

# A Study on Consumer Preferences for Electric Vehicle Charging and Battery Swapping Based on User-Generated Content Analysis

Yaoxuan Li\*, Ming Lu, Juntong Lu, Zhuoran Li, Ying Li#

School of Business, East China University of Science and Technology, Shanghai, China

Email: \*qdlyxyx@163.com, 1459683324@qq.com, 3252347809@qq.com, jacklyn.li@qq.com, #liying@ecust.edu.cn

**How to cite this paper:** Li, Y. X., Lu, M., Lu, J. T., Li, Z. R., & Li, Y. (2026). A Study on Consumer Preferences for Electric Vehicle Charging and Battery Swapping Based on User-Generated Content Analysis. *Journal of Service Science and Management*, 19, 203-231.

<https://doi.org/10.4236/jssm.2026.192011>

**Received:** March 28, 2026

**Accepted:** April 27, 2026

**Published:** April 30, 2026

Copyright © 2026 by author(s) and Scientific Research Publishing Inc. This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## Abstract

With the rapid expansion of the new energy vehicle (NEV) market, charging and battery swapping have emerged as the two principal energy replenishment modes, and differences in consumer preferences toward these two options have attracted increasing attention from both academia and industry. Drawing on user-generated content (UGC), this paper investigates consumers' preference characteristics for electric vehicle charging and battery swapping and identifies the key factors influencing their choices. A total of 3,009 user comments were collected from three major online platforms, namely Autohome, Dongchedi, and Zhihu. Latent Dirichlet Allocation (LDA) topic modeling was employed to extract key themes, and the optimal number of topics was determined to be 9. In addition, Snow NLP was used to examine sentiment distributions across topics, while random forest feature selection was applied to identify the most influential factors shaping consumer preferences. The results show that consumer concerns regarding charging and battery swapping are mainly concentrated in 9 dimensions: Battery Performance, Vehicle Structure, Equipment Safety, Energy Replenishment Efficiency, Service Experience, Range-related Cost, Policy Subsidies, Infrastructure Construction, and Brand Ecosystem. Sentiment analysis further indicates that consumers hold more positive attitudes toward battery swapping in terms of Energy Replenishment Efficiency, while expressing greater concern about charging with respect to Equipment Safety and Infrastructure Construction. Moreover, the random forest results reveal that Equipment Safety, Brand Ecosystem, and Infrastructure Construction are the most critical factors affecting consumer preferences. These findings provide useful insights for optimizing energy replenishment services and

\*First author.

#Corresponding author.

---

promoting the coordinated development of charging and battery swapping systems.

### Keywords

User-Generated Content Mining, Consumer Preferences, Latent Dirichlet Allocation, Snow NLP, Random Forest

---

## 1. Introduction

During the “14th Five-Year Plan” period, China’s new energy vehicle (NEV) industry has entered a stage of accelerated development. The construction of charging and battery swapping infrastructure has continued to advance, while charging technologies, standard systems, and the industrial ecosystem have been progressively improved. In this context, the energy replenishment system, as a critical support for the promotion and adoption of EVs, not only affects users’ consumption experience (Wang et al., 2021), but also directly influences the further expansion of the NEV market. At present, charging and battery swapping, as the two primary energy replenishment modes for EVs, exhibit distinct characteristics in terms of construction cost, replenishment efficiency, applicable scenarios, and user accessibility (Chen et al., 2024), and have gradually formed differentiated development paths. Moreover, the EV market demonstrates significant indirect network effects, where consumers’ vehicle purchase decisions and the construction of charging infrastructure interact with and reinforce each other (Li et al., 2017). Therefore, consumers’ choices between charging and battery swapping not only affect the market diffusion of these two replenishment modes, but also further shape the layout of infrastructure and the evolution of technological pathways.

Currently, both charging and battery swapping modes have established a certain foundation for development in the market; however, consumers’ acceptance and preference for the two modes still show divergent trends. Charging remains the dominant mode due to its broader application base and relatively mature infrastructure network, yet it still faces challenges such as longer replenishment times, queuing delays, uneven facility distribution, and pressure on the power grid. In contrast, battery swapping demonstrates advantages in efficiency and convenience but is constrained by factors such as the lack of unified battery standards, high construction and operational costs of swapping stations, and an as-yet-immature business model (Wang & Zhang, 2025). Under a market structure where both modes coexist and compete with distinct comparative advantages, consumers face multidimensional trade-offs when choosing energy replenishment methods. Consequently, differences in consumer preferences become a key factor influencing market selection and the evolution of these modes (Zhao et al., 2022). Existing studies on consumer preferences, both domestically and internationally, predominantly rely on traditional methods such as questionnaire surveys (Chen

et al., 2024). Although these methods facilitate variable control and quantitative analysis, they suffer from limitations in sample size, timeliness, and dynamic tracking, making it difficult to fully capture consumers' natural expressions and evolving attitudes in real usage contexts. Meanwhile, consumer preferences for replenishment modes are essentially shaped by the joint effects of objective constraints and subjective perceptions, which are difficult to comprehensively reveal through traditional approaches alone. In contrast, user-generated content (UGC) from social media and automotive community platforms is characterized by high authenticity, large data volume, and timely updates (Liang et al., 2024), providing a new research avenue to focus on the consumer perspective and to deeply identify their concerns, sentiment tendencies, and preference differences regarding charging and battery swapping modes.

Based on the above practical background and theoretical gaps, this study collects 3009 pieces of UGC data from three major platforms—Autohome, Dongchedi, and Zhihu. By integrating topic identification, sentiment analysis, and feature importance evaluation methods, this paper systematically explores the factors influencing consumer preferences for EV charging and battery swapping modes. The aim is to identify the core concerns, emotional attitudes, and preference differences of consumers toward the two replenishment modes from real user-generated textual data, and to further extract the key factors that influence their mode selection.

Focusing on the above research objectives, this paper addresses two core questions: First, what dimensions do consumers primarily focus on regarding charging and battery swapping modes in UGC texts, and what structural characteristics do these concerns exhibit? Second, are there significant differences in consumers' sentiment attitudes toward the two modes, and which key factors shape the formation of their preferences? Compared with existing studies, the contributions of this paper are threefold. First, by integrating comment data from Autohome, Dongchedi, and Zhihu, this study covers multiple contexts ranging from daily usage experiences to discussions on technology and policy, achieving multi-source heterogeneous data fusion and enhancing the representativeness and generalizability of the findings. Second, this paper employs BERT-based semantic encoding to improve the representation of short user-generated texts, followed by LDA topic modeling and Snow NLP sentiment analysis, thereby constructing a dual-dimensional "topic-sentiment" preference identification framework, enabling the automatic extraction of core concern topics and sentiment tendencies of consumers under charging and battery swapping modes, thereby overcoming the limitations of traditional survey-based approaches. Finally, from a consumer perspective, this paper reveals differentiated preference characteristics between charging and battery swapping modes across dimensions such as efficiency, economic performance, technological trust, and psychological perception. It defines the emotional polarity of these preferences as "consumer preferences" in this paper, providing behavior-based insights derived from real user data for government in-

infrastructure planning and enterprise service optimization, with important policy and industrial implications.

## 2. Literature Review

### 2.1. Research on the Energy Supplement Mode of Electric Vehicles

With the continuous expansion of application scenarios of electric vehicles, the focus of existing research on electric vehicle energy supplement mode has gradually changed from “static comparison of the advantages and disadvantages of charging and exchange point” to “competitive relationship between the two modes under different constraints.” From the perspective of infrastructure investors, Zhang et al. (2024) pointed out that the relative advantages of fast charging and power changing would be affected by factors such as valley power price, service cost, and potential market size. Chu et al. (2024) further introduced an evolutionary analysis framework, emphasizing that subsidy policies and consumer preferences would jointly shape the evolution direction of supplementary energy services. On this basis, the research perspective gradually shifts from the single mode judgment of “one of two choices” to the network configuration of “synergy and complementarity.” Lai and Li (2024) proposed a joint planning framework for multi-mode charging and recharging networks, arguing that charging has the advantages of cost and expansion, while recharging can improve vehicle turnover efficiency in high-frequency operation scenarios. Cao and Chu’s review (2025) also showed that the focus of future research has gradually shifted from single facility location to strategy-operation collaborative decision-making of charging, power changing, and integrated facilities. The overall evolution direction of EV energy supplement mode research is shifting from pure mode substitution discussion to collaborative placement and dynamic optimization for different scenarios.

In terms of technical performance, existing research is mainly carried out along two paths: “high-power charging and intelligent charging” and “standardization and automation of charging.” For the charging mode, Arif et al. (2021) and other systems sort out the existing technologies and future directions, and point out that the key to the evolution of charging technology is to improve the power level, improve the interface and system scheduling optimization capabilities. In contrast, Hu et al. (2020) pointed out that the power exchange technology emphasizes the standardization requirements such as realization conditions, structural interchange and electrical interface matching. In addition, Feng and Lu (2021) concluded from the perspective of station construction planning and operation that the substation has been extended from simple facility construction to scheduling optimization and business model design. Therefore, we can conclude that electricity swapping does not completely replace charging, but is more likely to exist for a long time as a complementary infrastructure in specific scenarios, and its sustainability depends on factors (Alhazmi, 2025) such as standard compatibility, automation level, safety regulation and consumer acceptance. In general, the evolution of supplementary energy technology has shown a basic trend of “contin-

ous acceleration of charging and strengthening of charging standards,” which also provides direct literature support for the identification of the theme of “technical performance” in the following article.

At the level of policy support, relevant studies generally believe that the diffusion of supplementary energy mode does not simply depend on the quality of technology, but is deeply affected by government support (Tian et al., 2024), infrastructure construction, subsidy mechanism (Zhao et al., 2024), and standard system construction. For the electricity exchange mode, Yang et al. (2023) pointed out from the perspective of social welfare that the implementation of incentives for electricity exchange service can improve overall welfare, and supply-side incentives are often more market efficient than pure demand-side incentives. Chu et al. (2024) analyzed the diffusion of battery changing service based on the three-decision evolutionary game model, and showed that the government’s subsidies to the builders were often more effective than the subsidies to the users. Tan et al. (2022), based on the Hangzhou sample, found that the public generally supported the construction of battery swapping stations and had a certain willingness to pay. The evolution direction of policy support has shifted from user subsidies to the supply side, which also lays a foundation for the identification of the theme of “policy support” later in this paper.

## 2.2. Consumer Preferences

Consumer preference is a personalized preference that reflects the degree of preference of consumers for different products and services, and is an important factor affecting market demand. With the popularity of online social media, consumers’ needs and preferences have an increasing influence on the business and sales of products. There are a lot of research and analysis methods for consumer preferences at home and abroad. For example, Li J. and Li (2024) used the conjoint analysis method to measure the relative importance of multiple product attributes and the utility of attribute levels, and obtained the product attribute preferences of wine consumers. Kiefer and Szolnoki (2024) constructed a discrete choice model of the key parameters affecting consumer perception to gain insight into preferences and willingness to pay for each product attribute.

It is crucial to find the specific influencing factors that affect consumers’ preferences for charging and switching modes. Shao et al. (2023) proposed a research method for public “charging” infrastructure demand for electric vehicles based on the random forest algorithm, and found that consumer preferences affect charging infrastructure by its existing location, capacity, and supply. In the consumer survey of “WL” company, Xu (2024) found that consumers’ use evaluation and acceptance of the power changing mode are greatly affected by the convenience of the power changing station, the cost of the power changing, and the safety and reliability of the battery.

However, in the research field of “consumer preference for electric vehicle charging and changing infrastructure,” previous studies either focus on policy re-

search (Wang & Liang, 2021), or focus on the location distribution (Han et al., 2024) of electric vehicle charging and changing devices, or focus on the analysis of consumer characteristics. For example, consumers are mainly (Li et al., 2019) high-tech talents and office workers in economically developed areas. Consumers' charging behavior is mainly concentrated during the daytime on working days (Liu et al., 2019). In the research on the relationship between charging and changing modes of electric vehicles and consumer preferences, few studies not only focus on consumers but also accurately focus on the factors that make consumers prefer charging and changing modes. In addition, in terms of research methods, taking the conjoint analysis and discrete choice model as examples, although they can measure the relative importance of multiple product attributes and the utility of attribute levels, the mining of consumers' complex preferences and influencing factors in the context of electric vehicle charging and charging infrastructure may not be comprehensive and in-depth. It is difficult to accurately focus on the key factors that make consumers prefer the charging and switching mode.

Based on this, we integrate the LDA model and BERT semantic coding, combined with existing literature, aiming to comprehensively explore the main concerns of consumers regarding the charging and switching mode, and focus the research on consumers and the factors that promote consumers to prefer the charging and switching mode.

### 2.3. Application of UGC Text Mining in Consumer Research

With the rapid development of Internet platforms and social media, UGC has shown explosive growth. UGC usually includes forum comments, social media discussions, product reviews, and other text data, which record consumers' usage experiences, attitude changes, and opinion expressions in real situations. Therefore, many scholars analyze the online review data of various platforms to study product evaluation, demand preference, and emotional attitude. Compared with traditional questionnaire survey methods, which are difficult to use for obtaining a large enough sample and cannot capture long-term changes in emotions, UGC data has the advantages of large data volume, timely information updates, more authentic expression, and low research cost, which can more objectively reflect consumers' real cognition and emotional tendencies. Therefore, it has been widely used in the field of consumer research in recent years.

To identify consumers' concerns and emotional attitudes from UGC text, most of the existing studies use text mining methods to analyze review data, among which BERT semantic coding, LDA topic recognition, and Snow NLP sentiment analysis are widely used. Among them, in terms of semantic coding, compared with the traditional bag-of-words model and static word vector, BERT can more fully capture the deep semantic information in the review text through bidirectional context modeling, and has gradually become an important tool for online review mining. In recent years, there have been a small number of studies applying the BERT model to the analysis of new energy vehicle reviews. For example,

Cheng and Zhao (2025) combined BERT with VADER rules for sentiment analysis of new energy vehicle user reviews. Wang G. (2025) used RoBERTa word embedding combined with CNN-LSTM and PLSA to identify user needs. In general, existing studies have proved that pre-trained language models have strong applicability in new energy vehicle review text processing, but related results mostly focus on single tasks such as sentiment classification or demand recognition, and apply BERT semantic coding as a text representation enhancement method prior to the LDA topic model, and the literature on charging mode preference research is still relatively few. Therefore, this paper introduces BERT semantic coding to enhance the semantic representation ability of UGC text and describe consumers' preferences for electric vehicle charging mode in a more fine-grained dimension. **In terms of topic recognition**, previous studies have shown that LDA topic modeling can effectively identify the core discussion topics in the text. Wu and Lu (2024) extracted four topics of "technical characteristics," "social activities," "driving experience," and "distribution strategy" based on the LDA topic model when studying the media communication strategy of Chinese car brands overseas home. Wang (2024) constructed the LDA topic model, clustered the recruitment skill requirements, and analyzed the word sets under each topic to obtain the topic summary words. Compared with the joint analysis, discrete choice, and other methods mentioned above, which are not deep enough to mine the influencing factors, the LDA document topic generation model can more systematically identify the core concern dimensions in consumer discussions by probabilistic modeling of text corpus. Therefore, this paper uses the LDA topic model to extract the topic of UGC comment text, to identify the main issues that consumers pay attention to in the discussion of electric vehicle charging and changing mode. **In terms of sentiment analysis**, Snow NLP is widely used in sentiment analysis of online reviews and Internet public opinion texts due to its high efficiency and stability (Xia & Shan, 2018) in Chinese corpus processing. Wang T. et al. (2023) used the Snow NLP model to analyze the emotional changes in online public opinion texts, so as to identify the evolution characteristics of Internet users' emotions at different stages. Cai et al. (2024) also used Snow NLP to analyze the sentiment of online public opinion data. These studies show that compared with traditional questionnaires, sentiment analysis technology can directly identify consumers' emotional tendencies from many unstructured text data, to more realistically reflect consumers' attitude changes in actual discussion situations. Therefore, this paper uses Snow NLP to analyze the sentiment of UGC review texts and further identify the differences in consumers' emotional attitudes when discussing the charging mode and changing mode of electric vehicles.

In recent years, UGC text mining methods have been gradually applied to the field of electric vehicles. Ashby et al. (2025) analyzed public views on electric vehicle charging based on X platform tweets, focusing on issues such as charging infrastructure, home charging accessibility, and cost disputes, reflecting consumers' perception of a single energy supplement scenario. He et al. (2025) used the

LDA model to study user focus topics and their evolution characteristics based on online reviews of electric vehicles, and effectively extracted consumer concerns. [Özkara et al. \(2025\)](#) focus on the public's overall evaluation or purchase intention of electric vehicles, and the research object is more concentrated on the overall evaluation and market acceptance process of electric vehicles. However, there are still some shortcomings in the existing research on data sources, analysis methods, and research objects. First of all, some research data sources are relatively single, and most of them focus on a single platform or single source of reviews, which makes it difficult to fully reflect the discussion content of consumers with different focuses. Secondly, existing studies tend to focus on a single text analysis method, such as only topic identification or sentiment analysis, and there is less comprehensive analysis of consumer discussion content from two dimensions of "topic-sentiment." In addition, most studies mainly focus on the overall evaluation or purchase intention analysis of new energy vehicles, and there are relatively few studies on the differences in consumer cognition of electric vehicle supplementary energy modes. Therefore, it is necessary to combine multi-platform UGC data, and comprehensively use BERT semantic coding, LDA topic recognition, and sentiment analysis methods to systematically mine consumers' concerns and emotional tendencies in the charging and changing mode of electric vehicles, and further identify the key factors that affect consumers' preferences in the supplementary energy mode.

### 3. Research Design

The research framework adopted in this study is illustrated in [Figure 1](#). First, the 3009 comments collected from the three major platforms—Autohome, Dongchedi, and Zhihu—are subjected to data preprocessing procedures, including Jieba word segmentation, stop-word removal, and BERT-based semantic vectorization. Subsequently, textual analyses such as word frequency statistics and word cloud visualization are conducted. Second, LDA is employed for topic modeling. The optimal number of topics is determined based on perplexity and coherence scores, as well as PyLDAvis visualization results. Based on this, the final key concerns of consumers regarding charging and battery swapping modes are identified, which constitute the ultimate influencing factors of consumer preferences. Third, the Snow NLP Chinese natural language processing library is used to compute preliminary sentiment scores. In addition, approximately 500 comments are randomly sampled for manual annotation. The distribution of sentiment tendencies across different topics is then visualized to assess whether consumers are satisfied with the current state of EV charging and battery swapping. Finally, feature importance analysis based on the Random Forest algorithm is conducted to evaluate consumers' satisfaction levels with various influencing factors, and corresponding future improvements and recommendations are proposed.

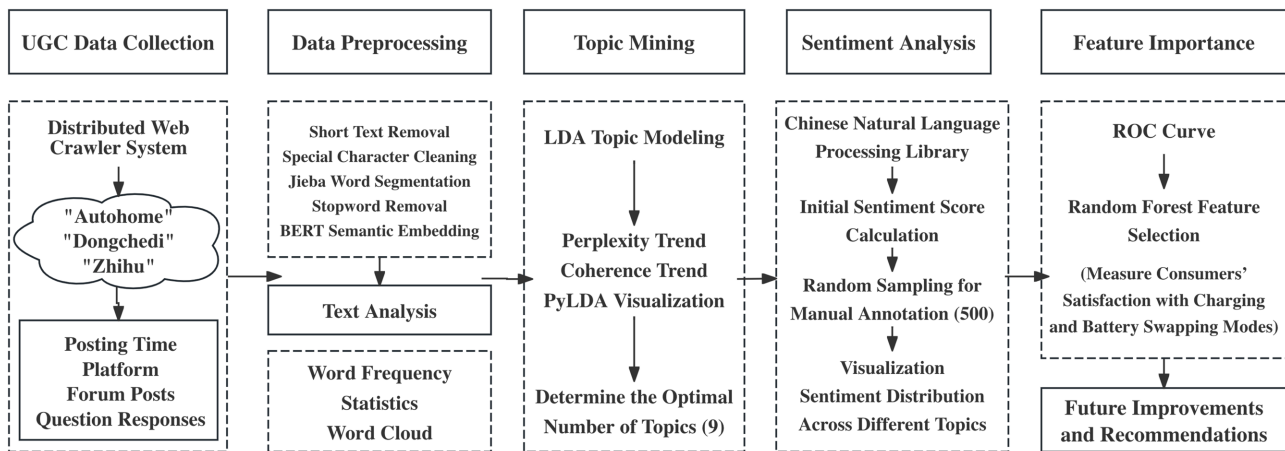


Figure 1. Research framework.

### 3.1. Data Sources and Preprocessing

#### 3.1.1. Data Sources

On many word-of-mouth and review platforms, user-generated comments serve not only as references for other potential consumers but also as a source of product market information essential for corporate development (Zhao et al., 2024). Big data has broken through the spatiotemporal limitations of traditional subjective preference research methods such as surveys and user interviews (Büschken & Allenby, 2016), offering greater objectivity, comprehensiveness, and faster processing. When combined with text mining techniques, it enables the extraction of massive amounts of online review data containing consumer opinions and behaviors, thereby allowing the extraction of user preferences and demands from the consumer perspective (Sun et al., 2020). For example, Ding et al. (2021) successfully analyzed changes in Twitter users' cognitive and emotional attitudes toward autonomous vehicles, providing new research insights for a deeper understanding of public perception.

This paper collects UGC data related to electric vehicle charging and battery swapping from three major platforms: the “Autohome” forum, the “Dongchedi” car owners’ circle, and the “Zhihu” topic community. The aim is to balance the diversity of user groups and the complementary nature of discussion content, thereby reducing sample bias caused by reliance on a single platform at the data level. This ensures that the UGC sample covers both genuine usage feedback and multidimensional perceptions encompassing technology, cost, experience, policy, and branding. The selection rationale includes:

- 1) As China’s largest automotive vertical portal, “Autohome” has a high proportion of certified car owners, covering practical usage feedback for most new energy vehicle brands, including special scenarios such as low-temperature charging in winter and long-distance energy replenishment.
- 2) Leveraging ByteDance’s algorithm-based recommendation system, “Dongchedi” boasts one of the highest user engagement rates in the industry, with over 60% of its users aged 25 - 40, representing the mainstream consumer

demographic. The site's data include GPS coordinates, enabling matching with charging and swapping facilities across multiple cities nationwide.

- 3) On “Zhihu”, some answers are contributed by professionals such as electrical engineers and battery R&D personnel, enhancing the depth of technical discussions. Users also voluntarily reference policy documents (e.g., 2024 Ultra-Fast Charging Facility Construction Guidelines).

### 3.1.2. Data Overview

The data for this study were collected using a distributed crawler system, gathering UGC data from the three platforms spanning from March 2015 to December 2025. The collection covers core scenarios of electric vehicle charging and battery swapping. To clarify the sample composition, we performed a human-assisted rule-based classification on the 3009 valid comments after cleaning, categorizing them based on their content focus. The classification rules are as follows: if a comment discussed only charging-related topics (e.g., charging piles, charging time, home charging) without mentioning battery swap, it was categorized as “Charging Only”; conversely, if it discussed only battery swap-related topics without mentioning charging, it was categorized as “Battery Swap Only”; if a comment compared or discussed both modes in relation to each other, it was categorized as “Both Involved”. The final count of comments in each category per platform is shown in **Table 1** (Zhang et al., 2024).

**Table 1.** Data overview.

Platform	Initial Crawl Volume	Valid Data Volume	Charging Only	Battery Swap Only	Both Involved
Zhihu	2012	1983 (65.9%)	632 (31.9%)	714 (36.0%)	637 (32.1%)
Dongchedi	742	715 (23.7%)	305 (42.7%)	228 (31.9%)	182 (25.5%)
Autohome	328	311 (10.4%)	183 (58.8%)	89 (28.6%)	39 (12.5%)
Total	3082	3009	1120 (37.2%)	1031 (34.3%)	858 (28.5%)

Following the approach of Paneru et al. (2025), a total of 3,000 unique data entries were collected after deduplication. Each entry includes the publication time, platform source, and main text content. Detailed information is presented in **Table 2**.

**Table 2.** Examples of collected data.

Release Time	Platform	UGC content (including forum posts, Q&A, etc.)
2025-08-12	Autohome	Electric vehicles offer better handling and comfort than fuel-powered cars. The only drawback is that charging is too slow.
2025-08-02	Autohome	How should the battery charge be managed for an electric vehicle that will be parked long-term? For an EV parked for an extended period, it is advisable to recharge it promptly to keep the battery level above 50%. Alternatively, starting the car and driving it for a few kilometers can help maintain the vehicle in good condition.

**Continued**

2023-06-16	Zhihu	As the range of electric cars continues to increase and home charging becomes more widespread, the demand for expensive third-party fast charging will only decrease. The need for quick energy replenishment...
2025-07-09	Zhihu	Charging should be the main method, with battery swapping as a supplementary option. Without a price advantage, if manufacturers' costs are not high, the cost-effectiveness remains low. To put it simply...
2025-08-02	Dongchedi	New energy vehicle benefits in the Beijing-Tianjin-Hebei region are being rolled out intensively. From direct purchase subsidies to the dense network of charging stations, this comprehensive approach not only brings real benefits to consumers but also makes green travel increasingly convenient.
2025-04-17	Dongchedi	Electric vehicles are becoming more common. Most still rely on charging, though battery swapping is also available. With so many EVs charging, leaving that aside, battery swapping is indeed fast. Personally, I feel you should use most of the battery before swapping; otherwise, you would be wasting some electricity costs.

**3.1.3. Data Preprocessing**

Data preprocessing is a core step to ensure the reliability of subsequent text analysis. For the 3009 pieces of UGC data on electric vehicle charging and battery swapping collected in this study, the following preprocessing procedures were implemented. First, exact deduplication was performed based on the unique comment ID and the combination of “user ID—posting time.” For text content that could not be deduplicated via ID, the SimHash algorithm was employed to compute text fingerprints, and entries with identical fingerprints were removed as duplicates. Subsequently, a three-tier filtering mechanism was applied to eliminate interfering information. This included: 1) removal of short texts: discarding fragmented content with fewer than 15 characters (such as “good” or “not bad”), as such texts lack sufficient informational density; 2) identification of commercial advertisements: filtering out promotional content using a predefined blacklist of keywords (including phrases like “contact for verification,” “click the link,” and “promotional offer”); and 3) cleaning of special symbols: using regular expressions to strip HTML tags, emoticons, and irregular characters, retaining only clean Chinese text.

As the review texts were collected from Chinese websites, the Jieba tool was chosen for word segmentation. To improve segmentation accuracy, this study integrated the Harbin Institute of Technology stopword list, the Baidu stopword list, and the Sichuan University Machine Intelligence Laboratory stopword library. Domain-specific keywords (e.g., “a certain brand,” “reportedly”) (Wang et al., 2025) from the EV field were added. Standard term entries were incorporated from the Electric Vehicle Engineering Terminology (Tsinghua University 2024 Edition), and unregistered words appearing more than 50 times in the original corpus (e.g., “battery swap compatibility,” “range anxiety”) were extracted. Finally, technical neologisms (e.g., “ultra-fast charging C-rate,” “State of Health”) were supplemented manually. Ultimately, a professional dictionary containing

2512 word entries was constructed. The effectiveness of the word segmentation is shown in **Table 3**.

**Table 3.** Examples of word segmentation results.

Online Review Data	Word Segmentation Results
Some battery swap stations have already added a small lounge nearby, which will serve as a pleasant rest and social space for NIO owners and EV enthusiasts during their journeys.	battery swap station/lounge/NIO/owner/electric vehicle/enthusiast/journey/pleasant/rest/social/space
However, NIO's battery-swappable model and battery leasing design cleverly circumvent the issue of battery lifespan.	NIO/battery swap mode/battery/leasing/design/cleverly/battery/lifespan
Achieving interoperability of charging networks helps electric vehicle owners easily find charging facilities during long-distance travel.	charging network/interoperability/helps/electric vehicle/owner/long-distance/travel/easily/find/charging/facility
The quality of charging piles must be excellent, as it directly relates to the safety and performance of electric vehicle batteries.	charging pile/quality/excellent/quality/relates/electric vehicle/battery/safety/performance

## 3.2. Research Methods

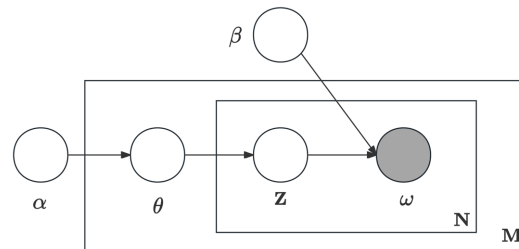
### 3.2.1. LDA Topic Clustering

To identify consumers' focal concerns and preference characteristics in the contexts of electric vehicle charging and battery swapping from large-scale UGC, this study employs the LDA model to perform topic clustering on the textual corpus. LDA is a widely used unsupervised text analysis method that can uncover a latent "semantic space" with underlying structure from large volumes of unlabeled UGC comments (Blei et al., 2003), thereby extracting potential themes and shared concerns of consumers regarding energy replenishment from unstructured text.

In line with the research objectives of this study, the LDA topic model is primarily used to address the question: "Which aspects of charging and battery swapping modes do consumers focus on in UGC?" Specifically, LDA models the data through a three-level hierarchical structure (as shown in **Figure 2**). First, a topic distribution is sampled for each document, denoted as  $\theta_d \sim Dir(\alpha)$ , representing the weight of consumers' attention across different energy replenishment topics (e.g., waiting time, cost, driving range). Then, for each word, a topic assignment is sampled as  $z_{dn} \sim Multinomial(\theta_d)$ , indicating the latent preference topic to which each comment belongs. Finally, a word is generated from the corresponding topic-specific word distribution  $\Phi_k \sim Dir(\beta)$ , i.e.,  $\omega_{dn} \sim Multinomial(\Phi_{z_{dn}})$ , simulating the probabilistic mechanism by which users choose vocabulary when describing their energy replenishment experiences.

It should be noted that, in practical modeling, traditional LDA primarily relies on the bag-of-words representation, which has limited ability to capture semantic structures. This limitation is particularly evident in short UGC texts such as Zhihu responses, which are typically brief, colloquial, and sparsely distributed in semantic structure. As a result, topics are easily affected by high-frequency words, leading to suboptimal performance (Zhao et al., 2011). To enhance semantic repre-

sensation in short texts, this study applies the BERT model for semantic encoding as a preprocessing step. BERT generates contextualized word or sentence vectors from the UGC corpus, which help disambiguate polysemous or colloquial expressions commonly found in short user comments. These vectors are primarily used for exploratory analysis and to assist manual interpretation of ambiguous terms before topic modeling. The actual topic extraction is then performed by the standard LDA model, which operates on term frequency matrices derived from the pre-processed and segmented corpus.



**Figure 2.** LDA topic model.

When applying LDA for topic mining, two key issues must be addressed: 1) how to determine the optimal number of topics  $K$ , and 2) whether the extracted topics exhibit sufficient interpretability. If the number of topics is too small, discussions from different semantic domains may be merged, resulting in insufficient topic differentiation. Conversely, if the number is too large, it may lead to excessive topic fragmentation, reducing interpretability. To improve topic interpretability, this study follows mainstream practices by jointly considering Perplexity, Topic Coherence, and PyLDAvis visualization results to determine the optimal number of topics (Chen & Li, 2019).

Among these, Perplexity is mainly used to measure the model's goodness of fit to the text. A lower perplexity value indicates a stronger ability of the model to explain the corpus, meaning that the learned topic structure is more consistent with the true data distribution. Its calculation formula is:

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

where  $p(\omega_d)$  is the probability of generating document  $d$  under the model,  $N_d$  is the number of words in document  $d$ , and  $M$  is the total number of documents.

Topic coherence is used to measure the semantic relatedness among keywords within the same topic and serves as an indicator of topic interpretability. A higher coherence value indicates that high-frequency words within a topic are more likely to co-occur in the corpus, implying stronger internal consistency and facilitating subsequent topic naming and interpretation. Its calculation formula is (Röder et al., 2015):

$$C_V(T) = \frac{1}{|T|(|T|-1)} \sum_{i \neq j} NPMI(\omega_i, \omega_j) \quad (2)$$

where  $\omega_i$  and  $\omega_j$  are keywords within the same topic, and NPMI (Normalized Pointwise Mutual Information) represents the normalized pointwise mutual information between word pairs.

### 3.2.2. Snow NLP Sentiment Analysis

To further investigate the emotional attitude tendency of consumers towards the two supplementary energy modes of electric vehicle charging and changing, this paper uses the Snow NLP module of Python to calculate the preliminary emotional scores of 3009 reviews, defining the resulting sentiment trends as the “consumer preferences” in this study. Snow NLP is a natural language analysis tool (Wang et al., 2022) specially designed for Chinese text processing. It mainly relies on a custom sentiment dictionary to identify the sentiment orientation of the text. The basic principle is to first pass the crawlable UGC text to the module, go through the words in each review, and find whether there is a matching item in the sentiment dictionary. If it belongs to a custom sentiment dictionary, its polarity (positive, negative, neutral) is added to the sentiment score of the review. Finally, the sentiment score is divided by the number of sentiment words to obtain the revised sentiment score (Chen et al., 2018), and the sentiment score is usually between 0 and 1 (the closer to 1, the more positive the text is). Therefore, compared with manual interpretation one by one, Snow NLP can more efficiently quantify the sentiment of large-scale review texts, especially the UGC short texts with a high proportion of Chinese corpus, colloquial expression, and strong emotional color. At the same time, in order to further improve the accuracy of the sentiment classification model in this paper, we randomly select about 500 comments from the total sample for manual annotation, reconstruct the corresponding sentiment dictionary, and classify the comments with a sentiment score higher than 0.7 as “positive”, those with a sentiment score lower than 0.3 as “negative”, and those with a sentiment score between 0.3 and 0.7 as “neutral”. By comparing the accuracy rates without using a custom dictionary, it was found that the improvement effect of using a custom dictionary was 0.0042. The accuracy was calculated for four groups of thresholds: 0.3/0.7, 0.4/0.6, 0.45/0.55, 0.35/0.65, among which the accuracy of 0.4/0.6 was the best at 0.4312.

### 3.2.3. Random Forest Feature Selection

Based on topic identification and sentiment analysis, this paper further uses the random forest method to evaluate the feature importance of key factors affecting consumer preferences, and further answers the question, “Which factors are more important for the formation of consumer preferences?” The random forest is an ensemble learning method, which is composed of multiple independent decision trees, and each tree outputs a classification result. After all the decision trees give their results, the random forest will select the feature with the most classification results as the final classification result. Therefore, the random forest model can

evaluate the importance of features through the judgment of decision trees. The “Autohome”, “Understand Car Emperor”, and “Zhihu” platforms selected in this paper usually mark the “selected” or “recommended” labels on UGC comment texts. The random forest model comprehensively evaluates these labels, which can make feature selection in a targeted way and rank the importance of the relevant influencing factors extracted in the topic identification stage. This can then identify the core issues that consumers pay more attention to in the selection of supplementary energy mode, which is of great significance for automobile enterprises to understand consumer needs and market research.

In the specific implementation process, based on the results of previous text mining, this paper takes relevant topics and feature values as input, and uses the random forest model to calculate the importance of different features to the target variable. The input features ( $X$ ) are generated by TF-IDF vectorization of preprocessed user-generated comment texts. The dependent (target) variable ( $y$ ) is defined as a multi-class composite label consisting of two energy replenishment modes and nine core discussion topics, forming 18 mutually exclusive classification categories. In random forest, the commonly used feature selection methods include information gain, gain rate, Gini coefficient, and chi-square test. The dataset is split into training and test sets, and the Random Forest model is constructed and trained with the Gini impurity criterion as the node splitting rule, which is the core metric for measuring feature purity. In this paper, the GINI coefficient (GINI) is used to select the criterion with the highest purity of each child node for feature selection (Zhao et al., 2024), and the importance of each feature is ranked. For a general decision tree, when the sample belongs to the class  $k$ , the probability distribution of the Gini index is as follows:

$$Gini(p) = \sum_{i=1}^i p_i (1 - p_i) = 1 - \sum_{i=1}^i p_i^2 \quad (3)$$

In general, the more obvious the node purity improvement brought by a feature in the process of decision tree splitting, the greater the contribution of the feature to the classification result and the higher the importance of the feature. Based on this idea, this paper ranks the importance of different topic-related features to identify the influencing factors that consumers pay more attention to in the process of forming their charging and swapping mode preferences.

## 4. Analysis of Preference Factors for EV Charging and Battery Swapping Based on the LDA Topic Model

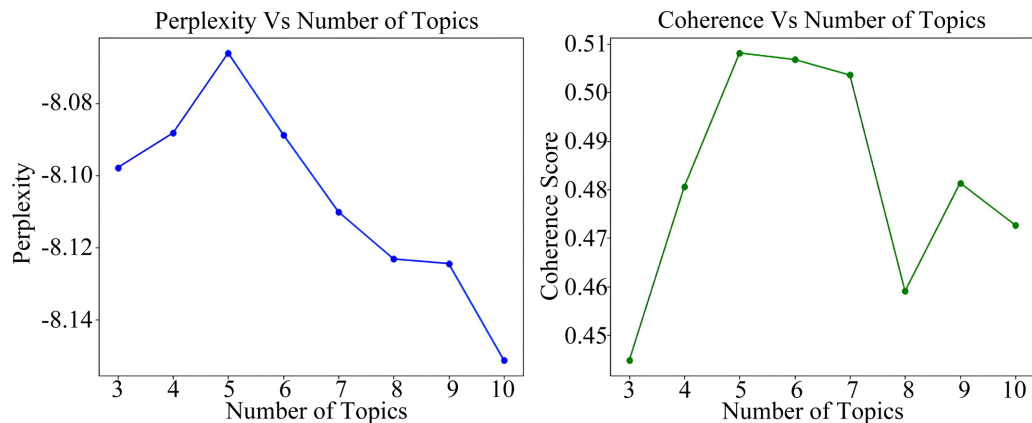
### 4.1. Text Analysis

To distill the core dimensions of user concern regarding EV energy replenishment modes, this study performed word frequency statistics on the 289,913 segmented words. From the perspective of replenishment efficiency, the “queueing” issue (354 mentions) emerged as the primary pain point for the charging mode, whereas the battery swap mode gained favor for its “speed” (477 mentions). From an economic dimension, the battery swap mode attracted short-term user interest



“cost,” “electricity price,” and “subsidy” among the high-frequency words indicates that consumer perception of the pricing system and reliance on policy incentives remain key dimensions influencing their preferences. Notably, the frequent occurrence of technical attribute terms such as “800 V,” “standardization,” and “compatibility” suggests that consumer discussions extend beyond user experience to encompass technical and industrial coordination, reflecting their concern for the future energy replenishment ecosystem. This reveals the overall characteristics of EV energy replenishment modes. Subsequent LDA topic modeling will further refine the semantic clusters corresponding to these high-frequency words to identify the latent preference structures and influencing factors among consumers regarding charging and battery swapping modes, thereby achieving complementarity and deepening between word frequency features and topic semantics.

#### 4.2. Determination of the Optimal Number of Topics



**Figure 4.** Trends of perplexity and coherence.

Before conducting topic modeling, determining the optimal number of topics is crucial to ensure model quality. To this end, this study adopts two mainstream evaluation metrics—Perplexity and Topic Coherence—to assess model performance. Specifically, the number of topics is set within the range of [3] [10], with 8246 feature words selected and 500 iterations performed. Based on this, the trends of perplexity and coherence are plotted. As shown in **Figure 4**, when the number of topics  $K = 9$ , the model achieves a local optimum in coherence (0.48) while maintaining a relatively low perplexity ( $-8.135$ ). In contrast, when  $K = 8$ , the coherence score drops to 0.46, and when  $K = 10$ , perplexity increases noticeably without significant gain in coherence. Moreover, the PyLDAvis inter-topic distance map (see **Figure 5**) shows that with  $K = 9$ , the circles representing each topic are well-separated with minimal overlap, indicating that the nine topics are semantically distinct and interpretable. When  $K$  is set to 8, some topics are forced to merge, leading to ambiguous keyword distributions; when  $K$  is increased to 10, several topics become fragmented, producing repetitive high-frequency words.

Therefore, considering both quantitative metrics and qualitative interpretability,  $K = 9$  is selected as the optimal number of topics.

In addition, this study employs the PyLDAvis visualization tool in Python to generate an inter-topic distance map (see Figure 5) (Sievert & Shirley, 2014), which intuitively presents the distribution of words within each topic and the relevance of each word to the corresponding topic (Nastiti et al., 2021). The results indicate that when  $K = 9$ , there is relatively little overlap among topics and the classification performance is more satisfactory, further supporting the selection of the optimal number of topics.

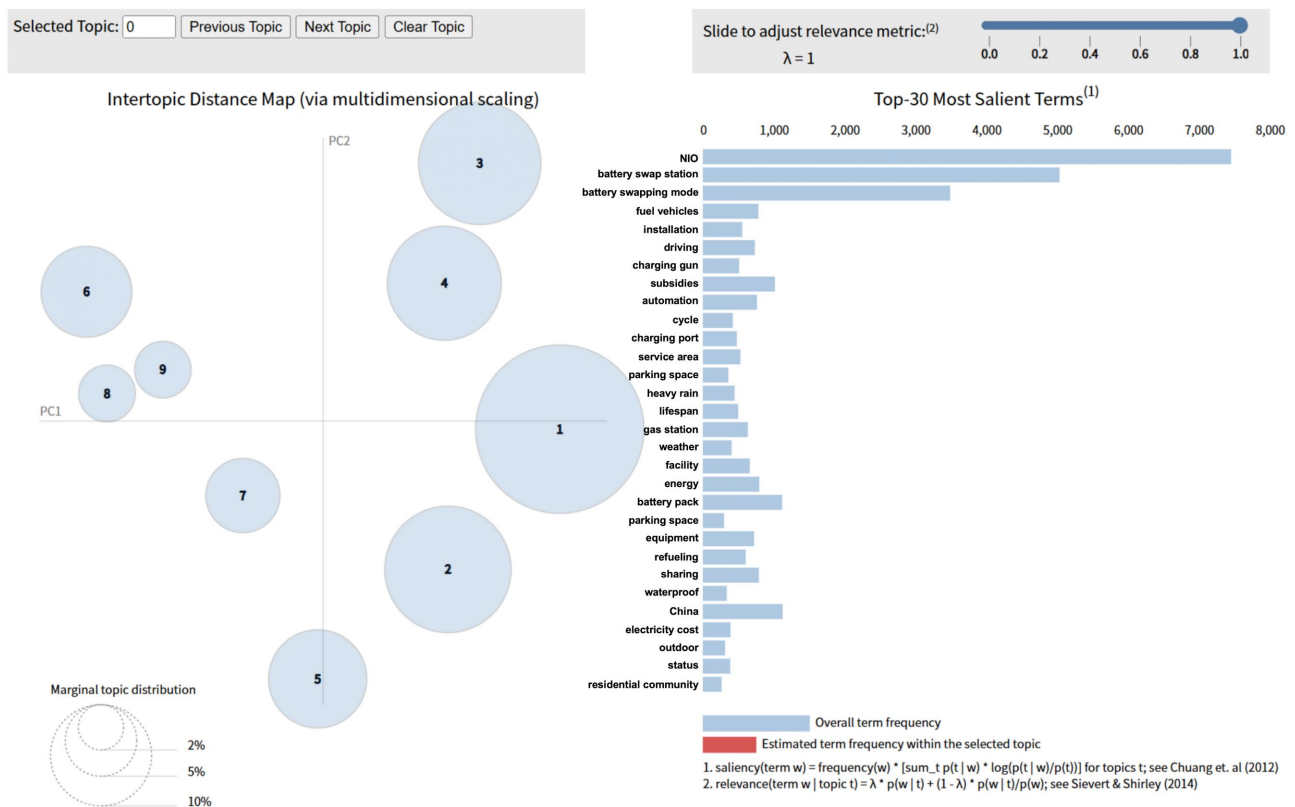


Figure 5. PyLDAvis inter-topic distance map.

### 4.3. Output of LDA Topic Modeling Results

Based on the topic mining results, each of the 9 topics was named by following a two-step procedure. First, the top 15 weighted keywords per topic were extracted. Second, two co-authors independently read a random sample of 30 representative UGC texts that had the highest topic probability for each candidate topic. The topic label was then assigned by synthesizing the semantic field of the keywords and the recurring themes in the representative texts. For example, in Topic T1.1, the keywords “cycle,” “lifespan,” “discharge,” “protection,” “status,” “damage,” and “lithium-ion battery” all point to the technical condition and aging of the battery, so it is labeled Battery Performance. In Topic T4.2, keywords such as “installation,” “parking space,” “residential community,” “distribution network,”

and “property rights” frequently appear together in user complaints about the difficulty of installing private charging piles, hence the label Infrastructure Construction. Inter-coder agreement on the nine labels exceeded 85% after discussion, confirming the reliability of the naming. **Table 4** presents the final nine secondary topics, their proportions in the corpus, and the associated keywords with weights.

**Table 4.** Text topic distribution and keyword weights.

Primary Topic Name	Secondary Topic Name (Proportion)	Keywords (Weights)
<b>T1</b> Technical Performance	<b>T1.1</b> Battery Performance (10.56%)	cycle (0.028), lifespan (0.018), discharge (0.015), protection (0.015), status (0.014), damage (0.012), lithium-ion battery (0.009)
	<b>T1.2</b> Vehicle Structure (16.96%)	chassis (0.009), body (0.008), NIO (0.007), size (0.007), BYD (0.007), structure (0.006), Tesla (0.006), high-end (0.006)
	<b>T1.3</b> Equipment Safety (12.56%)	charging gun (0.020), heavy rain (0.017), weather (0.016), waterproof (0.013), equipment (0.013), rainwater (0.011), current (0.009), voltage (0.009)
	<b>T1.4</b> Energy Replenishment Efficiency (9.07%)	battery swapping mode (0.016), power battery (0.009), fast (0.007), CATL (0.006), driving range (0.006), swapping technology (0.005), charging time (0.005)
<b>T2</b> Service Experience	<b>T2.1</b> Service Experience (8.17%)	driving (0.025), automation (0.021), service area (0.020), configuration (0.013), fast charging (0.010), queuing (0.010), car owners (0.009), used cars (0.009)
<b>T3</b> Range-related Cost	<b>T3.1</b> Range-Related Cost (18.81%)	fuel vehicles (0.026), refueling (0.014), electricity cost (0.012), inexpensive (0.010), long-distance travel (0.009), range extender (0.007), mileage (0.006), service fee (0.006)
<b>T4</b> Policy Support	<b>T4.1</b> Policy Subsidies (10.22%)	subsidies (0.016), infrastructure (0.013), China (0.012), government (0.012), electricity (0.008), power consumption (0.008), energy storage (0.008), technology (0.008)
	<b>T4.2</b> Infrastructure Construction (2.76%)	installation (0.042), parking space (0.030), parking lot (0.024), residential community (0.018), distribution network (0.018), malfunction (0.012), property rights (0.011), setup (0.009)
<b>T5</b> Brand Ecosystem	<b>T5.1</b> Brand Ecosystem (10.91%)	NIO (0.080), battery swapping station (0.053), Tesla (0.011), automakers (0.010), CATL (0.006), profitability (0.006), assets (0.005), cooperation (0.005)

**Topic T1 (Technical Performance)** mainly includes four dimensions: battery performance, vehicle structure, equipment safety, and energy replenishment efficiency, reflecting consumers’ continuous concern about battery durability, vehicle structural design, equipment safety, as well as the efficiency and convenience of energy replenishment.

**Topic T2 (Service Experience)** captures consumers’ subjective perceptions and demands regarding supporting services and driving experience, focusing on practical issues such as vehicle intelligence level, personalization, and completeness of service facilities, facility distribution density, and queuing time. Notably, this topic is often accompanied by emotional tendencies, as consumers tend to evaluate existing services based on their personal experiences. This phenomenon reflects the ongoing transition of the EV industry from a “product-oriented” to a “user-oriented” paradigm.

**Topic T3 (Range-related Cost)** accounts for the highest proportion of user discussions (18.81%), indicating that consumers place particular emphasis on the economic aspects of electricity versus fuel costs, as well as practical concerns such as long-distance driving range when choosing energy replenishment methods.

**Topic T4 (Policy Support)** encompasses both macro-level policy guidance and micro-level infrastructure construction, reflecting that subsidy policies and practical barriers—such as the installation of private charging piles—have a direct guiding effect on consumer preferences.

**Topic T5 (Brand Ecosystem)** mainly involves representative automakers such as NIO and Tesla, as well as industry competition patterns, technological pathways, and business models. Consumers' attention to enterprises often demonstrates strong forward-looking awareness and industry insight, forming a community culture based on brand identity. This is also one of the distinctive characteristics that differentiate the EV sector from the traditional automotive industry.

Furthermore, the keyword weight distributions within each topic are relatively balanced, with high semantic consistency. For example, in the “battery performance” topic, terms such as “cycle”, “lifespan”, and “discharge” exhibit strong technical relevance. This indicates that the LDA model effectively captures the latent topic structure in the text, providing a solid foundation for subsequent sentiment analysis to compare attitude differences across topics and for the application of Random Forest to identify key influencing factors.

## 5. Snow NLP Sentiment Analysis

### 5.1. Manual Annotation

As mentioned above, in order to accurately study the emotional attitude tendency of consumers towards the two supplementary modes, the Snow NLP library is used to calculate the preliminary emotional scores of all 3009 reviews. We also randomly selected about 500 reviews from the total sample and had two annotators label them (Cohen's Kappa = 0.89), as shown in **Table 5** below:

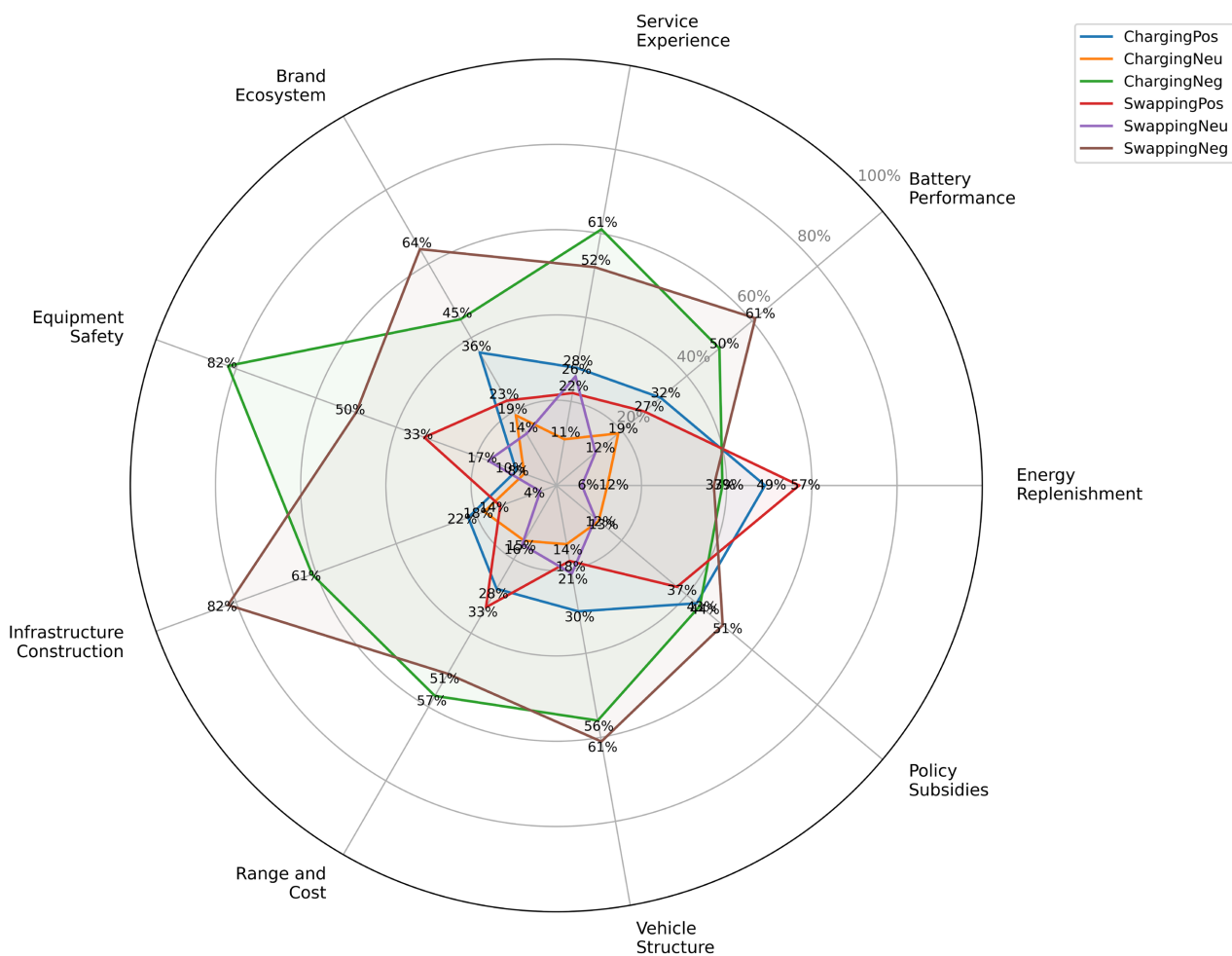
**Table 5.** Examples of representative reviews for each topic (two decimal places are retained).

Supply mode	Topic	Comment content	Sentiment score	Type	Opt. Sentiment type	Opt. Sentiment score
Charging	Battery performance	It should be said that when BYD made huge profits in the past few years, it was even more profitable than Tesla, and it still did not build a charging station. How much do you think the possibility will be...	0.71	Pos	Neg	0.43
Charging	Equipment safety	Charging stations are large power consumers, so several days have been limited to power, many owners run several charging stations can not be charged... Car companies are developing better battery cooling technology... The government is also looking at ways...	0	Neg	Neu	0.46

Continued

Swapping	Range and cost	At present, the only pain point of electric vehicles is that they can't go long distances. If the point of electric change on long distances is the same as that of gas stations, then electric cars can completely replace oil cars, and the cost may be lower than that of gas stations.	0.21	Neg	Pos	0.40
Swapping	Policy subsidies	A 50,000 yuan subsidy for replacing the old with the new is superimposed on incentives for car companies, and the actual purchase price of entry-level models can be reduced to about 30,000 yuan.	0.06	Neg	Pos	0.51

### 5.2. Distribution of Affective Tendencies



Radar Chart of the emotional distribution of Charging and Swapping entities

Figure 6. Distribution of sentiment tendencies for different topics.

To accurately capture the emotional tendency of different topics, this paper draws

a comparison map of the emotional distribution of different topics based on the above manual annotation (see **Figure 6**). Overall, users had a more positive attitude towards the “energy supplement efficiency” of the power exchange mode, while they showed deeper concerns about the “equipment safety and facility construction” of the charging mode.

The results show that equipment safety is the “hardest hit area” of negative emotions in the charging mode, which accounts for 82% of the negative emotions, much higher than that in the changing mode (50%), reflecting that users have more concerns about the reliability and installation conditions of outdoor charging equipment. In contrast, the power changing mode had a higher proportion of positive emotions on the topic of energy recharge efficiency (57% vs 49%), indicating that its feature of “fast energy recharge” was more easily recognized by users than the charging mode. However, its negative emotions on the topics of facility construction and brand ecology are stronger (82% vs 61%; 64% vs 45%), indicating that consumers still have doubts about its “installation” difficulty, “parking space” occupation, and the sustainability of its business model. At the same time, under the two themes of battery life cost and service experience, the proportion of negative emotions in the two modes is relatively high, indicating that economy and actual use experience are still the common pain points of consumers.

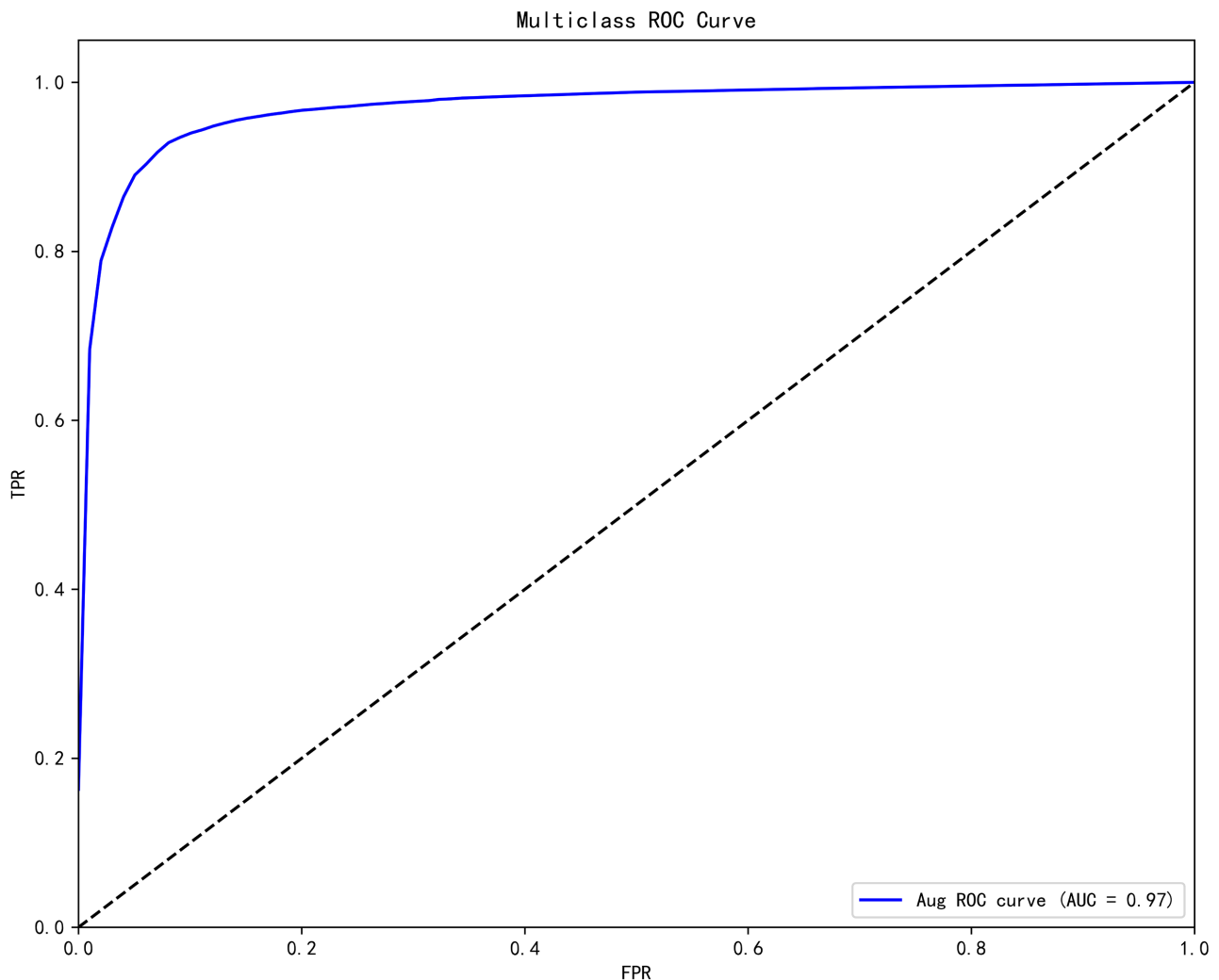
In summary, the emotional distribution map clearly reveals the structural differences in user preferences: the power exchange mode is good at “efficiency”, but is trapped in the trust crisis of “brand ecology”. Although the charging mode has high popularity, it has obvious shortcomings in “equipment safety” and “facility construction”. “battery life cost” and “service experience” are industry problems faced by both.

## 6. Random Forest Feature Selection

To answer the question “which topics can best represent consumers’ concerns about the two supplementary energy modes”, this paper uses the random forest model for feature selection, and ranks the importance of each type of feature. Firstly, we fully trained the random forest classifier in the Python code. The complete workflow includes text vectorization via TF-IDF, train-test split, model fitting, multi-class ROC curve evaluation, and feature importance calculation. When performing feature selection, the AUC (Area Under Curve) value is 0.97 (see **Figure 7**), indicating that the model is an effective classification model, and the user’s discussion content for charging and electricity changing has a large degree of discrimination.

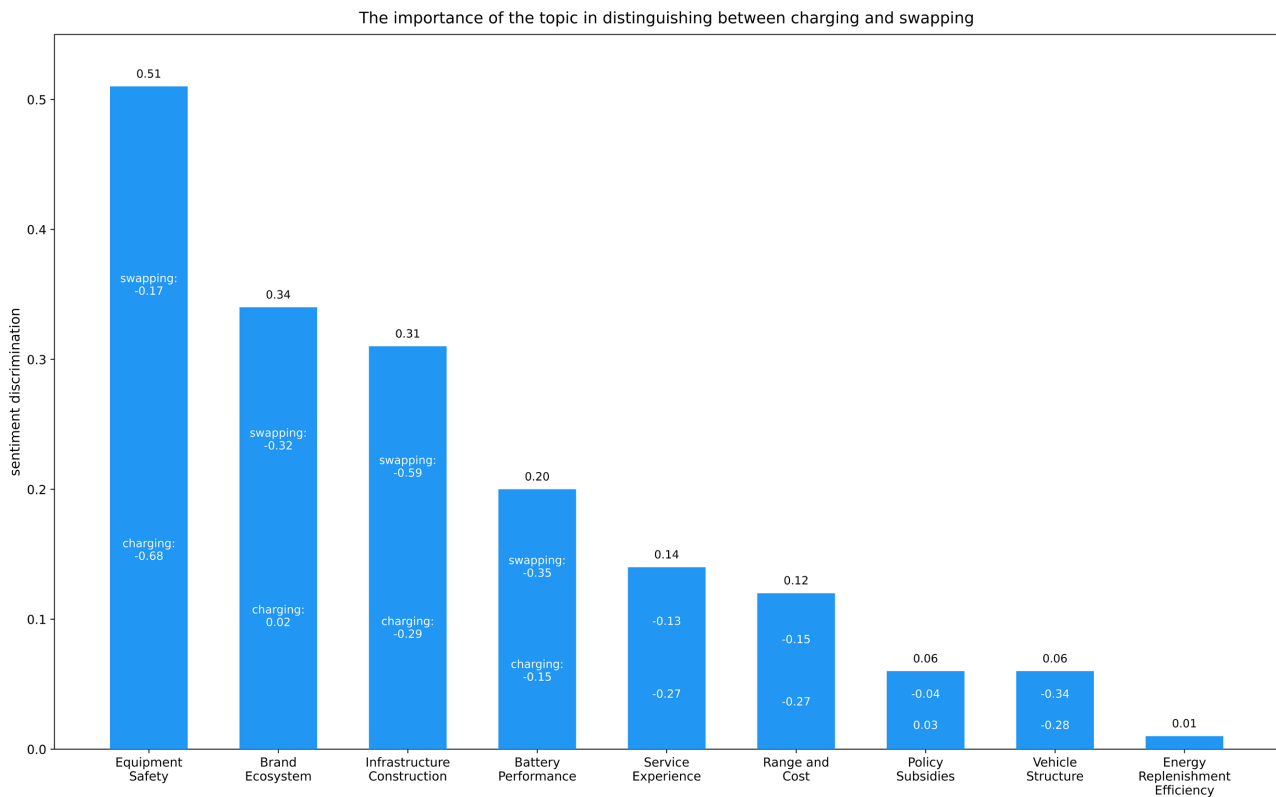
Secondly, the importance of each type of feature was calculated (see **Figure 8**). The higher the value of each bar, the greater the influence of this element on the classification result, that is, the higher the probability that this attribute feature is a value attribute that consumers consider more important.

The analysis found that the importance of equipment safety was far greater than that of other feature attributes, followed by brand ecology and facility construction, indicating that equipment safety, brand ecology, and facility construction were highly important in users' minds. The importance of energy supplement efficiency is low, indicating that this feature attribute is not an important attribute that affects consumer preferences.



**Figure 7.** Distribution of sentiment tendencies for different topics

In addition, the results of this part and sentiment analysis mutually confirmed the results: for example, the negative sentiment of users for the charging mode (82%) was much higher than that for the charging mode (50%). In the theme of brand ecology, the situation is the opposite. On the theme of energy efficiency, the difference between positive and negative emotions of charging mode is not large. This significant difference in sentiment orientation directly leads to highly discriminative text features of the corresponding topic between the two modes, which are captured by the random forest model.



**Figure 8.** Topic discrimination ranking based on the feature importance of the random forest.

## 7. Conclusions and Recommendations

### 7.1. Conclusions

Based on user review data from the “Autohome,” “Dongchedi,” and “Zhihu” platforms, this paper conducts an in-depth analysis of “consumer preferences regarding the charging and battery swap modes for electric vehicles” through LDA topic modeling, Snow NLP sentiment analysis, and random forest feature importance assessment.

**Based on LDA topic modeling**, the following conclusions can be drawn: 1) Calculations of perplexity, coherence, and PyLDAvis visualization indicated that the optimal number of topics is nine, namely Battery Performance, Vehicle Structure, Equipment Safety, Energy Replenishment Efficiency, Service Experience, Range-related Cost, Policy Subsidies, Infrastructure Construction, and Brand Ecosystem. 2) Based on the topic modeling results, these nine topics can be categorized into five groups: Technical Performance, Service Experience, Range-related Cost, Policy Support, and Brand Ecosystem. Specifically, Battery Performance, Vehicle Structure, Equipment Safety, and Energy Replenishment Efficiency are classified under Technical Performance, while Policy Subsidies and Infrastructure Construction are grouped under Policy Support.

**Based on sentiment analysis**, the following conclusions are drawn: 1) Overall, users hold a more positive attitude towards the “Energy Replenishment Effi-

ciency” of the battery swap mode, while expressing deeper concerns about the “Equipment Safety” and “Infrastructure Development” of the charging mode. 2) Energy Replenishment Efficiency is the core reason for gaining consumer favor, directly reflecting the importance consumers place on this factor for both modes. 3) The negative sentiment proportion is relatively high for both modes within the Range and Cost topic (Charging: 57%, Battery Swap: 51%), indicating that cost-effectiveness remains a prevalent core pain point for consumers. Neither mode fully satisfies user expectations for being “inexpensive” and suitable for “long-distance travel.” 4) The Equipment Safety topic highlights consumers’ deep-seated concerns about the reliability of outdoor charging piles in harsh environments. The Infrastructure Development topic shows a similar trend, where practical obstacles like installation difficulties, occupied parking spots, and property management restrictions are major issues plaguing charging users. 5) The high negative sentiment associated with keywords like “carmakers,” “profitability,” and “cooperation” under the Brand Ecosystem topic reveals widespread user skepticism regarding the sustainability of the battery swap business model, the standardization of batteries, and the progress of inter-brand collaboration.

**Based on feature selection**, the calculated importance scores are: Equipment Safety topic (0.51), Brand Ecosystem topic (0.34), and Infrastructure Development topic (0.31). This indicates that consumers place greater emphasis on attributes like “Equipment Safety, Brand Ecosystem, and Infrastructure Development.” Consequently, electric vehicle companies should increase investment and technological support in these areas.

## 7.2. Managerial Implications

- 1) **Strengthen Infrastructure Development and Equipment Safety Assurance.** To address practical issues such as difficult charging pile installation and occupied parking spots, automakers should collaborate with the government to actively promote the optimization of charging facility layout, solve installation difficulties and parking space shortages. Regarding problems associated with heavy rain, waterproofing, etc., automakers should enhance the safety design of charging piles and battery swap station equipment, improve reliability under extreme weather conditions, and increase investment in safety protection technologies.
- 2) **Optimize energy replenishment efficiency and reduce range-related costs.** Automakers should increase R&D and improvements in battery swap technology, enhance the coverage of battery swap stations, and reduce waiting times during the swapping process. Simultaneously, they should continue to promote battery technology innovation to extend driving range, and optimize electricity pricing structures to alleviate the economic burden on users during EV usage.
- 3) **The government should continue to intensify policy support for the EV industry**, particularly by providing more subsidies and tax incentives for in-

frastructure construction, and encouraging innovation in technological R&D and infrastructure development. Furthermore, the government can promote the standardization process for batteries, addressing the issue of non-uniform battery specifications among different manufacturers. This would lower consumer trust barriers towards the battery swap mode while ensuring user safety across various usage environments.

- 4) **Establish a comprehensive consumer feedback mechanism to encourage interaction between enterprises and consumers.** Through regular user surveys and data analysis, companies can promptly understand consumer preferences and pain points regarding charging and swap modes, continuously optimizing products and services. Additionally, enterprises should maintain close communication with consumers via social platforms and online communities, leveraging big data analysis to better meet consumer needs, thereby enhancing brand loyalty and market competitiveness.

Simultaneously, this study has certain limitations regarding data sources. Future research could collect data from more social media platforms to obtain a more comprehensive view of public opinion and sentiment feedback. It could also combine quantitative and qualitative methods to delve deeper into the underlying reasons and influencing mechanisms behind consumer preferences for EV charging and battery swap modes, providing more targeted suggestions and strategies for the future development direction of electric vehicle enterprises.

## Funding

Supported by the *Undergraduate Training Program on Innovation and Entrepreneurship grant* 202510251095.

## Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

## References

- Alhazmi, Y. A. (2025). Electric Vehicle Battery Swap Stations: An Overview and Critical Review. *Journal of Umm Al-Qura University for Engineering and Architecture*. <https://doi.org/10.1007/s43995-025-00215-z>
- Arif, S. M., Lie, T. T., Seet, B. C., Ayyadi, S., & Jensen, K. (2021). Review of Electric Vehicle Technologies, Charging Methods, Standards and Optimization Techniques. *Electronics*, 10, Article No. 1910. <https://doi.org/10.3390/electronics10161910>
- Ashby, C., Weir, D., & Fussey, P. (2025). Understanding Public Views on Electric Vehicle Charging: A Thematic Analysis. *Transportation Research Interdisciplinary Perspectives*, 29, Article ID: 101325. <https://doi.org/10.1016/j.trip.2025.101325>
- Blei, D. M., Andrew Y., N., & Michael I., J. (2003). Latent Dirichlet Allocation. *Journal of Machine Learning Research*, 3, 993-1022.
- Büschken, J., & Allenby, G. M. (2016). Sentence-Based Text Analysis for Customer Reviews. *Marketing Science*, 35, 953-975. <https://doi.org/10.1287/mksc.2016.0993>
- Cai, Z., Han, Y., Zhang, J., Jiang, N., & Feng, Y. (2024). MicroBlog Network Public Opinion

- Analysis System Based on SnowNLP. *Science Technology and Engineering*, 24, 5457-5464. (In Chinese)
- Cao, B., & Chu, C. (2025). A Review: Location of Charging and/or Battery Swapping Stations for Electrical Vehicles. *International Journal of Production Research*, 63, 10633-10676. <https://doi.org/10.1080/00207543.2025.2554949>
- Chen, J., & Li, Y. (2019). Social Media Fatigue Research Based on LDA Topic Model—Take WeChat as An Example. *Information Science*, 37, 78-86. (In Chinese)
- Chen, C., Chen, J., & Shi, C. (2018). Research on Credit Evaluation Model of Online Store Based on SnowNLP. *E3S Web of Conferences*, 53, Article ID: 03039. <https://doi.org/10.1051/e3sconf/20185303039>
- Chen, Z., Xia, S., Shao, Z., Gu, Z., Wu, Z., Wu, L. et al. (2024). Charging or Swapping? A Study on the Private Consumers' Acceptance of the Electricity Replenishment Mode of New Energy Vehicles. *Transportation Research Part F: Traffic Psychology and Behaviour*, 107, 903-923. <https://doi.org/10.1016/j.trf.2024.10.015>
- Cheng, C., & Zhao, J. (2025). Sentiment Analysis of New Energy Vehicle Users Comments Based on BERT and VADER Rules. *Intelligent Computer and Applications*, 15, 1-8. (In Chinese)
- Chu, Y., Yu, H., & Li, Y. (2024). How Do Supply- and Demand-Side Dynamics and Subsidies Affect the Prospects for Electric Vehicle Battery Swapping Services? Evidence from an Evolutionary Analysis. *Energy*, 308, Article ID: 132922. <https://doi.org/10.1016/j.energy.2024.132922>
- Ding, Y., Korolov, R., (AI) Wallace, W., & Wang, X. (. (2021). How Are Sentiments on Autonomous Vehicles Influenced? An Analysis Using Twitter Feeds. *Transportation Research Part C: Emerging Technologies*, 131, Article ID: 103356. <https://doi.org/10.1016/j.trc.2021.103356>
- Feng, Y., & Lu, X. (2021). Construction Planning and Operation of Battery Swapping Stations for Electric Vehicles: A Literature Review. *Energies*, 14, Article No. 8202. <https://doi.org/10.3390/en14248202>
- Han, S., Yu, X., Li, D., & Dong, J. (2024). Optimized Layout of New Energy Vehicle Charging Station Based on Improved Quantum Particle Swarm Optimization. *Science Technology and Engineering*, 24, 11720-11725. (In Chinese)
- He, S., Xue, B., & Luo, D. (2025). Research on the Evolution of Online User Reviews of New Energy Vehicles in China Based on LDA. *World Electric Vehicle Journal*, 17, Article No. 21. <https://doi.org/10.3390/wevj17010021>
- Hu J., Xu X., & Hao W. (2020). Study on Technical and Standard Demands for Electric Vehicle Battery Swap. *China Auto*, No. 6, 47-51+64. (In Chinese)
- Kiefer, C., & Szolnoki, G. (2024). Exploring Market Potential and Consumer Preferences for Fungus-Resistant Grape Varieties: A Discrete Choice Analysis in Germany. *Future Foods*, 10, Article ID: 100449. <https://doi.org/10.1016/j.fufo.2024.100449>
- Lai, Z., & Li, S. (2024). Towards a Multimodal Charging Network: Joint Planning of Charging Stations and Battery Swapping Stations for Electrified Ride-Hailing Fleets. *Transportation Research Part B: Methodological*, 183, Article ID: 102928. <https://doi.org/10.1016/j.trb.2024.102928>
- Li, J., & Li, J. (2024). Consumers' Preferences for Wine Attributes Based on Conjoint Analysis. *Sino-Overseas Grapevine & Wine*, No. 5, 103-110. (In Chinese)
- Li, Y., Huang, B., & Xie, D. (2019). Visual Analysis on the Behavior Characteristics of Electric Vehicle Users Based on k-means Clustering. *Electrical Automation*, 41, 12-15, 81. (In Chinese)

- Li, S., Tong, L., Xing, J., & Zhou, Y. (2017). The Market for Electric Vehicles: Indirect Network Effects and Policy Design. *Journal of the Association of Environmental and Resource Economists*, 4, 89-133. <https://doi.org/10.1086/689702>
- Liang, C., Yang, Q., Sun, H., & Ma, X. (2024). Unveiling Consumer Satisfaction and Its Driving Factors of EVs in China Using an Explainable Artificial Intelligence Approach. *Humanities and Social Sciences Communications*, 11, Article No. 1575. <https://doi.org/10.1057/s41599-024-04120-z>
- Liu, Y., Li, G., Liu, Y., Hong, Y., & Liu, Y. (2019). Clustering Technology of Electric Vehicle Charging Behavior Based on Random Forest. *Electric Power Engineering Technology*, 38, 115-121. (In Chinese)
- Nastiti, K. R., Hidayatullah, A. F., & Pratama, A. R. (2021). Discovering Computer Science Research Topic Trends Using Latent Dirichlet Allocation. *Jurnal Online Informatika*, 6, 17-24. <https://doi.org/10.15575/join.v6i1.636>
- Özkara, Y., Bilişli, Y., Yildirim, F. S., Kayan, F., Başdeğirmen, A., Kayakuş, M. et al. (2025). Analysing Social Media Discourse on Electric Vehicles with Machine Learning. *Applied Sciences*, 15, Article No. 4395. <https://doi.org/10.3390/app15084395>
- Paneru, C. P., Tarigan, A. K. M., & Toşa, C. (2025). Exploring User Engagement and Perceptions towards Smart Energy Applications in Norway: Insights from a Text Mining Analysis of User Reviews. *Energy Efficiency*, 18, Article No. 25. <https://doi.org/10.1007/s12053-025-10309-1>
- Röder, M., Both, A., & Hinneburg, A. (2015). Exploring the Space of Topic Coherence Measures. In *Proceedings of the Eighth ACM International Conference on Web Search and Data Mining* (pp. 399-408). ACM. <https://doi.org/10.1145/2684822.2685324>
- Shao, J., Chen, X., Dai, Y., Huang, Z., Wu, C., Wang, Q., Guo, Y., & Xu, W. (2023). Research on Public Charging Demand of Electric Vehicles Based on Random Forest. *Distribution & Utilization*, 40, 58-64. (In Chinese)
- Sievert, C., & Shirley, K. (2014). LDAvis: A Method for Visualizing and Interpreting Topics. In *Proceedings of the Workshop on Interactive Language Learning, Visualization, and Interfaces* (pp. 63-70). Association for Computational Linguistics. <https://doi.org/10.3115/v1/w14-3110>
- Sun, H., Guo, W., Shao, H., & Rong, B. (2020). Dynamical Mining of Ever-Changing User Requirements: A Product Design and Improvement Perspective. *Advanced Engineering Informatics*, 46, Article ID: 101174. <https://doi.org/10.1016/j.aei.2020.101174>
- Tan, Y., Fukuda, H., Li, Z., Wang, S., Gao, W., & Liu, Z. (2022). Does the Public Support the Construction of Battery Swapping Station for Battery Electric Vehicles? Data from Hangzhou, China. *Energy Policy*, 163, Article ID: 112858. <https://doi.org/10.1016/j.enpol.2022.112858>
- Tian, J., Wang, P., & Zhu, D. (2024). Overview of Chinese New Energy Vehicle Industry and Policy Development. *Green Energy and Resources*, 2, Article ID: 100075. <https://doi.org/10.1016/j.gerr.2024.100075>
- Wang, G. (2025). User Demand Identification and Co-Occurrence Analysis Based on Online Comments—Taking New Energy Vehicles as an Example. *E-Commerce Letters*, 14, 754-766. <https://doi.org/10.12677/ecl.2025.143765>
- Wang, L., Li, Y., Wang, Z., Zhang, L., Xing, Y., & Ma, W. (2025). Analysis of Influencing Factors and Severity of Chain Rear-End Collision Using Text Mining. *Journal of Tongji University (Natural Science)*, 53, 1074-1083. (In Chinese)
- Wang, N., & Liang, Y. (2021). Policy Analysis and Prospects of New Energy Vehicle Charging Infrastructure under the Background of New Infrastructure Development. *Auto In-*

- dustry Research, No. 1*, 8-15. (In Chinese)
- Wang, T., Qin, J., & Wang, Y. (2023). SnowNLP-LDA Based Online Public Opinion Topic Analysis of Public Health Emergencies. *Journal of Shangqiu Normal University*, *39*, 11-14. (In Chinese)
- Wang, Y. (2024). Multi-Attribute Decision-Making Evaluation of Talent Demand and Competitiveness in Information Management and Information Systems Based on LDA-TIFN-TOPSIS. *Marketing Management Review, No. 30*, 149-151. (In Chinese)
- Wang, Y., Zhu, J., Wang, Z., Bai, F., & Gong, J. (2022). Review of Applications of Natural Language Processing in Text Sentiment Analysis. *Journal of Computer Applications*, *42*, 1011-1020. (In Chinese)
- Wang, F., & Zhang, Q. (2025). The Development of China's New Energy Vehicle Charging and Swapping Industry: Review and Prospects. *Energies*, *18*, Article No. 4548. <https://doi.org/10.3390/en18174548>
- Wang, Y.-Y., Chi, Y.-Y., Xu, J.-H., & Li, J.-L. (2021). Consumer Preferences for Electric Vehicle Charging Infrastructure Based on the Text Mining Method. *Energies*, *14*, Article No. 4598. <https://doi.org/10.3390/en14154598>
- Wu, P., & Lu, G. (2024). Research on the Overseas Social Media Communication Strategy of Chinese Auto Brands Based on LDA Model. *Foreign Economic Relations & Trade, No. 10*, 47-52. (In Chinese)
- Xia, Y., & Shan, X. (2018). Sentiment Analysis of Simple Text Based on Python. *Yinshan Academic Journal (Natural Science Edition)*, *32*, 58-62. (In Chinese)
- Xu, T. (2024). Implementation Effect and Development Suggestions of "Battery Swapping" Mode for New Energy Vehicles. *Trade Fair Economy, No. 13*, 156-160. (In Chinese)
- Yang, Z., Hu, X., Sun, J., Lei, Q., & Zhang, Y. (2023). Is It Worth Promoting Battery Swapping? A Social Welfare Perspective on Provider- and Consumer-Side Incentives. *Journal of Environmental Management*, *330*, Article ID: 117157. <https://doi.org/10.1016/j.jenvman.2022.117157>
- Zhang, Y., Wang, X., & Zhi, B. (2024). Strategic Investment in Electric Vehicle Charging Service: Fast Charging or Battery Swapping. *International Journal of Production Economics*, *268*, Article ID: 109136. <https://doi.org/10.1016/j.ijpe.2023.109136>
- Zhao, J., Xie, W., Lv, X., & Zhao, J. (2024). User Demand Identification and Development Trend Prediction Based on RF-BERT and UGC. *Information Science*, *42*, 132-142. (In Chinese)
- Zhao, W. X., Jiang, J., Weng, J., He, J., Lim, E., Yan, H. et al. (2011). Comparing Twitter and Traditional Media Using Topic Models. In P. Clough, C. Foley, C. Gurrin, G. J. F. Jones, W. Kraaij, H. Lee, & V. Mudoch (Eds.), *Advances in Information Retrieval* (Vol. 6611, pp. 338-349). Springer. [https://doi.org/10.1007/978-3-642-20161-5\\_34](https://doi.org/10.1007/978-3-642-20161-5_34)
- Zhao, X., Li, X., Jiao, D., Mao, Y., Sun, J., & Liu, G. (2024). Policy Incentives and Electric Vehicle Adoption in China: From a Perspective of Policy Mixes. *Transportation Research Part A: Policy and Practice*, *190*, Article ID: 104235. <https://doi.org/10.1016/j.tra.2024.104235>
- Zhao, X., Ma, Y., Shao, S., & Ma, T. (2022). What Determines Consumers' Acceptance of Electric Vehicles: A Survey in Shanghai, China. *Energy Economics*, *108*, Article ID: 105805. <https://doi.org/10.1016/j.eneco.2021.105805>