World Electric Vehicle Journal

MDPI

Article

Research on the Evolution of Online User Reviews of New Energy Vehicles in China Based on LDA

Su He [1,*], Bo Xue [2] and Dejiang Luo [3]

[1] College of Information Engineering, Sichuan Post and Telecommunication College, Chengdu 610067, China
[2] China Mobile Communications Group Sichuan Yibin Branch, Yibin 644600, China
[3] Geomathematics Key Laboratory of Sichuan Province, Chengdu University of Technology, Chengdu 610059, China
* Correspondence: hesu@sptc.edu.cn

Abstract

To achieve China's carbon peak and carbon neutrality goals, it is essential to increase the market penetration of New Energy Vehicles (NEVs) and understand consumer attitudes. Based on a big data set of over 20,000 online user reviews, this study employs the Latent Dirichlet Allocation (LDA) model to extract themes, popular brands, and focal points across different time windows. The research constructs a data-driven threshold filtering mechanism that integrates topic probability, frequency, keyword weight, and cross-temporal topic similarity to quantify consumer reviews, enabling an in-depth analysis of the dynamic evolution of attitudes in the NEV market. The findings reveal a dual shift in consumer sentiment: first, a transition in focus from basic configurations and aesthetics toward quality experience; and second, a shift in purchasing decisions toward a socially driven model dominated by word-of-mouth and family collaboration.

Keywords: new energy vehicle; user reviews; LDA model; big data; evolutionary path

1. Introduction

The Intergovernmental Panel on Climate Change (IPCC) noted that carbon emissions are the main cause of temperature increases, and it has become a common goal of the international community to control the temperature increase within 1.5 °C and achieve carbon neutrality by 2050. Compared with other countries, China, as a major carbon emitter, faces greater pressure to reduce carbon emissions. In 2019, China's CO_2 emissions per unit of GDP decreased by 48.1% compared with those in 2005, and the proportion of nonfossil energy sources was 15.3%; however, there is still a certain gap to be addressed for carbon emissions to peak by 2030 and carbon neutrality to be achieved by 2060, and nonfossil energy accounted for 25% of the total in the NDC plan submitted by China in 2015. The carbon emissions of new energy vehicles, especially the most representative pure electric vehicles, are significantly lower than those of traditional fuel vehicles during use. Therefore, the development and promotion of new energy vehicles are conducive to the realization of China's carbon peak and carbon neutrality goals, the implementation of a national energy development strategy and the establishment of a resource-saving and environmentally friendly society.

However, in 2019, the market ownership of new energy vehicles accounted for only 1.46% of the total number of vehicles [1] the average market penetration rate in the first three quarters of 2021 was 14.5%, and that in September was 17.3% [2]. In 2022, the

Check for updates

Academic Editor: Vladimir Katic

Received: 15 October 2025
Revised: 26 December 2025
Accepted: 28 December 2025
Published: 31 December 2025

Copyright: © 2025 by the authors. Published by MDPI on behalf of the World Electric Vehicle Association. 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.

production and sales volume of new energy vehicles were 7.058 million and 6.887 million, respectively, with a market share of 25.6% [3]. There is still a gap between the output and sales of new energy vehicles, and the market penetration rate needs to be improved. To promote the development of the new energy automobile industry and increase market ownership, the government has introduced a series of subsidy policies. The central government provided 33.435 billion yuan in subsidies from 2009 to 2015 [1]. However, due to subsidy fraud, the subsidy policy has gradually declined, and the government canceled subsidies for new energy vehicles in 2023. Therefore, the "New Energy Automobile Industry Development Plan (2021–2035)" proposes to take the market as the leading role, give full play to the decisive role of the market in resource allocation, strengthen the main position of enterprises in technology route selection and production service system construction, and better play the role of the government in green consumption guidance and other aspects. Therefore, mastering market signals and dynamically understanding consumers' intentions have become the main means to improve the production and sales of new energy vehicles.

In recent years, with the development of text mining technology, domestic and foreign scholars have carried out intelligent analyses based on big data through text and data mining to identify public opinion topics [4,5], mining user preferences [6–8], and analyzing popular topics in different research fields. Among them, the mining of user preferences mainly involves identifying user diet preferences, mobile phone purchase preferences and medical service quality recognition through online comments. To promote the production and sales of new energy vehicles, a large number of studies have been carried out at home and abroad. In domestic studies, ref. [9] conducted research on the factors and emotions that consumers attribute most to new energy vehicles through online reviews. Ref. [10] studied the national policy of new energy vehicles. In foreign studies, Norway studied the influence of energy efficiency on the choice of automobiles by consumers through the interview data of survey companies [11]. Ref. [12] generated a policy mining dictionary for new energy vehicles through more than 260 policy documents to find the shortcomings of policies through text mining; ref. [13] surveyed 1000 Malaysian private car drivers through a questionnaire to study electric vehicle usage intentions in Malaysia. The international community has a strong desire to study user preferences and government policies of new energy vehicles, while most foreign studies have not been conducted on the basis of big data, and most domestic studies have only extracted the theme of comments or policies of new energy vehicles without studying the evolution process of the theme over a time window. Therefore, it is not possible to analyze the changing path of user preferences more clearly.

This paper is mainly based on the big data of online reviews of new energy vehicle users and uses the three-layer Bayesian principle of LDA to study the theme of consumers' attention to new energy vehicles, the evolution of the theme, and the main driving force for the use of new energy vehicles. This will provide a market basis for new energy vehicle enterprises, improve the market penetration rate of new energy vehicles, and promote the technological progress of new energy vehicles, which will be conducive to the realization of China's carbon peak and carbon neutrality goals. The main contributions of this paper are as follows: (1) Using Python 3.10, more than 20,000 online evaluation data points of new energy vehicle users on Auto Home (a mainstream Chinese automobile review website) were crawled from 2016 to 2023. This study provides big data support for user preference research on new energy vehicles. (2) The LDA topic model was constructed by parameter optimization, the comment topics of different time windows were extracted, and the evolutionary paths of the inheritance, disappearance and rebirth of the comment topics were determined by calculating the similarity between the topics of different time windows. Through visualization technology, an evolutionary path was presented to analyze how the

preferences of new energy vehicle users change with different policy backgrounds and technical backgrounds.

The remainder of this paper is structured to address the deficiencies identified in previous studies, specifically the lack of temporal theme evolution analysis and insufficient big data support. Section 2 constructs the comprehensive research framework. Section 3 details the big data-collected online review dataset. Section 4 elaborates on the specific research methodology, encompassing the LDA topic model and the calculation of theme evolution paths using topic similarity. Section 5 presents and discusses the dynamic evolution results and driving forces of new energy vehicle consumption themes. Section 6 concludes the paper, and finally, Section 7 outlines the study's limitations and suggests directions for future research.

2. Research Design

The research framework of this paper is shown in Figure 1. The main research steps are as follows:

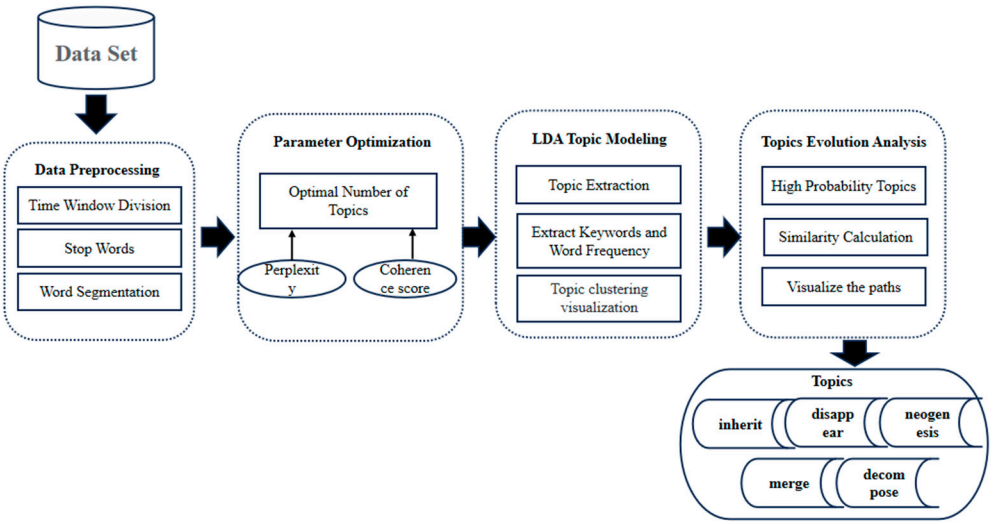


Figure 1. Research framework.

First, Python crawler technology was used to access the online review data of new energy vehicle users on the China Auto Home website, and the data were preprocessed to a certain extent. The data were mainly divided according to the time window. Based on the stop word list of the Harbin Institute of Technology (HIT) University, common but meaningless words that may appear in the review text of new energy vehicles were added through the experiment, and then the review text was divided into words. At the same time, word frequency of text vectorization and word frequency–inverse file frequency (TF–IDF) technology were used to vectorize the review text. Second, through experiments, the optimal number of topics was determined according to the degree of perplexity and coherence score to construct the LDA topic model and extract the comment topic, the keywords of the topic and the corresponding word frequency in different time windows. The most discussed brand and focus of new energy vehicle users were analyzed. Finally, the similarity between topics with high probability in different time windows was calculated to form an evolutionary path of theme inheritance, disappearance and rebirth, and the evolutionary path was presented by visualization technology to analyze how the preferences of users of new energy vehicles change with different policy and technical backgrounds.

3. Data

The evaluation data of new energy vehicles on the Auto Home website were collected by crawler technology, and the text comments on "reasons for choosing this car" in user comments were crawled, with a total amount of data reaching 25,196 pieces, covering the period from 2016 to February 2023. The collected vehicle model data included 167 types of new energy vehicles, such as Qin PLUS, Tang New Energy, Han, Song PLUS, Dolphin, Destroyer 05, ID.4, Xpeng P7, and BYD e2. The brands involved were mainly BYD, Chery, NiO, Xpeng, Changan, Volkswagen, GAC, Weima, and Tesla. In terms of time, there were 38 user evaluation data points in 2016, 42 in 2017, 281 in 2018, 739 in 2019, 1639 in 2020, 5223 in 2021, 13,946 in 2022, and 3288 in January–February 2023. Among them, the amount of review data in 2022 increased significantly, and the data volume was full. The dataset was segmented into time windows defined by years, primarily considering policy consistency and the establishment of macro-level development phases.

4. Methods

This section details the complete technical process employed to extract, model, and analyze consumer concerns from the raw user comment data. Initially, the data was prepared for subsequent modeling through text preprocessing and TF-IDF vectorization.

Subsequently, the LDA topic model was constructed to effectively identify the core topics of concern within different time windows. Finally, thematic evolution analysis was performed to quantitatively assess the relationships and transitions between topics across different years, thereby revealing the dynamic shifting trends in consumer demand.

4.1. Text Preprocessing

In text preprocessing, after removing common words through the stop list, the Chinese word segmentation tool jieba was applied for word segmentation, vectorization was performed with TF–IDF, and then the text was stored. The stop list of this study was based on the stop list of HIT University, and it included common but meaningless words in the evaluation of new energy vehicle users, such as first, car, father, woman, wife, vehicle, new energy, feeling, thinking, feel, now, choice, etc. TF–IDF is a text vectorization technique, and its formula is as follows (1)–(3):

$$TF_{ij} = \frac{n_{ij}}{|d_i|} \tag{1}$$

where TF_{ij} represents the word frequency of term j in the ith file, n_{ij} represents the number of occurrences of term j in the ith file, and $|d_i|$ is the number of all terms in the ith file.

$$IDF_j = \log \frac{|D|}{|d_j| + 1} \tag{2}$$

where IDF_j represents the inverse file frequency of term j, $|D|$ is the total number of files in the document set, and $|d_j|$ is the number of files containing term j.

$$TF - IDF = TF \times IDF \tag{3}$$

Therefore, the main principle of TF–IDF is as follows: TF represents the frequency of the term appearing in this article, and IDF represents the logarithm of the frequency of the term appearing in all articles in the entire document. If a term is rare but appears more often in a particular article, then the term is likely to reflect the characteristics of that article. Therefore, the high frequency of a term in a particular article and the low frequency of that term in the overall document collection can result in a high TF–IDF value.

4.2. LDA Topic Model

The basic idea of the LDA topic model to extract topics from online reviews of new energy vehicle users is as follows: Each comment is a probability distribution of a series of topics, denoted P(z); each topic is a probability distribution of a series of words, denoted P(w | z). Therefore, the probability distribution of each word in the review is as follows [14]:

$$P(w_i) = \sum_{j=1}^{k} P(w_i|z_i = j)P(z_i = j) \quad (4)$$

The probability graph model of LDA is shown in Figure 2, where α, β is the parameter of the Dirichlet distribution, K is the number of topics, α is a K-dimensional vector, and α is the prior probability of the subject, and α is a hyperparameter of θ, and θ is a polynomial distribution of the document-subject; β is the prior probability of the word, and is the hyperparameter of φ, and φ is the polynomial distribution of the subject-word. M represents the number of words in each review, N represents the number of comments in the corpus, and the arrows represent the probabilistic dependencies of each node. Therefore, the process of generating K topics with LDA is as follows [15]:

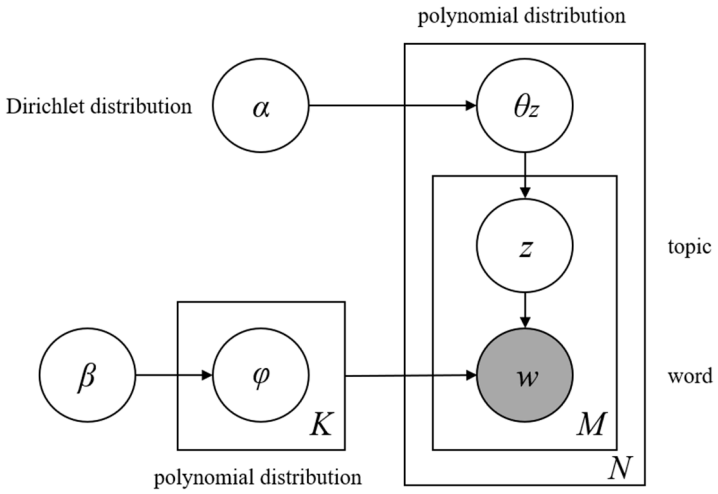


Figure 2. LDA probability graph model.

(1) Generate the distribution of the comment document on the subject, which follows α;

(2) Generate topic z randomly by θ;

(3) Generate the distribution φ of the words corresponding to the topic, which is subordinate to β;

(4) According to the theme z and φ, generate the word w;

(5) Loop (1)–(4) to generate a document containing M words;

(6) According to the set number of topics K, generate N documents with K topics.

To determine the optimal topic number K, the evaluation criterion of the LDA subject model, perplexity degree, was used, and the topic number when the perplexity degree was no longer reduced or at the inflection point was found through experiments. The lower the perplexity is, the less confused the model is about the subject representation. The formula for the perplexity degree is as follows [16]:

$$\text{Perplexity} = \exp\left[-\frac{\sum_{d=1}^{D} \log P(w_d)}{\sum_{d=1}^{D} N_d}\right] \quad (5)$$

where $P(w_d)$ is the probability of each word in document d, N_d is the number of words in the document, and D is the document set.

The Coherence Score (C_V) method assesses the quality of a topic by calculating the co-occurrence probability of word pairs within that topic. To compute coherence C_V, normalized pointwise mutual information or NPMI is calculated between each word and its corresponding context vector. The NPMI values are then used to construct context vectors for all top words in the topic. For instance, let us consider a topic with the top word "$word_1$". The context vector $\vec{v}word_1$ may be represented as:

$$\vec{v}word_1 = [NPMI(word_1, word_1), NPMI(word_1, word_2), \ldots\ldots] \quad (6)$$

where $word_2$ and $word_3$ represents the second and third top word in the topic. Once the context vectors are computed, C_V measures the similarity between these vectors using cosine similarity [17].

4.3. Thematic Evolution Analysis

The corresponding topics are extracted from different time windows by LDA, and the topics between adjacent time windows may be related. In this paper, cosine similarity was used to measure the similarity between topics in adjacent time windows. The formula for calculating the cosine similarity is shown in Equation (7) [18].

$$\cos(\theta) = \frac{\sum_{i=1}^{n}(x_i \times y_i)}{\sqrt{\sum_{i=1}^{n}(x_i)^2} \times \sqrt{\sum_{i=1}^{n}(y_i)^2}} \quad (7)$$

where x_i, y_i are the vectorized topics, the value range of $\cos(\theta)$ is [0, 1], and the larger the value is, the smaller the angle between the two vectors and the greater the topic similarity. The topic similarity threshold was set. If the cosine similarity between the front adjacent time window topic t_i and the back adjacent time window topic t_{i+1} is greater than the threshold value, t_i and t_{i+1} have a thematic inheritance relationship; if the similarity between t_{i+1} and all the topics in the forward time window is less than the threshold, then t_{i+1} is a new topic. If the similarity between t_i and all backward time windows is less than the threshold, t_i is the vanishing topic. If the similarity between t_{i+1} and two or more topics of the forward time window is greater than the threshold value, then t_{i+1} has a merging relationship with the corresponding topics of the forward time window. If the similarity between t_i and two or more topics in the backward adjacent time window is greater than the threshold value, then t_i has a decomposition relationship with the backward corresponding topics.

5. Results and Discussion

Prior to LDA modeling, the optimal value for K was determined by simultaneously employing the perplexity metric and referencing the methodology of [19], which integrates the C_V coherence score based on word co-occurrence. The governing principle for K selection was: when the coherence score was high, the K value corresponding to the higher score was selected; conversely, when the coherence score was low, the K value corresponding to the minimum perplexity metric was chosen.

5.1. Determination of the Optimal Number of Topics

The optimal number of topics, K, serves as a hyperparameter for the LDA topic model. Its optimal value is determined prior to topic extraction by calculating both Perplexity and the Coherence Score corresponding to different K values. [16] tested the perplexity with 1–25 topics based on 139,962 total data points. [5] tested the perplexity with 1–40 topics; however, recognizing that too many topics are prone to overfitting, only the top-10 topics with high probability were selected for research. This paper took a year as the time window

and tested the perplexity with 1–30 topics in different time windows. The model's iteration count was set to 50, and the random seed was fixed at 42. The K value corresponding to the minimum perplexity was selected if the coherence score was less than 0.5; conversely, the K value associated with the maximum coherence score was chosen if the coherence score exceeded 0.5. This strategic choice is made because a low coherence score implies poor semantic interpretability of the topics, necessitating the pursuit of the minimum perplexity value to represent the theoretically optimal statistical fit of the model. Ref. [17] noted that when assessing model quality using the coherence score, a score of 0.3–0.4 represents low coherence compared to 0.6–0.7. Table 1 summarizes the maximum coherence score, minimum perplexity, and the final K value determined after comprehensive evaluation for each time window. The perplexity and coherence score curve for 2022 is illustrated in Figure 3. To evaluate the topical stability of the LDA model on the relatively small 2016 dataset, 50 independent model runs were performed. Subsequently, the results of these 50 runs were compared using cosine similarity, totaling 1225 comparisons, i.e., C(50,2). The analysis yielded an average cosine similarity of 0.757125, a minimum value of 0.646929, and a maximum value of 0.843639, which collectively demonstrate satisfactory stability of the model.

Table 1. Optimal Topic Number, Perplexity, and Coherence Score for Different Time Windows.

Year	K	Min Perplexity	Max Coherence Scores
2023	30	452.30	0.4442
2022	16	492.89	0.5158
2021	30	432.15	0.4958
2020	30	331.97	0.3772
2019	29	242.55	0.3504
2018	30	139.07	0.3510
2017	26	44.17	0.7409
2016	20	47.79	0.6666

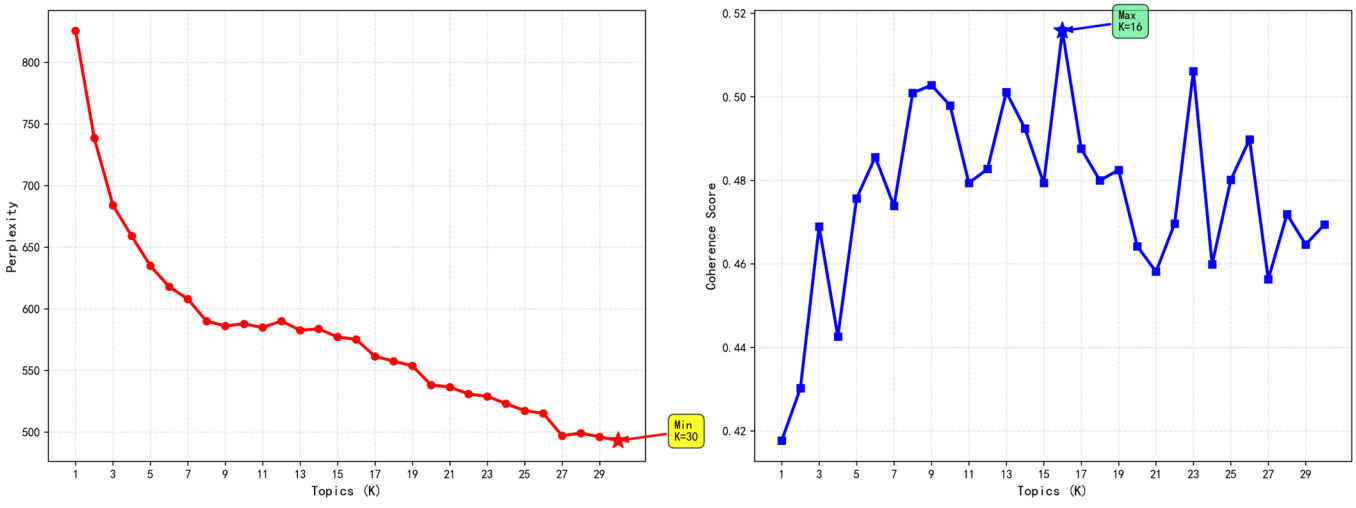


Figure 3. Topic–perplexity and coherence score curves for 2022.

5.2. Topic Extraction Results

Through a text preprocessing technique and based on the optimal K values of different time windows, the LDA topic model was established, and K topics were extracted. At the same time, the word frequency of all topic keywords for different time windows was extracted, and the popular brand or vehicle model for the corresponding time window and the popular spot of consumers for new energy vehicles were obtained by analyzing

the word frequency. As shown in Table 2, all the popular brands in China are domestic brands. The most popular brand in the 2021–2023 reviews was BYD. BYD is a Chinese enterprise that started with the battery industry and was also the first company in the world to officially announce and implement the cessation of fuel vehicle production to focus on the production of pure electric and plug-in hybrid vehicles. The most popular brand according to comments in 2019–2020 was Xpeng Automobile, which is a leading domestic intelligent electric vehicle design and manufacturing brand with the positioning of "young, fashion and technology". In 2018, the most popular brand was LEVDEO Car, and LEVDEO is known as the king of low-speed electric vehicles, but it faced bankruptcy in 2023. In 2017, the most popular brand was ROEWE Automobile, which is the self-owned brand of SAIC Motor, China's largest automobile group, with the positioning of "taste, technology and realization". In 2016, the most popular brand was Red Banner. Red Banner is an old domestic brand that has entered the civilian luxury car market; although it was not listed in the year of pure electric vehicles, it has received increased attention. In the focus of new energy vehicles, mainstream media will mainly investigate aspects such as space, driving experience, control, range, comfort, appearance, interior, cost performance, and intelligence. As shown in Table 2, among the extracted comment topics, the space, appearance and price of new energy vehicles were the most concerning to the Chinese. Figure 4 is a visualization of the 2022 topic extraction. Among them, the number of bubbles is the topic number; the larger the bubble is, the greater the probability of the topic, and the distance between the bubbles indicates the difference between the topics. Figure 4 shows that there is a certain distance between the first 10 topics, and the topic overlap is small. The right side is the topic words in this topic, blue is the weight of the topic words in the document, and red is the weight of the topic words in the topic. Therefore, to avoid too many topics and disperse the similarity analysis among topics, only the 10 topics with the greatest probability in each time window were selected for analysis in subsequent research.

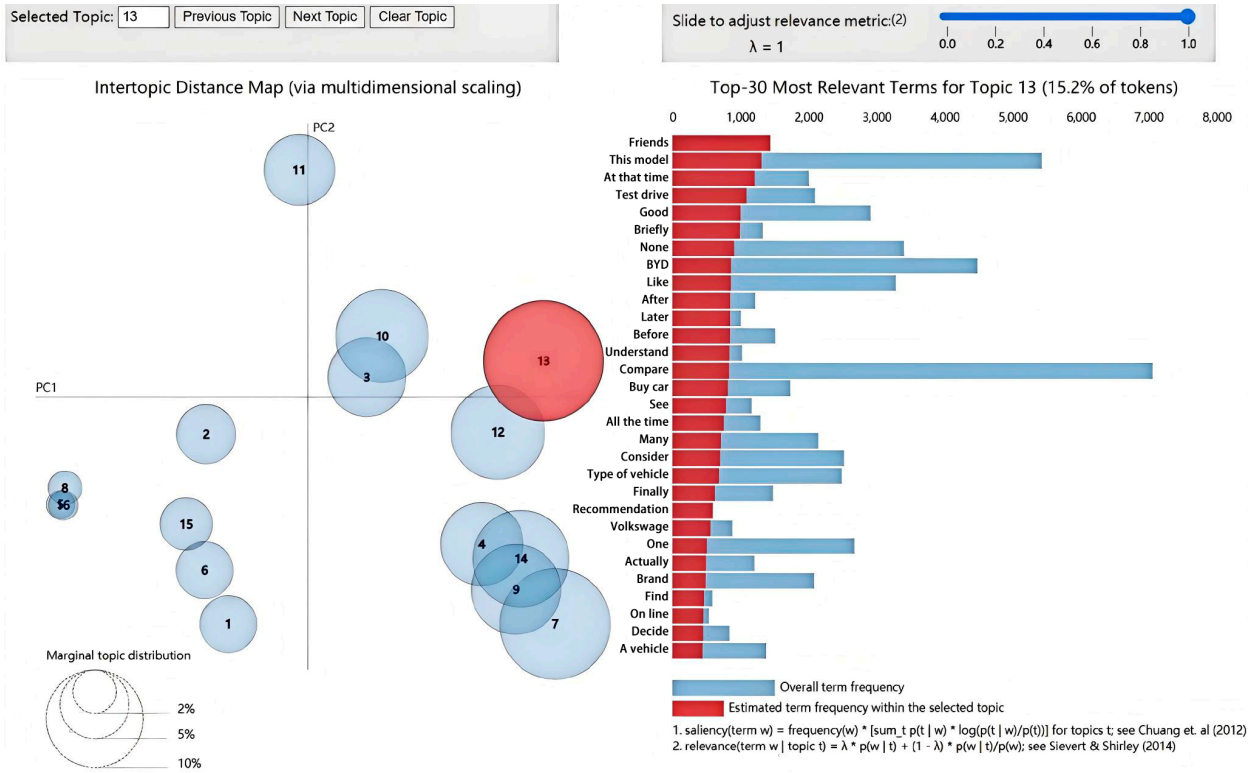


Figure 4. Topic visualization for 2022.

Table 2. Popular brands and focus of attention for different time windows.

Year	Hot Brand	Focus of Attention
2023	BYD	space
2022	BYD	appearance
2021	BYD	appearance
2020	Xpeng	appearance
2019	Xpeng	space
2018	LEVDEO.Mango	price
2017	ROEWE. Ei5	appearance
2016	Red Banner	space

5.3. Similarity Analysis Between Topics

For the topic similarity analysis, the 10 topics with the highest probability in each time window were designated as the analysis objects. Cosine similarity was used to measure the similarity between topics in adjacent time windows. When studying the correlation between topics, ref. [20] used a similarity of 0.5 as the threshold to filter out some correlations between topics. Ref. [21] used the mean similarity as a threshold to filter the correlation between topics. The topic data in this article yields 700 total similar associations, with a mean (μ) of 0.1015 and a standard deviation (σ) of 0.1082. A sensitivity analysis was then conducted across thresholds of [0.2, 0.3, 0.4, 0.5, 0.6], resulting in [102, 44, 21, 4, 1] retained associations, respectively. Ultimately, in this study, 0.3 was selected as the threshold for topic similarity. The reasons for this choice are twofold: (1) Statistical Rationality: The value 0.3 is statistically rational because it is relatively close to the anchor point $\mu + 0.5\sigma = 0.1556$, which serves as an objective reference calculated from the overall similarity distribution. (2) Effective Noise Reduction and Core Preservation: This threshold successfully filtered out 91.5% of weakly correlated elements, retaining 44 core correlations. This significantly reduced the complexity of the topic network. Compared to the excessive filtering and information loss caused by thresholds of 0.5 and 0.6, the 102 associations under the 0.2 threshold make the evolution path overly dense and difficult to interpret. This threshold ensures the coherence and completeness of the topic evolution path and can effectively capture the inheritance and development of the core knowledge in the field.

The maximum similarity of the 2016 theme was 0.2508, which fell below the predefined threshold; thus, it was forcibly incorporated into the topic evolution path to ensure temporal continuity. Conversely, the maximum similarity between the 2019 and 2020 themes reached a high of 0.6721. This high correlation suggests that comment activity regarding NEV has been more active since 2019, and the thematic relevance between adjacent time windows has significantly strengthened. Throughout the thematic evolution, key terms such as “test drive experience” (2019), “hybrid” (2021), and “blade batteries” (2022) began to appear in the high-probability topics of the respective time windows. The topics extracted for each time window, based on the correlation filtering rules, are summarized in Table 3. For the detailed topic, please refer to Appendix A.

Table 3. Related topics extracted for each time window.

Year	2016	2017	2018	2019	2020	2021	2022	2023
Topic	**Top 1**: Aesthetic Preference & Brand (Red Banner)	**Top 2**: Brand (Tesla, ES8, GLS) & Power Preference **Top 3**: NEV Configuration, Appearance, Price, and "Image" Demand	**Top 4**: Space, Price, and Usage Convenience (Parking) **Top 10**: NEV Performance: Range, Li Auto, Value, Configuration	**Top 1**: Aesthetic, Configuration, and Brand (Xpeng) Preference **Top 2**: EV Scenario: Space, Range, Charging, Home Convenience **Top 3**: Analysis of Exterior Configuration **Top 4**: Purchasing Factors: Brand, Price, Configuration, Space, Power **Top 7**: Purchasing Consideration: Value, Range, Space, Charging **Top 8**: Purchase Decision **Top 9**: Interior Quality and Cost-Performance **Top 10**: Range and SUV	**Top 1**: Purchasing Factors: Price, Configuration, Value, Range, Xpeng **Top 2**: Model Selection: Space, Price, Hybrid, Friend Recommendation **Top 3**: NEV Commuting and Home Convenience **Top 4**: Exterior Design & Power Preference **Top 5**: Brand (Xpeng P7, BYD), Range, Appearance, Interior, Configuration **Top 7**: NEV Home Use: Space, Convenience, Configuration, Battery, Power **Top 10**: NEV Daily Use: Convenience, Range, Space, Xpeng	**Top 1:** Performance: Power, Configuration, and Overall Feeling **Top 2**: NEV Usage Needs, Cost, and Usage Consideration **Top 3**: Peer Recommendations **Top 4**: Brand (BYD), Hybrid Tech, Battery, and Fuel Consumption **Top 5**: Price, Configuration, Value, and Budget **Top 6**: Exterior/Interior Design and Aesthetic Preference **Top7**: Family Commuting **Top 9**: Range and Charging **Top 10**: Brand (BYD), Test drive	**Top 1:** Appearance, Interior Design, Configuration, and Aesthetic Preference **Top 2**: Friend Rec., Test Drive, and Brand (BYD) **Top 3**: Purchasing Factors: Space, Price, Value, Configuration, Budget **Top 4**: NEV Daily Commuting and Home Convenience **Top 5**: Brand (BYD), Model, Technology, and Quality **Top 6**: Driving Feeling and Experience **Top 7**: NEV Brand Comparison (Xpeng, Tesla, NIO, Weltmeister) **Top 8**: Convenience of Charging, Range, and Maintenance	**Top 1**: Purchasing Factors: Appearance, Price, Value, Configuration, BYD **Top 2**: NEV Range, Commuting, Charging, and Space **Top 3**: Purchasing Factors: Space, Power, Appearance, Fuel Economy, Aesthetic **Top 4**: Purchasing Consideration: Fuel, Home Use, Commuting, Cost **Top 5**: Friend Rec. and Research Process (MG) **Top 6**: Family Preference (Wife), Test Drive, and Brand (VW)

5.4. Topic Evolution Analysis

The Sankey diagram was drawn according to the similarity between topics and the set threshold, as shown in Figure 5. The topics in the diagram are connected by similarity, where line thickness corresponds to the degree of correlation. Each node represents a topic within the corresponding time window. A newly emergent node (e.g., 2017 Top 3) has no incoming links from the previous time window, while a disappeared node (e.g., 2017 Top 2) has no outgoing links to the subsequent time window. A continuous theme evolution path—such as the inheritance of 2019 Top 1 by 2020 Top 1, 2021 Top 1, 2022 Top 1, and 2023 Top 1—is consistently centered on aesthetic design, configuration, and brand preference. To better analyze the thematic evolution, the maximum weights of key terms have been summarized in Table 4; these weights are derived from various topics, with only the maximum values extracted for each term. Simultaneously, the topic frequency—defined as the number of occurrences of a specific topic in the comments divided by the total number of comments—is presented in Table 5; notably, 'TOP 1' identifies the first topic for each year, the nature of which varies annually.

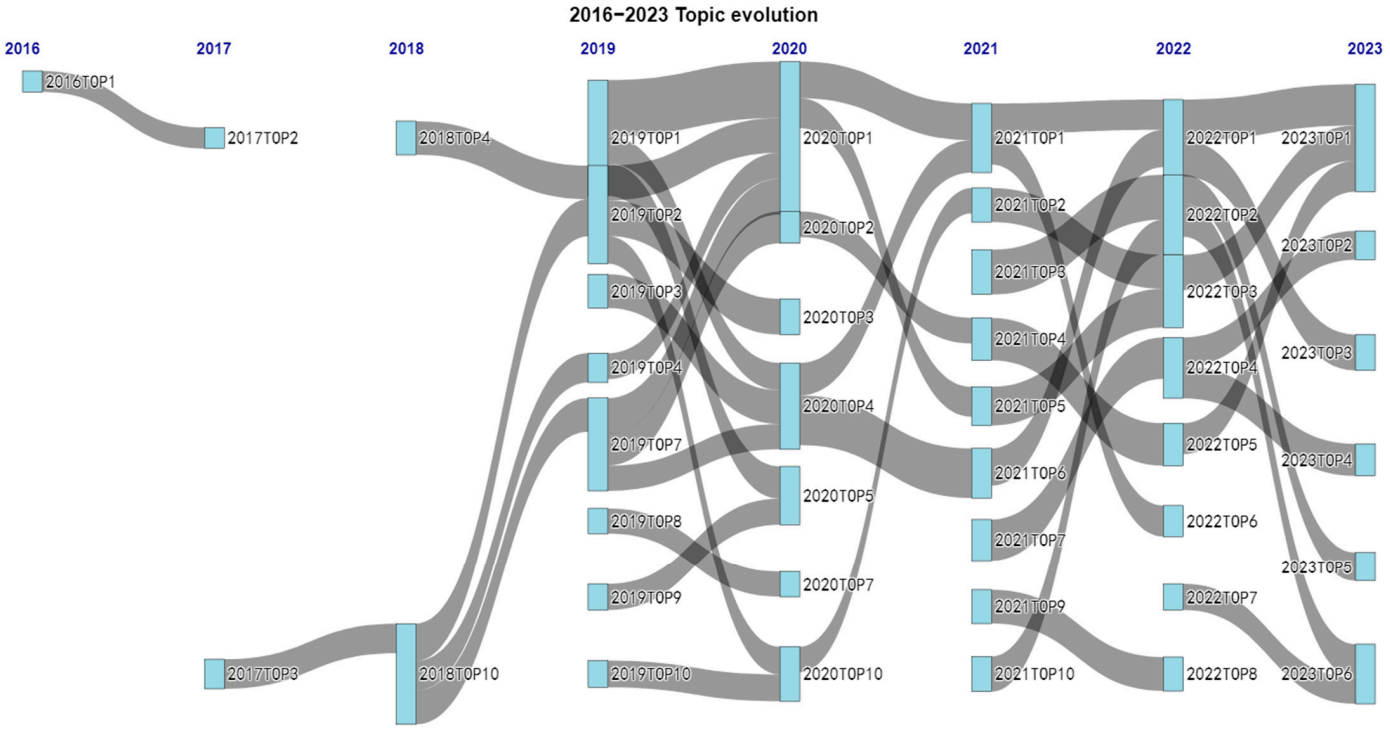


Figure 5. Topic evolution Sankey diagram.

Table 4. Key Term Weights.

Key Terms	2016	2017	2018	2019	2020	2021	2022	2023
BYD	/	/	/	0.006	0.014	0.112	0.072	0.019
Tesla	/	0.03	/	/	/	/	0.018	/
NIO	/	/	/	/	/	/	0.01	/
Xpeng's	/	/	/	0.024	0.061	/	0.019	/
P7	/	/	/	/	0.029	/	/	/
Space	/	/	0.011	0.02	0.012	/	0.049	0.033
Range	/	/	0.011	0.026	0.021	0.102	/	0.027
Configuration	0.04	0.04	0.008	0.035	0.025	0.049	0.024	0.012
Appearance	0.021	0.027	0.009	0.027	0.05	0.109	0.057	0.021

Table 4. *Cont.*

Key Terms	2016	2017	2018	2019	2020	2021	2022	2023
Technology	/	/	/	/	/	0.021	0.019	/
Charging	/	/	/	0.016	0.02	0.009	0.057	0.014
Driving	/	/	/	0.014	0.057	/	0.022	/
Brand	/	/	/	/	/	0.032	0.037	/
Convenience	/	/	0.016	0.007	0.009	0.051	0.035	0.022
Commuting	/	/	/	/	/	0.047	0.034	0.014
Little Ant	/	/	/	/	/	0.018	/	/
Hybrid	/	/	/	/	0.01	0.03	/	/
Quality	/	/	/	/	/	/	0.013	/
Experience	/	/	/	/	/	/	0.018	/
fuel	/	/	/	/	/	0.029	/	0.02
cost	/	/	/	/	/	0.032	/	0.014
friends	/	/	/	0.008	0.009	/	0.023	0.03
recommendation	/	/	/	/	/	/	/	0.013
family	/	/	0.014	0.007	0.012	0.02	0.023	0.015
wife	/	/	/	/	/	0.016	/	0.017
Volkswagen	/	/	/	/	/	/	/	0.009
Test Drive	/	/	/	0.01	0.007	0.017	0.017	0.01
Smart	/	/	/	0.007	0.027	/	/	/
Issues	/	/	/	/	0.01	0.01	0.018	/

Table 5. Topic Frequencies from 2016 to 2023.

Topic	2016	2017	2018	2019	2020	2021	2022	2023
TOP 1	13.16%	/	/	9.74%	12.26%	15.66%	20.11%	24.33%
TOP 2	/	6.67%	/	6.77%	8.79%	11.22%	18.74%	12.80%
TOP 3	/	6.67%	/	6.09%	8.42%	8.75%	11.11%	8.91%
TOP 4	/	/	4.98%	5.01%	6.41%	7.66%	12.53%	7.60%
TOP 5	/	/	/	/	6.22%	6.85%	11.54%	7%
TOP 6	/	/	/	/	/	5.32%	6.69%	6.60%
TOP 7	/	/	/	4.47%	3.97%	5.86%	6.60%	/
TOP 8	/	/	/	4.60%	/	/	4.76%	/
TOP 9	/	/	/	4.06%	/	3.62%	/	/
TOP 10	/	/	3.56%	4.06%	3.66%	5.32%	/	/

The 2017 Topic 2 (“Brand (Tesla, ES8, GLS) & Power Preference”), inherited from the 2016 Topic 1 (“Aesthetic Preference & Brand (Red Banner)”), subsequently disappeared. This reflects the evolution of early luxury car brands and the sustained preference for power. The 2018 Topic 10 was inherited from the new 2017 theme (“NEV Performance: Range, Li Auto, Value, Configuration”), and their similarity is 0.3520. The 2018 Topic 4 (“Space, Price, and Usage Convenience (Parking)”) newly emerged. The 2018 Topic 10 was further differentiated into the 2019 Topic 2 (“Car Buying Elements”) and the 2019 Topic 7 (“Car Buying Considerations”).

The 2019 Topic 2 (“NEV Scenario: Space, Range, Charging, Home Convenience”) is a synthesis of the 2018 Topics 4 and 10, with its similarity to the 2018 Topic 10 reaching 0.4433, where the weight of the word “Space” is the highest at 0.020, and the weight of “Range” is mainly inherited from the 2018 Topic 10. The 2019 Topic 2 is driven by ‘Configuration’ (0.035) and aesthetic terms like ‘Aesthetics’ (0.029), with the frequency of this topic rising from 3.56% to 6.77%. This shift establishes Xpeng’s (0.024) brand leadership based on product design and feature focus. The 2019 themes, namely: the 1st, 3rd, 8th, 9th, and 10th, were all new. Concurrently, the distinct weighting of key

purchasing factors—'Test Drive' (0.010) and 'Smart' (0.007) in Topic9 and Topic8 confirms the emergence of new buying trends driven by experiential factors and intelligent features. During this stage, there is a significant differentiation and re-aggregation of topics, and the considerations for purchasing a vehicle begin to become more detailed.

The 2020 Topic 5 (Prob = 0.0669) consolidated brand and product attributes, evidenced by the high weight of 'Xpeng' (0.061), its model 'P7' (0.029), and core factors like 'Appearance' (0.021) and 'Configuration' (0.018), while also aggregating 'BYD' (0.014). Concurrently, the 2020 Topic 10 (Prob = 0.0315) synthesized the use-case focus of the 2019 themes, with high weights on 'Convenience' (0.009), 'Range' (0.008), and 'Space' (0.007), confirming a reinforced focus on NEV daily usage scenarios. The thematic evolution during this phase was characterized by massive consolidation. The frequency of Topic 1, centered on vehicle purchasing factors, rose from 9.74% to 12.26%, indicating that a large volume of segmented themes coalesced around new emerging brands such as 'Xpeng/BYD' and essential vehicle purchasing factors.

In 2021, two new salient themes emerged: Topic 7 (Prob = 0.0506) focused on "Daily Commuting & Home Use," indicated by the high weights for 'Convenience' (0.051), 'Commuting' (0.047), and the specific micro-NEV brand 'Little Ant' (0.018), signifying that consumers began to view NEV as the second family car or urban transportation tool. Simultaneously, Topic 9 (Prob = 0.0393) centered on performance concerns, dominated by 'Range' (0.102) and explicitly mentioning 'Charging' (0.009) and 'Issues' (0.010), reflecting a significant rise in user attention to actual range performance and charging infrastructure. Furthermore, the 2021 Topic 4 (Prob = 0.0629) evolved from the 2020 Topic 2, with a similarity score of 0.3064, by hyper-focusing on BYD (0.112 weight) and its 'Hybrid' (0.030) and 'Technology' (0.021) attributes, confirming a market shift toward recognizing the technological breakthroughs and resurgence of established manufacturers.

In 2022, Topic 7, with a probability of 0.0636, newly emerged. As shown in Table 4, several major brands appeared simultaneously with high weights, such as 'Xpeng' (0.019), 'Tesla' (0.018), and 'NIO' (0.010), and the frequency of this topic reached 6.6% in Table 5. This indicates that consumer focus shifted from "whether to buy an NEV" to conducting in-depth comparisons among specific NEV brands. Concurrently, Topic 5, which significantly increased its probability from 0.0629 (2021 Topic 4) to 0.0993, also saw its frequency rise from 7.66% to 11.54%, evolving to further discuss BYD. This theme reflects the brand's widespread market recognition, demonstrated by the consistent high weight for 'BYD' (0.072) and a growing emphasis on softer attributes like 'Brand' (0.037), 'Technology' (0.019), and 'Quality' (0.013). Additionally, Topic 6 (Prob = 0.0725) represents a shift in consumer demand from hard indicators to soft ones, inheriting the focus on 'Driving' (0.022) and 'Experience' (0.018), signifying an upgrade from basic configurations and range to actual driving pleasure and comfort.

Topic 4 in 2023, with a similarity of 0.3826, evolved from its 2022 counterpart. The critical shift in this topic is demonstrated by the emergence of 'fuel' (0.020) and 'cost' (0.014), explicitly reflecting consumers' meticulous calculation and comparison of energy usage costs between NEVs and fuel vehicles. Topic 5 (Prob = 0.0695) is a continuation of the social influence theme from the 2022 Topic 2, confirmed by the highest weights for 'friends' (0.030) and 'recommendation' (0.013), validating the crucial role of Word-of-Mouth Marketing (WOM) and social recommendation in car purchasing decisions. Furthermore, Topic 6 (Prob = 0.0643) highlights a trend towards family-oriented and collaborative decision-making, evidenced by high weights for 'family' (0.020) and 'wife' (0.017). The mention of 'Volkswagen' (0.009) also indicates that traditional joint-venture brands are beginning to regain influence in the NEV sector. This collective market shift suggests the segment has entered a phase focused on user cost calculation and family-centric decision-making.

Overall, the decrease in frequency and keyword weights for certain themes in 2023 compared to 2022 may be attributed to the incompleteness of the 2023 dataset. During the thematic evolution, the frequency of topics related to purchasing factors—such as aesthetics, configuration, price, and brand—increased steadily from 9.74% in 2019 to 24.33% in 2023. Meanwhile, thematic similarity fluctuated downward from 0.672 in 2019 to 0.4898 in 2023, indicating that while these topics receive sustained attention, their specific content is constantly evolving. In 2022, several major brands appeared simultaneously with high weights, signifying that consumers focused heavily on comparisons between these leading players. The weights of fundamental indicators like configuration (0.049) and aesthetics (0.109) declined after 2021 to 0.012 and 0.021, respectively. Concurrently, terms related to quality and experience began to emerge with higher weights in 2022, marking a shift from basic features and range toward actual driving pleasure and comfort. In 2023, terms such as "friends" (0.03) and "recommendation" (0.013) appeared with high weights, and the weight of 'friends' has also shown a year-on-year increase, highlighting the growing importance of word-of-mouth marketing in NEV sales. Additionally, family purchasing preferences emerged as a distinct theme in 2023, reaching a frequency of 6.6%.

6. Conclusions

The LDA topic model, optimized by perplexity and coherence scores, can effectively extract themes from Chinese comments. Topic extraction and evolutionary analysis of user comments on new energy vehicles can aid in understanding changes in consumer attitude under different policy and technical backgrounds. This paper draws the following conclusions through the analysis of big data on user reviews of new energy vehicles in China:

(1) The main concerns of Chinese consumers for new energy vehicles are space, appearance and price. The most discussed brand is BYD.

(2) Hierarchical Ascent: From Pragmatism to Experiential Economy

The focus of electric vehicle consumption has shifted from early fundamental functions (e.g., range, space, price) to higher-level experiential demands. This trend aligns with Maslow's Hierarchy of Needs, where consumers, having satisfied the basic attributes of a transport tool, now pursue the emotional value and self-actualization derived from "exterior design," "driving experience," and "interior aesthetics." The evolved 2022 Topic 5 (Prob. increasing from 0.0629 to 0.0993), with its emphasis on 'Brand' (0.037) and 'Quality' (0.013), as well as the focus of 2022 Topic 6 (Prob = 0.0725) on 'Driving' (0.022) and 'Experience' (0.018), all confirm this market shift. This market shift necessitates continuous innovation from automakers in design and driving pleasure. Automakers should increase investment in enhancing experiential attributes, such as vehicle tuning, interior materials, and Human–Machine Interface (HMI) development.

(3) Decision Model Transformation: Driven by Social Influence and Family Collaboration

In high-value car purchasing decisions, there is a distinct evolution in consumers' information sources and power structures. On one hand, the sustained prominence of the "friend recommendation and test drive" theme, evidenced by 2023 Topic 5 with the highest weights for 'friends' (0.030) and 'recommendation' (0.013), validates the central role of Word-of-Mouth Marketing (WOM) in high-risk choices, where existing users' evaluations are key factors influencing potential buyers. On the other hand, the increasing visibility of "family preference" suggests that car buying has become a form of Household Collaborative Decision-Making, requiring products to cater to diverse family requirements. Sales departments should establish comprehensive mechanisms for existing user incentives and experience sharing, reallocating marketing resources toward community operations and

targeted test drive campaigns, and designing multi-functional, high-safety, family-friendly configurations aimed at household decision-makers.

(4) Market Competition Intensification: New Entrants Pioneer Innovation, Incumbents Revert to Rationality

The thematic evolution reveals the dynamic competitive landscape of the NEV market. Initially, new players represented by Xpeng and NIO rapidly dominated conversations by aggregating themes around intelligence and configuration, serving as the pioneers of Innovation Diffusion. Xpeng appeared in multiple 2020 topics, with its highest weight being 0.061. As the market matures, established manufacturers, such as BYD with its hybrid technology, are regaining attention by leveraging technical accumulation, innovation, and existing brand trust (e.g., Volkswagen). Consequently, both new entrants and incumbent brands must maintain a balance between technological advancement and cost control. Traditional manufacturers should leverage their strengths in supply chain management and quality control while actively embracing intelligent features; conversely, new players must translate their technological leadership into measurable reliability and economies of scale.

(5) Market Segmentation Deepening: Pursuit of Specific Utility and Energy Cost Calculation

The latter thematic fragmentation illustrates the market's need for specialized utility, which is consistent with the Consumer Utility Maximization Principle. On one hand, the emergence of "daily commuting" and micro-vehicle themes signifies that consumers are positioning NEVs as tools for satisfying specific needs like urban transit. On the other hand, the contrast between "fuel vs. cost" has become a central focus, reflecting consumers' detailed calculations amid economic pressures, leading them to base purchasing decisions more on a rational trade-off regarding energy usage costs. Production departments should promote the diversification and precision of the product matrix. They must develop customized solutions for various sub-scenarios (e.g., urban commuting, family long-distance, premium experience) and clearly articulate the Total Life-Cycle Energy Cost Advantage through transparent data and financial instruments to meet consumers' rigorous calculation demands.

7. Limitations and Future Work

The main limitations of this study lie in the singularity of the data source and temporal constraints. Data collected solely from a single social media platform (Auto Home) may introduce inherent limitations and biases. The Auto Home user base is skewed toward active car purchase decision-makers who tend to focus on pre-purchase themes. Consequently, this specific data source may suffer from selection and demographic biases, limiting its ability to fully reflect the characteristics and opinions of the general consumer population. Furthermore, given the potentially incomplete nature of the 2023 data, the research findings offer a limited reflection of the overall market situation for that year. Additionally, this study focuses primarily on the evolution of thematic content, lacking an in-depth analysis of consumer sentiment, which ultimately limits a comprehensive understanding of complex consumer behavior.

Despite its inherent limitations, the topic trend analysis methodology employed in this study remains an important tool. This method is primarily based on LDA, a model that has demonstrated effectiveness in extracting meaningful topics from Twitter and other social media platforms [22,23]. Critically, LDA's statistical foundation provides strong cross-market generalizability [24]: as an unsupervised probabilistic model, it is inherently language-agnostic and operates on word co-occurrence patterns rather than culture-specific semantic features. One must acknowledge that the fast-paced evolution of the market

post-2023 might influence the long-term applicability of the conclusions drawn from the 2016–2023 data. Future research will build upon this topic trend analysis by specifically focusing on integrating multi-platform data to mitigate bias and incorporating both real-time monitoring and fine-grained sentiment analysis. This approach aims to construct a more accurate and comprehensive model for capturing and forecasting the impact of market events on consumer behavior.

Author Contributions: Conceptualization, S.H.; Methodology, S.H.; Software, S.H.; Validation, B.X. and D.L.; Formal analysis, S.H.; Investigation, S.H.; Resources, S.H.; Writing—original draft, S.H.; Writing—review & editing, S.H.; Visualization, S.H.; Supervision, S.H.; Project administration, S.H.; Funding acquisition, S.H. All authors have read and agreed to the published version of the manuscript.

Funding: This study was supported by the Sichuan Mineral Resources Research Center (No. SCKCZY2024-ZC004).

Data Availability Statement: The raw data supporting the conclusions of this article will be made available by the authors on request.

Conflicts of Interest: Author Bo Xue was employed by the company China Mobile Communications Group Sichuan Yibin Branch. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Appendix A. LDA Results of Topics in Thematic Evolution

2016 TOP 1 (Prob = 0.1231) 0.057 * "Compare" + 0.042 * "This model" + 0.034 * "Very" + 0.029 * "Good" + 0.025 * "Really" + 0.025 * "Like" + 0.025 * "Specially" + 0.022 * "Red Banner" + 0.022 * "Aspect" + 0.021 * "Exterior"
2017 TOP 2 (Prob = 0.0648) 0.059 * "This model" + 0.030 * "Tesla" + 0.020 * "ES8" + 0.020 * "Like" + 0.020 * "Compare" + 0.020 * "On the road" + 0.020 * "GLS" + 0.020 * "Specially" + 0.020 * "Decisively" + 0.020 * "Power" **2017 TOP 3 (Prob = 0.0628)** 0.040 * "Configuration" + 0.027 * "Exterior" + 0.027 * "Brand" + 0.027 * "This car" + 0.027 * "Price" + 0.027 * "Electric vehicle" + 0.014 * "Finally" + 0.014 * "Driving" + 0.014 * "Good" + 0.014 * "Keeping up appearances"
2018 TOP 4 (Prob = 0.0474) 0.016 * "Parking" + 0.016 * "Convenient" + 0.014 * "This model" + 0.014 * "Home/Family" + 0.014 * "Place" + 0.013 * "One" + 0.012 * "Price" + 0.011 * "Compare" + 0.011 * "Space" + 0.011 * "Very" **2018 TOP 10 (Prob = 0.0338)** 0.011 * "Electric vehicle" + 0.011 * "Li Auto" + 0.011 * "Range" + 0.009 * "This model" + 0.009 * "Good" + 0.009 * "Exterior" + 0.009 * "Mainly" + 0.008 * "Cost-performance" + 0.008 * "Performance" + 0.008 * "Configuration"
2019 TOP 1 (Prob = 0.0957) 0.044 * "Compare" + 0.037 * "This model" + 0.035 * "Configuration" + 0.029 * "Aesthetics" + 0.024 * "Xpeng" + 0.022 * "Very" + 0.022 * "Model" + 0.021 * "Exterior" + 0.017 * "Aspect" + 0.016 * "Like" **2019 TOP 2 (Prob = 0.0619)** 0.020 * "Space" + 0.018 * "Compare" + 0.016 * "Charging" + 0.015 * "Range" + 0.010 * "Electric vehicle" + 0.008 * "This model" + 0.007 * "Convenient" + 0.007 * "Family" + 0.007 * "Configuration" + 0.007 * "Cost-performance" **2019 TOP 3 (Prob = 0.0590)** 0.026 * "Exterior" + 0.025 * "Compare" + 0.014 * "Driving" + 0.012 * "This model" + 0.012 * "Space" + 0.011 * "Configuration" + 0.010 * "Assist" + 0.010 * "Like" + 0.010 * "Good-looking" + 0.010 * "None" **2019 TOP 4 (Prob = 0.0470)** 0.026 * "Range" + 0.022 * "Compare" + 0.012 * "This model" + 0.011 * "Electric vehicle" + 0.010 * "Space" + 0.009 * "Mileage" + 0.008 * "Ei5" + 0.008 * "None" + 0.007 * "Cost-performance" + 0.007 * "Like" **2019 TOP 7 (Prob = 0.0416)** 0.018 * "Very" + 0.013 * "Good" + 0.013 * "This model" + 0.012 * "Compare" + 0.012 * "Space" + 0.012 * "Cost-performance" + 0.010 * "None" + 0.008 * "Range" + 0.008 * "Consider" + 0.008 * "Charging" **2019 TOP 8 (Prob = 0.0403)** 0.011 * "Very" + 0.009 * "Xpeng" + 0.009 * "Buying a car" + 0.009 * "P5" + 0.008 * "None" + 0.008 * "Test drive" + 0.007 * "Intelligent" + 0.007 * "Consider" + 0.007 * "Needs" + 0.007 * "Space" **2019 TOP 9 (Prob = 0.0382)** 0.011 * "Interior" + 0.010 * "Test drive" + 0.009 * "Cost-performance" + 0.009 * "Space" + 0.009 * "Later" + 0.008 * "Friends" + 0.008 * "Exterior" + 0.007 * "None" + 0.006 * "Model" + 0.006 * "BYD" **2019 TOP 10 (Prob = 0.0378)** 0.017 * "One" + 0.013 * "Consider" + 0.012 * "Space" + 0.011 * "None" + 0.009 * "Mileage" + 0.009 * "Li Auto" + 0.009 * "Ei5" + 0.009 * "Range" + 0.009 * "SUV" + 0.008 * "Mainly"

2020 TOP 1 (Prob = 0.1236) 0.054 * "Compare" + 0.039 * "This model" + 0.030 * "Price" + 0.025 * "Configuration" + 0.021 * "Cost-performance" + 0.021 * "Model" + 0.019 * "Aspect" + 0.016 * "Very" + 0.016 * "Range" + 0.013 * "Xpeng"
2020 TOP 2 (Prob = 0.0760) 0.015 * "Model" + 0.014 * "Very" + 0.012 * "This model" + 0.012 * "Space" + 0.011 * "Compare" + 0.010 * "Price" + 0.010 * "Mainly" + 0.010 * "Hybrid" + 0.010 * "Good" + 0.009 * "Friends"
2020 TOP 3 (Prob = 0.0722) 0.030 * "Compare" + 0.015 * "One" + 0.014 * "A vehicle" + 0.012 * "Family" + 0.011 * "None" + 0.011 * "Convenient" + 0.010 * "Problem" + 0.009 * "This model" + 0.008 * "Commuting" + 0.008 * "Electric vehicle"
2020 TOP 4 (Prob = 0.0710) 0.050 * "Exterior" + 0.037 * "Very" + 0.026 * "Compare" + 0.023 * "This model" + 0.019 * "Good" + 0.019 * "Good-looking" + 0.018 * "Power" + 0.018 * "Like" + 0.017 * "Design" + 0.017 * "Specially"
2020 TOP 5 (Prob = 0.0669) 0.061 * "Xpeng" + 0.029 * "P7" + 0.021 * "Range" + 0.021 * "Exterior" + 0.018 * "Configuration" + 0.018 * "Price" + 0.015 * "Interior" + 0.015 * "Good" + 0.014 * "BYD" + 0.012 * "Aesthetics"
2020 TOP 7 (Prob = 0.0341) 0.010 * "None" + 0.009 * "Actually" + 0.009 * "Briefly" + 0.009 * "Consider" + 0.007 * "Really" + 0.007 * "Space" + 0.007 * "Xpeng" + 0.007 * "Finally" + 0.007 * "Decide" + 0.007 * "Test drive"
2020 TOP 10 (Prob = 0.0315) 0.010 * "Electric vehicle" + 0.009 * "Convenient" + 0.009 * "Consider" + 0.009 * "Usually" + 0.008 * "Range" + 0.008 * "None" + 0.008 * "Use" + 0.007 * "One" + 0.007 * "Xpeng" + 0.007 * "Space"
2021 TOP 1 (Prob = 0.1081) 0.055 * "Very" + 0.049 * "Compare" + 0.048 * "This model" + 0.031 * "Aspect" + 0.027 * "Power" + 0.020 * "Configuration" + 0.019 * "Good" + 0.018 * "Performance" + 0.015 * "Many" + 0.014 * "One"
2021 TOP 2 (Prob = 0.0829) 0.044 * "Vehicle use" + 0.039 * "Compare" + 0.032 * "Cost" + 0.024 * "Electric vehicle" + 0.022 * "Needs" + 0.022 * "Travel" + 0.022 * "Satisfy" + 0.021 * "This model" + 0.019 * "Consider" + 0.017 * "Use"
2021 TOP 3 (Prob = 0.0722) 0.030 * "Compare" + 0.015 * "One" + 0.014 * "Vehicle" + 0.012 * Family" + 0.011 * "None" + 0.011 * "Convenient" + 0.010 * "Issue" + 0.009 * "This Model" + 0.008 * "Commuting" + 0.008 * "EV"
2021 TOP 4 (Prob = 0.0629) 0.112 * "BYD" + 0.038 * "Battery" + 0.030 * "Hybrid" + 0.029 * "Fuel consumption" + 0.026 * "This model" + 0.024 * "Brand" + 0.023 * "Consider" + 0.023 * "Compare" + 0.022 * "Model" + 0.021 * "Technology"
2021 TOP 5 (Prob = 0.0589) 0.058 * "Price" + 0.049 * "Configuration" + 0.048 * "Cost-performance" + 0.046 * "Compare" + 0.039 * "Budget" + 0.027 * "Price point" + 0.016 * "Range" + 0.014 * "None" + 0.013 * "Exterior" + 0.012 * "Mainly"
2021 TOP 6 (Prob = 0.0544) 0.109 * "Exterior" + 0.060 * "Like" + 0.044 * "Interior" + 0.040 * "Very" + 0.039 * "Design" + 0.035 * "Good-looking" + 0.026 * "Fashionable" + 0.018 * "Specially" + 0.018 * "Attract" + 0.016 * "Atmospheric"
2021 TOP 7 (Prob = 0.0506) 0.051 * "Convenient" + 0.047 * "Commuting" + 0.030 * "Child" + 0.030 * "Commuting" + 0.025 * "Usually" + 0.022 * "Pick up" + 0.020 * "Family" + 0.018 * "Little Ant" + 0.017 * "Parking" + 0.016 * "Wife"
2021 TOP 9 (Prob = 0.0393) 0.102 * "Range" + 0.036 * "Kilometers" + 0.029 * "Mileage" + 0.012 * "Pure electric" + 0.010 * "Currently" + 0.010 * "Problem" + 0.009 * "One" + 0.009 * "Charging" + 0.009 * "None" + 0.009 * "EV"
2021 TOP 10 (Prob = 0.0364) 0.032 * "Brand" + 0.020 * "BYD" + 0.019 * "Finally" + 0.017 * "Test drive" + 0.015 * "Sales" + 0.015 * "At that time" + 0.012 * "None" + 0.011 * "Like" + 0.011 * "One" + 0.010 * "Finally"
2022 TOP 1 (Prob = 0.1588) 0.057 * "Exterior" + 0.045 * "Compare" + 0.043 * "This model" + 0.035 * "Very" + 0.031 * "Like" + 0.024 * "Configuration" + 0.018 * "Interior" + 0.017 * "Design" + 0.017 * "Aesthetics" + 0.016 * "Good"
2022 TOP 2 (Prob = 0.1359) 0.023 * "Friends" + 0.021 * "This model" + 0.019 * "At that time" + 0.017 * "Test drive" + 0.016 * "Good" + 0.016 * "Briefly" + 0.014 * "None" + 0.014 * "BYD" + 0.014 * "Like" + 0.013 * "After"
2022 TOP 3 (Prob = 0.1014) 0.049 * "Space" + 0.043 * "Price" + 0.027 * "Compare" + 0.026 * "Needs" + 0.022 * "Satisfy" + 0.021 * "Budget" + 0.020 * "Cost-performance" + 0.017 * "Vehicle use" + 0.017 * "Configuration" + 0.016 * "This model"
2022 TOP 4 (Prob = 0.1012) 0.035 * "Convenient" + 0.034 * "Commuting" + 0.026 * "Transport" + 0.024 * "Compare" + 0.023 * "Family" + 0.020 * "EV" + 0.020 * "Mainly" + 0.019 * "Usually" + 0.016 * "Consider" + 0.016 * "Vehicle use"
2022 TOP 5 (Prob = 0.0993) 0.072 * "BYD" + 0.037 * "Brand" + 0.031 * "Compare" + 0.020 * "Model" + 0.019 * "Technology" + 0.017 * "This model" + 0.016 * "Electric vehicle" + 0.015 * "Very" + 0.013 * "Quality" + 0.012 * "Aspect"
2022 TOP 6 (Prob = 0.0725) 0.025 * "This unit" + 0.022 * "Driving" + 0.021 * "Compare" + 0.020 * "None" + 0.018 * "One" + 0.018 * "Feeling" + 0.017 * "Starting up" + 0.016 * "Aspect" + 0.016 * "Very" + 0.015 * "Good"
2022 TOP 7 (Prob = 0.0636) 0.019 * "Xpeng" + 0.019 * "None" + 0.018 * "Tesla" + 0.012 * "WM Motor" + 0.011 * "Interior" + 0.011 * "EV" + 0.010 * "NIO" + 0.010 * "Finally" + 0.010 * "Like" + 0.008 * "Fuel car"
2022 TOP 8 (Prob = 0.0555) 0.057 * "Charging" + 0.027 * "No need" + 0.018 * "Problem" + 0.018 * "Convenient" + 0.018 * "Worry" + 0.015 * "Kilometers" + 0.014 * "One" + 0.011 * "Maintenance" + 0.010 * "None" + 0.010 * "Refueling"

2023 TOP 1 (Prob = 0.1899) 0.054 * "Compare" + 0.035 * "This model" + 0.030 * "Very" + 0.021 * "Exterior" + 0.020 * "Price" + 0.019 * "Aspect" + 0.019 * "BYD" + 0.017 * "Good" + 0.013 * "Cost-performance" + 0.012 * "Configuration"
2023 TOP 2 (Prob = 0.1068) 0.027 * "Range" + 0.024 * "Compare" + 0.018 * "Convenient" + 0.014 * "Commuting" + 0.014 * "Charging" + 0.014 * "Satisfy" + 0.013 * "This model" + 0.013 * "Transport" + 0.012 * "Use" + 0.012 * "Space"
2023 TOP 3 (Prob = 0.0863) 0.033 * "Space" + 0.031 * "Power" + 0.022 * "Exterior" + 0.021 * "This model" + 0.019 * "Fuel consumption" + 0.018 * "Like" + 0.016 * "Aesthetics" + 0.013 * "Very" + 0.013 * "Fit" + 0.011 * "Aspect"
2023 TOP 4 (Prob = 0.0742) 0.025 * "Consider" + 0.022 * "Convenient" + 0.020 * "Fuel" + 0.018 * "This model" + 0.017 * "Compare" + 0.015 * "Family" + 0.015 * "A unit" + 0.014 * "Commuting" + 0.014 * "Cost" + 0.014 * "A vehicle"
2023 TOP 5 (Prob = 0.0695) 0.030 * "Friends" + 0.016 * "Briefly" + 0.015 * "MG" + 0.015 * "At that time" + 0.014 * "Many" + 0.014 * "None" + 0.013 * "Understand" + 0.013 * "Later" + 0.013 * "Before" + 0.013 * "Recommendation"
2023 TOP 6 (Prob = 0.0643) 0.020 * "Family" + 0.017 * "Wife" + 0.013 * "Like" + 0.012 * "Briefly" + 0.011 * "None" + 0.010 * "Test drive" + 0.009 * "Volkswagen" + 0.009 * "Good" + 0.008 * "Finally" + 0.008 * "EV"

References

1. Ju, Q.-j.; Ju, P.; Dai, W.-q.; Ran, L. Adoption of new energy vehicles under subsidy policies: Unit subsidies, sales incentives and product differentiation. *J. Manag. Sci. China* **2021**, *24*, 101–116.
2. Dai, H. Review of New Energy Vehicle Subsidy Policy Effect and Future Adjustment Suggestions. *Price Theory Pract.* **2021**, *9*, 28–30.
3. Lan, Y. More than 6.8 million new energy vehicles will be sold in 2022. *Auto Rev.* **2023**, *2*, 106–107.
4. Wang, X.w.; Li, Y.q.; Liu, T.y.; Zhang, L. Research on the Collaborative Model of Sentiment Analysis and Topic Mining of Micro-blogging Users in the Context of COVID-19. *J. China Soc. Sci. Tech. Inf.* **2021**, *40*, 223–233.
5. Zhang, L.; Tan, H.-w.; Zhang, X.; Han, L. Evolution and Path of Public Opinions about Teachers' Ethics in Colleges and Universities Based on LDA Model. *Inf. Sci.* **2022**, *40*, 144–151.
6. Yue, Z.-j.; Zhang, C.-z.; Zhou, Q.-q. Mining User Diet Preference with Online Reviews—A Case Study of Beijing City. *Libr. Trib.* **2017**, *37*, 108–115.
7. Gao, H.-y.; Liu, J.-w.; Yang, S.-x. Identifying Topics of Online Healthcare Reviews Based on Improved LDA. *Trans. Beijing Inst. Technol.* **2019**, *39*, 427–434.
8. Zhao, Z.-j.; Liu, Y.; Zhang, Y.; Zhou, W.-T.; Meng, L.-Y. Hotel User Preference Model Based on LASSO-LDA. *Comput. Appl. Softw.* **2021**, *38*, 19–26.
9. Mao, X.-l.; Shi, B.-z. Big Data Research on the Participation of Ordinary Consumers in New Energy Vehicles: On the Basis of Text Mining and Deep Learning. *J. Hainan Univ. (Humanit. Soc. Sci.)* **2022**, *40*, 201–210.
10. Zhao, F.-f.; Wang, Y.-q.; Kong, Q.-c.; Zhou, Q.-S. Application Framework and Empirical Study of Web Media Content Analysis on Public Policy—New Energy Vehicle Policy. *Inf. Sci.* **2020**, *389*, 140–146.
11. Anton, O.; Steffen, K. The impact of consumer attitudes towards energy efficiency on car choice: Survey results from Norway. *J. Clean. Prod.* **2019**, *214*, 816–822. [CrossRef]
12. Liu, Q.; Jia, M.-t.; Xia, D. Dynamic evaluation of new energy vehicle policy based on text mining of PMC knowledge framework. *J. Clean. Prod.* **2023**, *392*, 136237. [CrossRef]
13. Yew-Ngin, S.; Hussain, A. Modelling electric vehicle usage intentions: An empirical study in Malaysia. *J. Clean. Prod.* **2015**, *92*, 75–83. [CrossRef]
14. Jiang, C.-q.; Lu, X.-z.; Duan, R. Analyzing topic evolution of online product forum based on topic model. *J. Syst. Eng.* **2019**, *34*, 598–609.
15. Tao, S.-y.; Xu, X.-h.; Yu, Y.-f. Topic Mining and Evolution Trend Analysis of Educational Technology Research Based on LDA Model. *Mod. Inf. Technol.* **2023**, *7*, 176–180.
16. Yan, Y.; Peng, W.; Zhi, Z.; Li, Z.M. Research on Online Medical Service Quality Identification Based on LDA-BiLSTM Model. *Inf. Stud. Theory Appl.* **2022**, *45*, 178–183.
17. Mohammad Hamid, A.; Anindya Apriliyanti, P.; Tutut, H.; Triyani, H. The Combination of Contextualized Topic Model and MPNet for User Feedback Topic Modeling. *IEEE Access* **2023**, *11*, 130272–130286. [CrossRef]
18. Zhou, J.; Zhang, J.; Qu, R.; Yan, S. Theme Mining and Evolution Analysis of Domestic and International Blockchain Based on LDA. *J. Intell.* **2021**, *40*, 161–169.
19. Korkmaz, A.; Bulut, S.; Kosunalp, S.; Iliev, T. Investigating Remote Work Trends in Post-COVID-19: A Twitter-Based Analysis. *IEEE Access* **2024**, *12*, 196954–196968. [CrossRef]
20. Guan, P.; Wang, Y.-f.; Fu, Z. Analyzing Topic Semantic Evolution with LDA: Case Study of Lithium Ion Batteries. *Data Anal. Knowl. Discov.* **2019**, *3*, 61–72.

21. Yan, D.; Su, D.; Zhang, X.-Y. Research on Frontier Detection in Scientific Field Based on Sequential Topic Association Evolution. *Inf. Stud. Theory Appl.* **2019**, *42*, 144–150.
22. Polyzos, E.; Wang, F. Twitter and market efficiency in energy markets: Evidence using LDA clustered topic extraction. *Energy Econ.* **2022**, *114*, 106264. [CrossRef]
23. Parker, M.A.; Valdez, D.; Rao, V.K.; Eddens, K.S.; Agley, J. Results and Methodological Implications of the Digital Epidemiology of Prescription Drug References Among Twitter Users: Latent Dirichlet Allocation (LDA) Analyses. *J. Med. Internet Res.* **2023**, *25*, e48405. [CrossRef]
24. Jelodar, H.; Wang, Y.; Yuan, C.; Feng, X.; Jiang, X.; Li, Y.; Zhao, L. Latent Dirichlet allocation (LDA) and topic modeling: Models, applications, a survey. *Multimed. Tools Appl.* **2019**, *78*, 15169–15211. [CrossRef]

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.