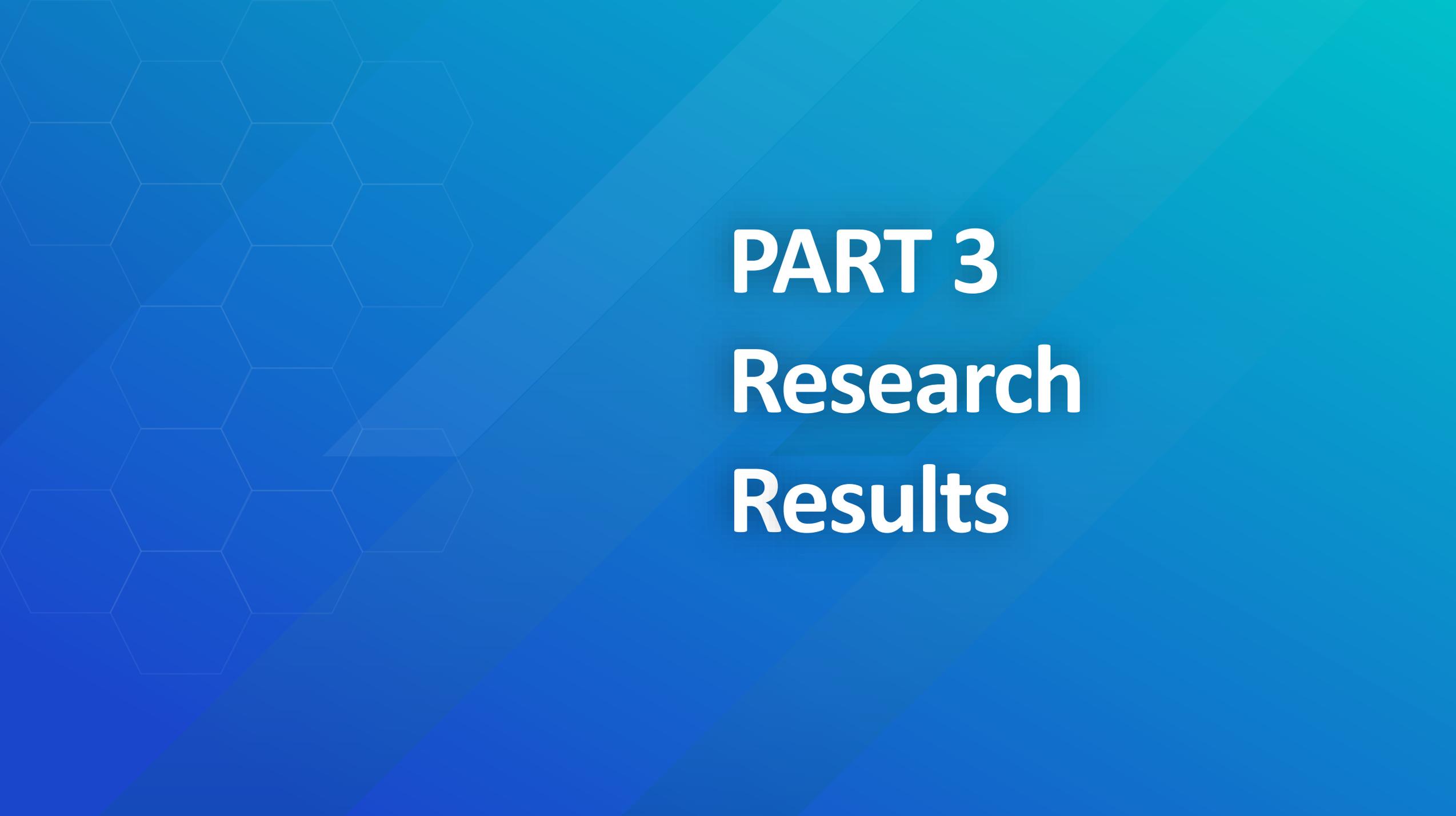


Figure 3 Schematic of SAO-BERTopic model principle

Comparative analysis between SAO-BERTopic and other conventional topic models

Model	Topic Representation	Reflects Relationships Between Technological Components	Topic Number Setting	Semantic Expression
LDA	Words or Phrases	No	Manual	Ignored
BERTopic	Words or Phrases	No	Automatic	Reflected
SAO-LDA	Triplets with SAO Structure	Yes	Manual	Reflected
SAO-BERTopic	Triplets with SAO Structure	Yes	Automatic	Reflected

Table 3 Comparison of the results of four topic models



PART 3
Research
Results

Research results

We use hierarchical topic modeling (as shown in Figure 4) to intervene on a plethora of topics, and at the same time, combining the knowledge and guidance of domain experts, the topics are finally refined into five core topics.

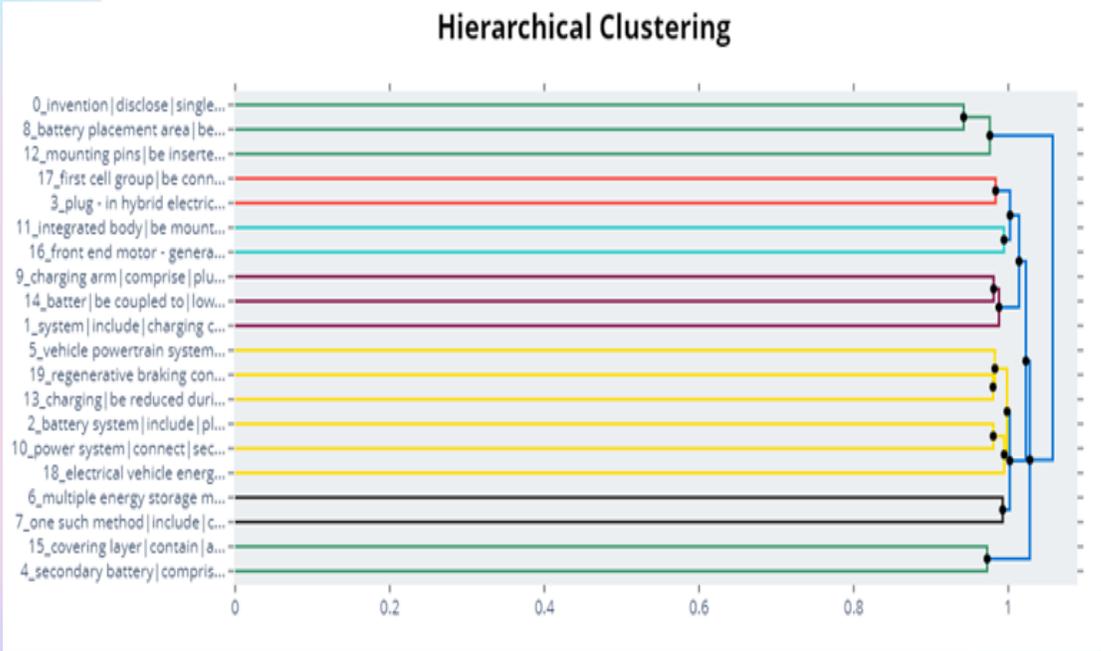


Figure 4 Hierarchical modeling diagram

Text	Result
novelty - a battery case comprises a ceramic coating layer, which is formed on the surface of the battery case for the electric vehicle.; use - battery case for an electric vehicle that is utilized in electric motors, hybrid electric vehicles driven by engines and driving motors, and fuel cells.; advantage - the battery case has improved heat dissipation performance and durability.	[('battery case', 'comprise', 'ceramic coating layer'), ('battery case', 'improve', 'heat dissipation performance'), ('battery case', 'improve', 'durability')]

Table 4 SAO-BERTopic clustering result

Identification result

	Name	Detail
Topic 0	Energy Management and Power Transmission Technology in Hybrid Electric Vehicles	encompasses the integration of internal combustion engines and electric motors, energy storage and recovery, as well as the control of power transmission systems.
Topic 1	Electric Vehicle Charging Systems and Energy Management	where the SAO triples contained in topic 1 focus on describing the functionalities of receiving, managing, and providing energy in charging panels and systems.
Topic 2	Battery System Integration and Energy Efficiency Management	where the SAO triples collectively depict various aspects of energy storage system design, management, and application, including modular composition, energy control and management, system power supply, and integration with electric vehicles.
Topic 3	Electric Vehicle Battery Pack Configuration and Structural Design	covering various aspects of the physical configuration, structural design of battery packs, and how these designs impact battery pack performance.
Topic 4	Cathode Materials and Composition for Secondary Batteries	focusing on the composition and design of cathode materials for batteries, especially secondary batteries, which directly impact battery performance aspects like capacity, energy density, charge/discharge speed, and lifespan.

Table 5 Identification result

SAO-BERTopic brings?

Topic A: Battery-Store-Energy

Topic B: Battery-Provide-Power

Existing topic models

- Individual words like “battery,” “store,” “energy,” “provide,” and “power” are common across multiple topics, creating ambiguity in topic distinction.

SAO-BERTopic

- SAO triples offer more specific information, enhancing topic clarity.
- Topic A clarifies that the “battery” is to “store energy”.
- Topic B indicates that the “battery” is used to “provide power”.
- Allows for precise differentiation between topics with similar content but different focus areas.

Comparison of clustering effect

	topic model	Calinski harabasz index	Davies bouldin index
k=10	bertopic+sao	154.8341188	4.135587902
	bertopic	110.2708578	7.249651957
	lda+sao	1.490386821	40.428595
	lda	113.2137258	5.730619019
k=5	bertopic+sao	191.1178148	4.476240803
	bertopic	178.4198845	5.83858005
	lda+sao	1.423594994	58.0559951
	lda	181.8154094	5.008411221

Table 6 Comparison of the results of four topic models

Comparative Visualization of Dimensionality Reduction for Four Models

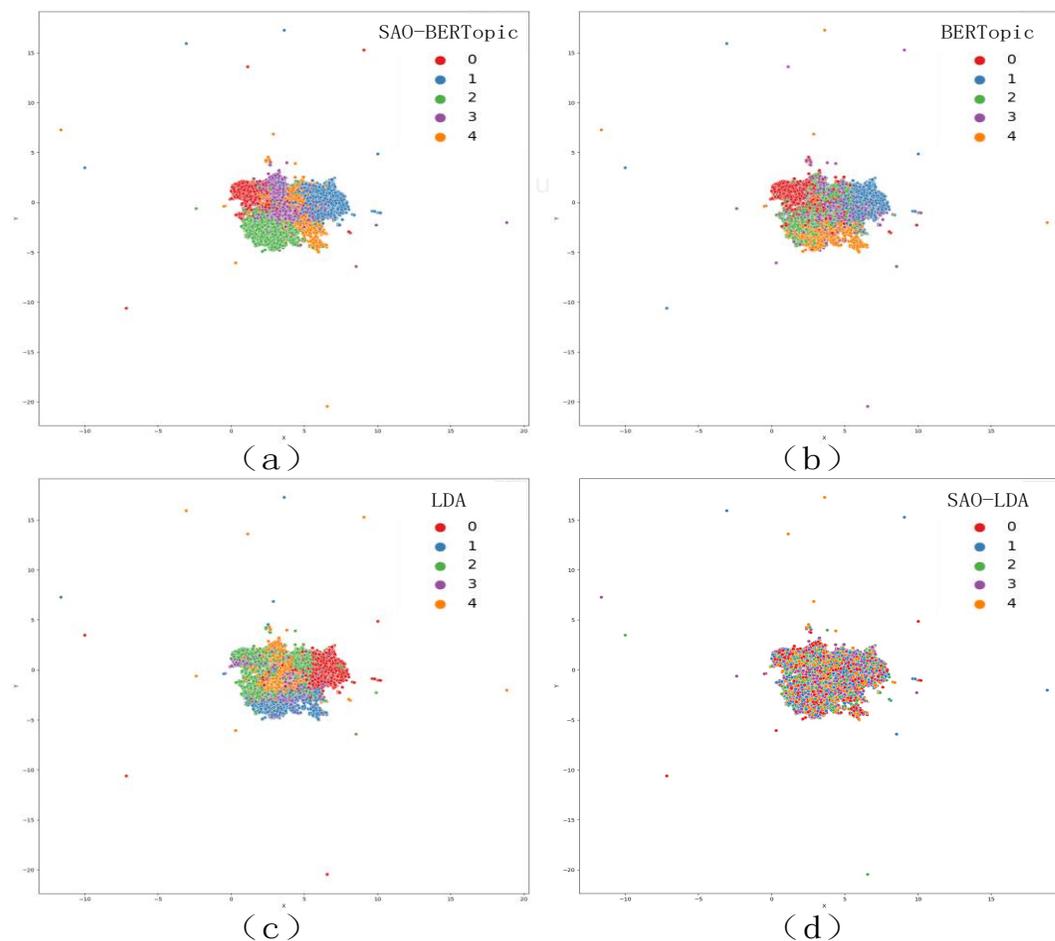


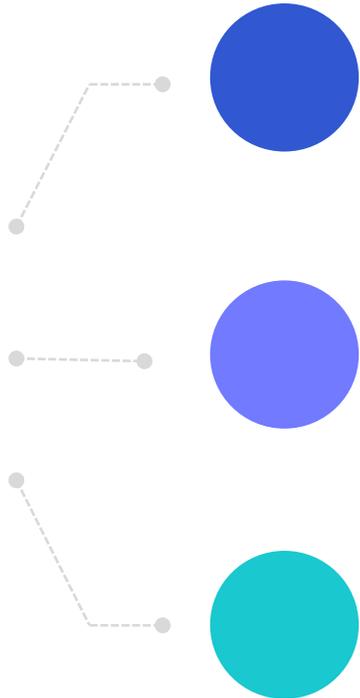
Figure 5 Visual comparison of clustering of four topic models

The background features a blue-to-teal gradient with a pattern of faint, light-colored hexagons on the left side. The text is centered on the right side of the image.

PART 4

Conclusions

Conclusions



Precision in Technology Analysis:

Demonstrated that the SAO-BERTopic model significantly enhances the precision of identifying core technologies within the vast patent data of the new energy vehicle industry.

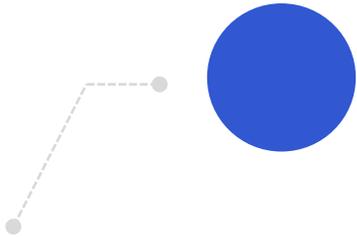
Innovative Combination:

The integration of SAO structures with the BERTopic model allows for the extraction of nuanced technological insights that are not possible with traditional models.

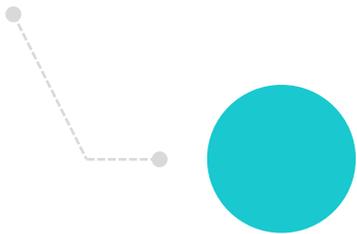
Impact on Industry Strategy:

The findings facilitate better technological forecasting and strategic alignment for companies within the new energy vehicle sector, directly influencing competitive tactics and innovation trajectories.

Innovations



The method can extract structured SAO triplet information from patent information, and realize deeper semantic understanding and structured data representation than traditional keyword extraction. This not only improves the accuracy and adaptability of the model, but also is particularly suitable for mining deep technology trends and key innovation points in patent texts containing rich technical descriptions.



Breaks through the limitations of traditional technology identification methods in processing big data, quantifies the information abundance of patent data through the information weight method, and selects high-quality patent data as the input of SAO-BERTopic model. This strategy not only enhances the ability of the model in revealing core technologies and market trends, but also significantly improves the efficiency and accuracy of technology identification.

The background is a gradient of blue shades. On the left side, there is a faint, light blue hexagonal grid pattern. Scattered across the background are several 3D geometric shapes: a teal pyramid at the top center, a purple cube at the top right, a purple cube at the middle left, a teal cylinder at the bottom left, a teal cube at the bottom center, and a teal pyramid at the bottom right. The text 'THANK YOU' is centered in a bold, white, sans-serif font.

THANK YOU