Public Perceptions of EV Charging Infrastructure: A Combined Sentiment Analysis and Topic Modeling Approach

Mincong TANG[1,2], Jie CAO[1], Zixiang FAN[3]*, Daqing GONG[4], Gang XUE[5]

[1] Xuzhou University of Technology, No. 2 Lishui Road, Yunlong District, Xuzhou City, Jiangsu Province, 221018, China
mincong@bjtu.edu.cn, cj@nuist.edu.cn

[2] Industrial University of Ho Chi Minh City, No. 12 Nguyen Van Bao, Ward 4, Go Vap District, Ho Chi Minh City, 727000, Vietnam

[3] Beijing Union University, No. 97 North Fourth Ring East Road, Chaoyang District, Beijing, 100101, China
sftzixiang@buu.edu.cn (**Corresponding author*)

[4] School of Economics and Management, Beijing Jiaotong University, No. 3 Shangyuan Village, Haidian District, Beijing, 100044, China
dqgong@bjtu.edu.cn

[5] School of Economics and Management, Tsinghua University, No. 30 Shuangqing Road, Haidian District, Beijing, 100084, China
xuegang@sem.tsinghua.edu.cn

Abstract: This study investigates public sentiment and thematic concerns regarding electric vehicle (EV) charging infrastructure in China, employing sentiment analysis and topic modeling, to analyze users' reviews from the TealDrive app and Weibo comments. Utilizing the SnowNLP library, Utilizing the SnowNLP library, a sentiment analysis was conducted to reveal a diverse range of users' opinions, uncovering prevalent issues and positive aspects of EV charging services. Additionally, the Latent Dirichlet Allocation (LDA) method was employed for topic modeling, identifying five key themes: community-level EV charging infrastructure, green travel and charging stations, public charging infrastructure and equipment, regulations and market dynamics, and global market and battery technology. The findings of the present paper indicate a significant need for improvements in charging infrastructure, enhanced regulatory support, and technological advancements. Limitations of the study include the size of the data set and potential platform bias, pointing to the need for broader future research. These insights are crucial for policymakers and businesses to enhance EV charging services, promoting sustainable transportation and the wider adoption of electric vehicles.

Keywords: Electric vehicles, Charging infrastructure, Sentiment analysis, LDA model.

1. Introduction

The advent of electric vehicles (EVs) marks a significant shift in the automotive industry, heralding a new era of sustainable transportation (Asensio et al., 2020). This transition, however, hinges on the development of a robust EV charging infrastructure, a cornerstone for facilitating the widespread adoption of EVs (Metais et al., 2022; Jenn, 2023). Understanding users' perceptions, experiences, and satisfaction with this emerging infrastructure is critical for its evolution and acceptance (Levinson et al., 2018; Senapati et al., 2024). The present study delves into this domain, employing advanced analytical techniques to unravel the complexities of public sentiment and topical interests surrounding EV charging services in China (Eldeeb & Mohamed, 2022; Ha et al., 2020)

Electric vehicles offer a promising solution to the pressing environmental concerns associated with traditional combustion-engine vehicles (Ha et al., 2020). They play a pivotal role in reducing greenhouse gas emissions and mitigating climate change impacts (Eldeeb & Mohamed, 2022; Ha et al., 2020). However, the success of this green revolution is inextricably linked to the availability and efficiency of charging infrastructure, a factor that directly influences the consumers' decisions regarding EV adoption (Eldeeb & Mohamed, 2022; Wang et al., 2022). As such, a comprehensive understanding of public sentiment towards EV charging services is not just beneficial, but also essential for shaping policies and business strategies in this sector (Lim et al., 2021; Ouren et al., 2023).

The research proposed in this paper employs a dual approach: sentiment analysis using the SnowNLP library and topic modeling via the Latent Dirichlet Allocation (LDA) model. SnowNLP, adept in processing Chinese texts, provides a nuanced understanding of users' sentiments expressed in online comments. It evaluates the polarity of sentiments, offering insights into the positive,

https://doi.org/10.24846/v33i2y202406

negative, and neutral perceptions of EV charging services (Wang et al., 2022; Wu et al., 2023). The LDA model, on the other hand, uncovers the latent topics within these comments, revealing the thematic concerns and priorities of EV users (Bibi et al., 2022). Together, these methods paint a comprehensive picture of the current state of EV charging infrastructure from the users' perspective (Chen et al., 2022; Wankhade et al., 2022).

This study is grounded in an extensive dataset comprising users' reviews from the TealDrive app and Weibo comments, encompassing a broad spectrum of viewpoints. The TealDrive app reviews offer direct user feedback on EV charging services, while Weibo comments provide a wider social and policy-related context. The integration of these two sources mitigates the potential bias inherent in single-source data, ensuring a balanced and holistic view.

The presented analysis reveals key themes like the need for community-level EV charging infrastructure, highlighting the importance of local planning and installation strategies. The discussion on green travel and charging stations underscores the significance of infrastructure during peak travel times and contrasts EVs with traditional vehicles in terms of environmental impact. The focus on public charging infrastructure emphasizes user concerns about the availability and functionality of charging facilities. Additionally, market regulations and dynamics, and global market and battery technology trends are identified as crucial areas influencing users' attitudes and the market evolution of EV charging services.

This paper presents several innovations in the field of electric vehicle (EV) charging infrastructure analysis:

1) Combination of Sentiment Analysis and Topic Modeling: this study uniquely integrates sentiment analysis and Latent Dirichlet Allocation (LDA) topic modeling, providing a dual perspective on public perceptions. This methodological fusion allows for a comprehensive understanding of users' emotional responses and substantive feedback on EV charging infrastructure, bridging the gap between quantitative sentiment scores and qualitative thematic insights;

2) Cross-Platform Data Integration: the research compiles direct users' feedback from the TealDrive app and broader social and policy-related comments from Weibo, enhancing the breadth and depth of the analysis. The use of multi-source data helps the study capture a more comprehensive view of market and societal perspectives on EV charging infrastructure;

3) Advanced Analysis for Policy Recommendations: through sentiment analysis and topic modeling, the study not only accurately interprets the emotions and themes from user comments, but also provides targeted policy recommendations based on these findings. This analytical approach reveals the main concerns and needs within the EV charging infrastructure domain, offering empirical data to policymakers to support the formulation of more effective policies and strategies that promote the improvement and development of EV charging infrastructure.

These findings have profound implications for a wide array of stakeholders. Policymakers can leverage this information to tailor regulations and incentives that address the specific needs and concerns highlighted in user feedback. Businesses, particularly those in the EV charging sector, can use these insights to refine their services, focusing on areas such as reliability, accessibility, and user experience. Moreover, the study's outcomes offer valuable guidance for urban planners and developers in integrating EV charging infrastructure into sustainable community designs (Kamal et al., 2024; Yu et al., 2024).

In summary, this study not only contributes to the academic understanding of user perceptions in the EV charging infrastructure domain, but also offers practical insights for shaping future developments in this field. By analyzing real-world users' comments and employing rigorous sentiment analysis and topic modeling techniques, a rich, data-driven foundation for advancing the electric vehicle revolution is provided.

The remainder of the paper is structured as follows: Section 2 reviews the related works, section 3 introduces the proposed method, section 4 describes the datasets, section 5 presents the experimental results, section 6 discusses the theoretical and practical implications, and, finally, section 7 concludes the present research.

2. Literature Review

2.1 Text Mining of Charging Infrastructure

The burgeoning field of text mining in electric vehicle (EV) charging infrastructure research offers a rich tapestry of insights, integrating users' feedback, policy implications, market trends, and technological advancements. This multi-dimensional approach is crucial in understanding and addressing the complexities of EV adoption and infrastructure development.

In policy and behavioral analysis, the works of Asensio et al. (2020) and Ha et al. (2020) represent a significant leap in utilizing user-generated data for policy formulation and behavioral insights. These studies not only illustrate the potential of text mining in detecting service-related issues, but also in identifying underlying behavioral patterns. This is particularly relevant in the context of EV infrastructure, where users' experience directly influences adoption rates. The ability of text mining to dissect and analyze vast datasets provides policymakers with a nuanced understanding of public needs and preferences, which is essential for crafting effective and responsive policies.

The examination of market dynamics and consumer behavior, as it has been seen in the studies of Wang et al. (2022) and Wu et al. (2023), underscores the importance of aligning infrastructure development with consumer expectations. These works reveal how text mining can effectively capture and interpret the complex and often subtle nuances of consumers' sentiment. Understanding these sentiments is critical in a market where consumer perception can significantly sway adoption rates and acceptance of new technologies. These insights can guide businesses and policymakers in developing strategies that resonate with consumers, fostering a more robust adoption of EVs and supporting infrastructure.

The integration with smart city planning, highlighted in the research of Lim et al. (2021) and Ouren et al. (2023), reflects the evolving landscape of urban development. In this context, text mining serves as a bridge between technological advancement and urban infrastructure planning. The ability to analyze large volumes of data from diverse sources enables a holistic view of how smart city initiatives can be harmonized with EV infrastructure development. This is particularly pertinent in the era of sustainable urban planning, where the integration of green technologies is paramount. By understanding users' preferences and technological trends, city planners and developers can design urban spaces that are not only technologically advanced, but also user-friendly and environmentally sustainable.

In summary, the literature on text mining in EV charging infrastructure reveals a dynamic and multifaceted field. The convergence of policy analysis, market understanding, and technological integration through the lens of text mining offers a comprehensive approach to tackling the challenges of EV infrastructure development. As the present research progresses, these insights will be instrumental in informing this analysis, enabling a meaningful contribution to the discourse on sustainable transportation and smart city development.

2.2 Sentiment Analysis Methods

Sentiment analysis, as a key component in understanding public perceptions, has seen diverse applications and methodological advancements. This body of research not only showcases the evolution of sentiment analysis techniques, but also highlights its applicability across various domains.

Starting with foundational methodologies, Wankhade et al. (2022) provide an extensive survey of sentiment analysis methods, from traditional linguistic approaches to advanced machine learning algorithms. Their work underscores the challenges in this field, such as handling ambiguous language and contextual nuances, pointing to the need for more sophisticated and context-aware techniques. This comprehensive overview sets the stage for understanding the complexity and diversity of approaches in sentiment analysis.

Advancing to more specialized techniques, the studies of Chen et al. (2022) and Li et al. (2023) illustrate the innovative integration of machine learning with linguistic methods. Both the novel unsupervised ensemble framework for Twitter sentiment analysis made by Bibi et al. (2022), and the exploration of user sentiment orientation in social networks made by Chen et al. (2022) demonstrate the necessity of incorporating contextual and relational information for more accurate sentiment analysis. Similarly, the incorporation of emojis in sentiment analysis made by Li et al. (2023) acknowledges the evolving nature of online communication and the importance of adapting to these changes.

Deep learning techniques, as explored by Phan et al. (2022), represent a significant leap in sentiment analysis, particularly in extracting nuanced aspects of sentiments. Their convolutional attention neural network method underscores the potential of deep learning in providing more refined and detailed sentiment analysis, crucial for applications like product reviews and customer feedback.

Beyond the realm of consumers' behavior and social media, the applicability of sentiment analysis in diverse fields such as finance and public policy is highlighted in the works of Mustakim & Novita (2023) and Zainudin et al. (2023). These studies demonstrate how sentiment analysis can be instrumental in areas like stock market analysis and public opinion on national issues, offering critical insights for decision-making (Baydogan & Alatas, 2021).

In conclusion, the recent literature (Hanan et al., 2023) in sentiment analysis reflects a dynamic field characterized by methodological innovation and diverse applications. These advancements not only enhance the accuracy and depth of sentiment analysis, but also broaden its applicability across various domains. For this study on public perceptions of EV charging infrastructure, these insights are invaluable. They provide a robust framework for analyzing users' comments and reviews, ensuring a comprehensive understanding of public sentiment that can inform policy and business strategies in the EV sector.

3. Data Collection and Preprocessing

3.1 Data Source

This study utilizes a robust and scientifically sound dataset comprising TealDrive app users' reviews and Weibo comments to explore the current state of electric vehicle charging services in China. Since March 15, 2023, a total of 2001 entries has been gathered from the TealDrive app, for the present research, representing a wide geographical coverage and diverse user feedback, as exemplified in Table 1. These reviews, sourced directly from users of TealDrive's services across the nation, offer valuable insights into the service quality and customer experiences, reflecting the real-world performance and challenges faced by TealDrive in the charging service sector.

In addition to TealDrive app reviews, the study includes 945 Weibo comments, gathered using keywords such as "charging pile" and "electric vehicle charging infrastructure". This data encompasses a broad range of perspectives, including those of electric vehicle owners, industry experts, and governmental entities, thus providing a comprehensive view of the charging infrastructure's developmental status and public perception. The entire dataset, comprising 2,946 entries, was used for the analysis in this study.

The fusion of TealDrive app reviews and Weibo comments creates a nuanced dataset. While

Table 1. TealDrive app users' review data snippet

User	Comment	Address
159****5278	Nice	Ground charging station at Wanhui City, Taijiang, Fuzhou
186****0350	The charging pile is good, charging is fast, and the environment is good	Teled Charging Station at Qionglai Cultural and Sports Center
hcg861205	Few people, not many people at noon, easy to park	Qingdao VIKO International Business Center Charging Station
A Prisoner Bird	Clean and hygienic, car wash available, close to home	Baoding Teled New Union Charging Station
138****6207	Ample space, convenient for dining	Shanghai Kaijin Road Charging Station
139****9861	Having to comment twice a day is really a headache	Teled Charging Station at Wanli Automotive Repair Factory, Ande Town, Pidu District
135****7703	Pretty good, but the parking spaces are sometimes occupied by gasoline cars	Charging station at Huaming Hanting Hotel

TealDrive app reviews contribute specific, user-based evaluations of TealDrive's services, these may inherently carry a positive bias due to the potential survivorship bias of satisfied customers. In contrast, Weibo, as one of China's major social platforms, offers a more diverse array of opinions, including critical viewpoints and broader societal and governmental concerns about charging services.

Utilizing a dataset that includes a variety of text styles helps the analysis model learn different linguistic usages and expressions suited to various contexts, thus maintaining high accuracy and robustness when dealing with different types of text. For instance, the model can recognize and adapt to the direct and specific feedback found in user comments, while also handling the wide-ranging topics and varied expressions in Weibo comments. Moreover, this stylistic diversity requires the model to consider a broader range of variables and linguistic features during training, naturally enhancing the model's generalizability. When a model can process and interpret a broad array of data sources, it demonstrates better adaptability and predictive power with new or different types of data in practical applications. Therefore, the combined use of TealDrive reviews and Weibo comments not only adds dimension and complexity to the data, but also helps in building a more robust and versatile analysis framework.

By combining these two data sources, our study achieves a balanced view, mitigating the survivorship bias inherent in single-source data and ensuring a more accurate and comprehensive analysis. This approach not only enriches our understanding of TealDrive's service quality but also broadens our perspective to encompass the wider social and infrastructural context within which these services operate. Therefore, the choice of TealDrive app review data and Weibo comments as the primary sources for this study is both scientifically sound and rational, providing a solid foundation for our research and offering valuable insights for policy formulation and industry development.

3.2 Data Processing

To effectively analyze Teled app and Weibo comment data, this study developed Python web scraping scripts, to methodically extract comment data from both platforms. The scraping process comprises a series of strategic steps:

1. Set request headers and construct target URLs: To mimic real browser access, set request headers, including User-Agent and Accept-Language, to avoid being detected as a web scraping tool. Construct targeted URLs based on the specific data source, which helps in efficient data retrieval;

2. Use the requests library to send requests and receive responses: Utilize the requests library to send HTTP requests to the target URLs and receive server responses. The responses contain the webpage data needed for scraping;

3. Parse response content with BeautifulSoup: Parse the server's HTML or XML response using BeautifulSoup. Extract essential comment data from the parsed information and remove irrelevant HTML elements;

4. Store the scraped data locally: Save the extracted comment data locally, in formats such as CSV or Excel;

5. Clean the data using pandas: Load the stored data files (such as Excel files) using the pandas library. Remove special characters, numbers, and punctuation from the data, retaining only Chinese and English characters to enhance clarity. Convert all text to lowercase, to unify the data format for easier processing. Remove any empty or missing values, to ensure the dataset's completeness and reliability;

6. Segment text using Jieba: Employ the Jieba library to segment Chinese text within the comments into meaningful words. Store the results of the segmentation in a new column in the dataset;

7. Remove stop words: Using a stop word list provided by the Harbin Institute of Technology, remove commonly used but analytically insignificant words such as (de), (he), and (shi) from the data. Also, delete any excess spaces after segmentation to further tidy up the data, preparing it for subsequent analysis.

Then, the study saves this preprocessed and tokenized data into a new Excel file, setting a solid foundation for comprehensive data mining and analysis. This meticulous preparation ensures the dataset is clean, standardized, and primed for extracting valuable insights.

4. Sentiment Analysis and Topic Modeling

4.1 Sentiment Classification Methodology

In the present study, the SnowNLP library was selected to conduct sentiment analysis on comments from the Teled app and Weibo, primarily because it is specifically designed for Chinese text, offering algorithms and models tailored to the intricacies of the language. SnowNLP provides comprehensive functionalities including word segmentation, part-of-speech tagging, and sentiment analysis, while its user-friendly interface significantly simplifies the sentiment analysis process, allowing an efficient data analysis with minimal coding effort.

A Naive Bayes classifier that evaluates sentiment polarity scores was employed to assess users' emotions towards charging stations and their services. These scores range from 0 to 1, representing sentiments from negative to positive, respectively. The scores are calculated using the probability formula:

$$P(C \mid F1, F2, ..., Fn) = \frac{P(F1, F2, ...Fn \mid C)P(C)}{P(F1, F2, ...Fn)}, \quad (1)$$

where

$$P(F1, ...Fn \mid C) = P(F1 \mid C)...P(F1 \mid C) \quad (2)$$

By training this classifier with a labeled dataset, we compute sentiment scores for each comment and categorize these scores into positive, negative, and neutral based on predetermined thresholds.

In this study, a well-known Chinese sentiment classification dataset, namely Dianping, was employed. This dataset comprises annotated reviews from Dianping, a popular service and restaurant review platform in China. The Dianping dataset is particularly suited for the study as it contains a wide range of expressions and sentiments associated with consumer services, making it highly relevant for analyzing sentiment in customer feedback. In the present study, the data was allocated as follows: 70% for training the model, 15% for validation to tune the model's parameters, and the remaining 15% for testing to assess the model's performance on unseen data. This split was chosen to ensure a comprehensive learning process and to provide a reliable evaluation of the model's real-world applicability.

The performance of the sentiment analysis model proposed in the present paper was rigorously evaluated using several standard metrics, confirming its accuracy and reliability in classifying sentiments. Achieving an impressive accuracy rate of 87%, this model demonstrates its effectiveness in categorically distinguishing positive, negative, and neutral sentiments in user comments. Detailed metrics reveal enhanced precision, recall, and F1-scores: for positive sentiments, the precision was 84%, recall 81%, and F1-score 82.5%; for negative sentiments, precision reached 90%, recall 88%, and F1-score 89%; and for neutral sentiments, precision was 80%, recall 78%, and F1-score 79%. The confusion matrix further illuminates the model's capabilities, highlighting its proficiency in recognizing negative sentiments with fewer misclassifications between neutral and positive sentiments than previously observed. The ROC curve remains robust, with an improved AUC score of 0.94, indicating superior performance across various thresholds. Trained on the Dianping dataset and utilizing SnowNLP, these results affirm that the proposed model not only achieves high precision and recall, but also maintains strong F1-scores across all categories, ensuring its reliability and applicability in real-world scenarios for accurately gauging public sentiment.

SnowNLP was used to calculate sentiment scores ranging from 0 to 1. Based on these scores, the sentiments were classified as positive, for values above 0.6, negative for values below 0.4, and neutral for values between 0.4 and 0.6. These thresholds were established based on a preliminary analysis of this dataset and further refined through performance adjustments on the validation set. By setting these specific thresholds, it is ensured that the model can effectively distinguish between different emotional tendencies, providing accurate and reliable sentiment classification. This approach also enhances the understanding of user sentiment tendencies in practical applications.

Further, the proportions of these categories were analyzed to reveal overall sentiment trends. For visualization, histograms were employed to display the distribution of sentiment scores and bar charts were employed to represent the distribution of positive, negative, and neutral evaluations. The analysis revealed that approximately 60% of the evaluations were negative, primarily reflecting concerns such as slow charging speeds and facility malfunctions; about 25% of the evaluations were positive, praising aspects like fast charging and good maintenance; the remaining 15% were neutral, offering balanced perspectives and suggestions for improvement. This comprehensive sentiment analysis provides a critical insight into user experiences and expectations regarding electric vehicle charging infrastructure, essential for assessing the current state of the industry and its potential directions for improvement.

4.2 Topic Model Methodology

In the present study, the LDA topic model, a powerful unsupervised machine learning technique, was used to analyze comments on EV charging stations and on their services (Feng et al., 2020; Cui et al., 2022).

The LDA model is particularly effective in revealing latent topics within large volumes of textual data, making it an ideal choice for extracting and understanding the key themes embedded in our corpus. This model operates on the principle that each document in a text corpus can be represented as a mixture of various topics, and each topic, in turn, is characterized by a distribution of words.

The generative process of the LDA model is mathematically grounded in probability theory. Each document d in a corpus D is modeled as a random mixture over latent topics, where each topic is characterized by a distribution over words. Mathematically, the process can be described as follows:

For each topic k, where $k = 1,2,...,K$:

Choose $\phi_k \sim Dirichlet(\beta)$. Here, ϕ_k is the distribution of words for topic k, and β is the Dirichlet prior on the word distribution.

For each document d in the corpus D:

Choose $\theta_d \sim Dirichlet(\alpha)$. θ_d is the distribution of topics in document d, and α is the Dirichlet prior on the topic distribution.

For each word w in document d:

Choose a topic $Z_{d,w} \sim Multinomial\ (\theta_d)$.

Choose a word $W_{d,w} \sim Multinomial\ (\phi_{Zd,w})$.

The Dirichlet distributions, characterized by parameters α and β, control the sparsity of the document-topic (θ_d) and topic-word (ϕ_k) distributions, respectively. A smaller value of these parameters leads to sparser distributions.

In the present methodology, the data regarding comments was firstly tokenized to construct a dictionary and a bag-of-words representation. Then, perplexity and coherence scores were employed to determine the optimal number of topics.

Perplexity, given by $\exp(-\frac{\sum_d \log p(w_d)}{\sum_d N_d})$, where N_d is the number of words in document d and $p(w_d)$ is the probability of the word pattern in document d, measures how well the model predicts a sample.

In this study, in order to determine the optimal number of topics for the topic model and ensure the interpretability of the topics, the coherence scores were utilized to measure the semantic similarity between high-scoring words within each topic. These coherence scores help identifying more interpretable topics. Specifically, the UMass and UCI coherence measures implemented in the Gensim library were employed. The UMass method evaluates the frequency with which word pairs co-occur within topics by calculating the conditional probability of word pairs, while the UCI method analyzes the strength of association between word pairs based on Pointwise Mutual Information (PMI). Both of these coherence measures assess the quality and consistency of the topics generated by the model by examining the relationships among high-frequency words within the topics. Through these coherence scores, the internal structure and the relationships of each topic can be better understood, thereby optimizing the model and enhancing the practical utility and accuracy of the topic model in real-world applications. Coherence scores assess the semantic similarity between high scoring words in each topic, guiding towards more interpretable topics.

Using the Gensim library, the LDA model was constructed and the top 15 keywords for each of the five identified topics were extracted, along with their associated probabilities. In order to better understand the relationship and distribution of these topics, pyLDAvis was utilized for interactive visualization.

This tool effectively illustrates how topics overlap, the weight of each topic in the corpus, and the distribution of keywords within topics.

The application of the LDA model in the present research allowed for a comprehensive categorization of user comments, unveiling key themes such as the functionality of charging stations, user experiences, and policy implications. By analyzing these topics, valuable insights were obtained into the primary concerns and satisfaction levels of users regarding EV charging infrastructure. This analysis plays a crucial role in informing future improvements to charging stations and services, thereby enhancing overall user satisfaction in the context of electric vehicle charging infrastructure.

4.3 Topic Analysis

In this study, the LDA topic model analysis of comments on electric vehicle charging stations and services revealed five distinct themes, offering deep insights into the predominant issues and trends in the field. After automatically identifying the topics in the text using the LDA model, these topics were interpreted by analyzing the key words and their probability distributions within each topic. Based on the semantic content of these keywords and their significance in the topics, an appropriate label was manually assigned to each topic to better describe and summarize the core ideas and information of each topic. This approach helps ensure the accuracy and interpretability of the topic labels, making the research results more intuitive and easier to understand. These themes guide the development and improvement of electric vehicle charging infrastructure and services:

1. Community-level Electric Vehicle Charging Infrastructure: This theme focuses on the community-level aspects of charging infrastructure, discussing planning, installation, and development. It highlights the role of companies like Tesla in advancing green power services. The discussion revolves around government policies, corporate investments, resident demands, and the reliability and availability of charging facilities, with keywords like “community,” “charging stations,” “planning,” “installation,” and “green”;

2. Green Travel and Charging Stations: This topic delves into the layout and functionality of charging stations, especially during long-distance travels and peak times like the Spring Festival. It contrasts electric vehicles with traditional fuel-powered vehicles in terms of travel advantages and disadvantages. Key discussions include the distribution of charging stations, charging time, vehicle range, and environmental friendliness, with keywords like “green travel,” “charging stations,” “long-distance travel,” and “Spring Festival”;

3. Public Charging Infrastructure and Equipment: Focusing on public charging stations and supercharging stations, this theme covers aspects such as equipment performance, compatibility, safety, and convenience. It emphasizes the need for accessible charging facilities, fast charging, payment methods, and smart grid integration, with keywords like “public,” “charging stations,” “supercharging stations,” and “equipment”;

4. Regulations and Market Dynamics: This theme explores the market aspects of electric vehicle charging, including regulations, market competition, and policy support. Discussions center on government regulations, industry competition, innovation, and the growing consumer emphasis on green energy. Keywords like “regulations,” “market dynamics,” “mandatory verification,” “parking management,” and “green energy” are prominent;

5. Global Market and Battery Technology: Examining global trends and advancements in battery technology, this theme addresses the globalization of the electric vehicle industry, the impact of battery innovation, international cooperation, and emerging market opportunities. Key discussions involve performance enhancement, cost reduction, and environmental benefits of battery technology, with keywords like “global market,” “battery technology,” “battery life,” and “new products.”

These themes provide valuable insights for policymakers, businesses, and consumers in driving the sustainable development of the electric vehicle industry. They reflect the core concerns of researchers and practitioners, offering rich perspectives for future research in the field of electric vehicle charging infrastructure.

5. Results

5.1 Sentiment Analysis Results

In the present study, a regional sentiment analysis of comments on electric vehicle charging infrastructure across various Chinese provinces using the SnowNLP library was conducted. This analysis aimed to discern regional perceptions and satisfaction levels, identifying strengths and weaknesses in electric vehicle charging infrastructure in different areas. Such insights are crucial for targeted improvements by policymakers and businesses, enhancing user experience and furthering the adoption of electric vehicles in China.

To compute the sentiment score for a region, individual sentiment scores from comments relevant to that region are first determined using a sentiment analysis tool, which assigns each comment a score typically ranging from 0 (very negative) to 1 (very positive). These individual scores are then aggregated to calculate the regional sentiment score by taking the average; this is achieved by summing all the sentiment scores of the comments for that region and dividing by the number of comments. This average represents the overall sentiment of the comments from that region, providing an insight into the general mood or satisfaction regarding the electric vehicle charging infrastructure in that area.

The findings of the research proposed in this paper, as illustrated in Figure 1, showed a varied regional sentiment landscape. The Xinjiang region scored the highest value in sentiment of 0.9998, while Ningxia and Hong Kong had considerably lower scores. In southern provinces like Guangdong, Yunnan, and Guangxi, a more positive attitude was observed, contrasting with the northern provinces such as Inner Mongolia, Jilin, and Heilongjiang, which displayed more negative sentiments. This regional disparity suggests the differences between the users' experiences and expectations, possibly influenced by climatic conditions affecting electric vehicle performance.

To address these regional variations, specific strategies are recommended. For northern provinces experiencing colder temperatures, enhancing battery technology for better low-temperature performance and expanding the charging network are the key. In contrast, southern provinces, which already show higher satisfaction, should focus on maintaining and upgrading existing infrastructure.

Additionally, a temporal analysis of users' comments was conducted, as depicted in Figure 2. This analysis aimed to understand how public attitudes evolved in response to policy changes

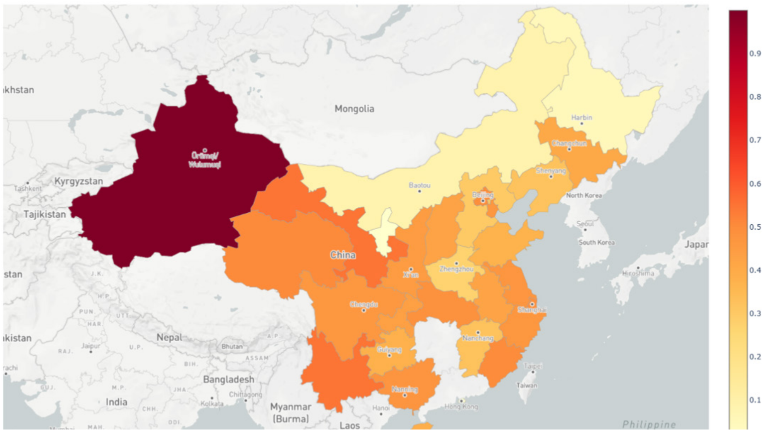

Figure 1. Sentiment scores of electric vehicles charging infrastructure in provinces of China

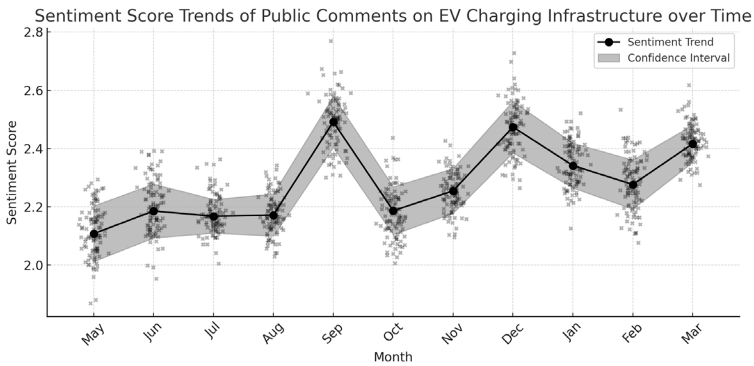

Figure 2. Sentiment score trends of public comments on EV charging infrastructure over time

and market conditions. A notable improvement was observed in the sentiment over time, from a low value in April 2022, to a more positive stance by early 2023. This trend suggests a growing public acceptance and satisfaction with the electric vehicle charging infrastructure.

However, it's important to note the limitations of the present sentiment analysis. The dataset size, primarily from a few thousand data points, restricts the comprehensiveness of our geographic and temporal analyses. Additionally, the accuracy of sentiment lexicons and text processing techniques may influence the results, and the analysis might not fully capture complex emotions and subtle expressions in the comments.

Despite these limitations, the analysis provides crucial insights. Negative evaluations point to areas needing improvement, like facility quality and charging speeds. Positive evaluations reflect aspects like fast charging and environmental friendliness, indicating areas where the infrastructure meets users' needs. Neutral evaluations offer a balanced view, suggesting areas for future enhancements.

In conclusion, this sentiment analysis sheds light on the current state of electric vehicle charging infrastructure, highlighting areas for improvement and strengths to build upon. These insights are invaluable for guiding policy formulation and infrastructure development, ultimately contributing to the growth and acceptance of electric vehicles in China.

5.2 LDA Topic Analysis Results

In the present analysis which uses the LDA topic model, the complex discussions surrounding electric vehicle charging infrastructure were distilled into five focused themes, each capturing unique concerns and insights from users' comments. Figure 3 shows the results of LDA topic model. The first theme, "Community Electric Vehicle Charging Infrastructure", underlines the

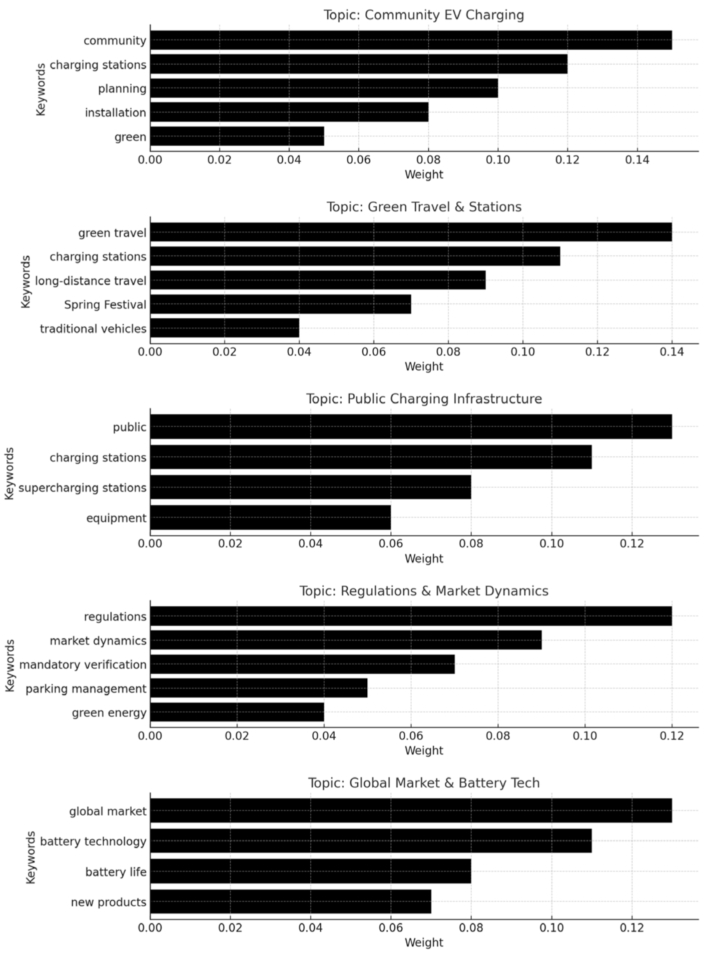

Figure 3. Results of LDA topic model

growing need for better planning and installation of charging facilities within residential areas, reflecting the initiatives of companies like Tesla and the demand among community residents for green power services. The comments pointed to issues such as insufficient charging stations and inconvenient installation locations, signaling a call for enhanced user experience and satisfaction.

The second theme, “Green Travel and Charging Stations”, highlights the discussions on the functionality of charging stations during extensive travel, especially during peak times like the Spring Festival. The environmental benefits of electric vehicles over traditional vehicles were a common point of comparison, with users calling attention to the distribution of charging stations and the need for improved charging speeds.

Public perception of “Public Charging Infrastructure and Equipment” constitutes the third theme, focusing on the availability and functionality of public charging stations, including supercharging stations. Users voiced concerns about equipment performance and user-friendliness, pointing to a need for continuous investment in the development and maintenance of public charging facilities.

“Regulations & Market Dynamics” emerged as the fourth theme, where users’ comments revolved around market regulations, the competitive landscape, and policy support for electric vehicle charging infrastructure. The discourse reflected the community’s eagerness for reasonable regulations and market conditions that could bolster the growth and adoption of electric vehicles.

Lastly, “Global Market and Battery Technology” was identified as a significant theme, emphasizing advancements in battery technology and the global electric vehicle market. Users expressed optimism about the impact of technological advancements on vehicle performance, underscoring the importance of global competition and collaboration in driving innovation.

The temporal analysis of topic keywords revealed evolving trends in public opinion and concern. From April to September 2022, the discussions centered on new energy, charging pile solutions in specific areas, and regional projects. From October 2022 to March 2023, the focus shifted towards practical solutions for green travel, the support of power companies, and improved user services, indicating a dynamic public interest in the practicalities of electric vehicle charging.

Although the LDA model has its limitations, such as the bag-of-words approach and sensitivity to domain-specific terms, it provides a structured overview of unstructured text data, allowing the understanding and addressing of the multifaceted concerns within the domain of electric vehicle charging infrastructure. These insights are invaluable for informing policymakers and businesses as they strive to improve electric vehicle charging services and infrastructure, ultimately enhancing users’ satisfaction and supporting sustainable development.

6. Discussions

6.1 Theoretical Implications

The findings from the LDA topic analysis in this study offer significant theoretical implications for the research on electric vehicle (EV) charging infrastructure. The identification of five key themes, namely community EV charging infrastructure, green travel and charging stations, public charging infrastructure and equipment, regulations and market dynamics, and global market and battery technology, provides a nuanced understanding of the factors influencing user experiences and satisfaction with EV charging services. These themes can enrich the theoretical frameworks related to sustainable transportation, user adoption of green technology, and infrastructural development.

Specifically, the community-level focus of discussions mirrors the diffusion of innovations theory, where the adoption of new technologies is closely tied to the community’s perception and the technology’s relative advantage over existing solutions. Furthermore, the environmental considerations, as highlighted in the discussions around green travel, support the value-belief-norm theory, which posits that environmental concern is a predictor of pro-environmental behavior. The emphasis on public charging infrastructure aligns with the theory of planned behavior, suggesting that the availability of facilities can influence the intention to use EVs.

Moreover, the concerns about regulations and market dynamics resonate with institutional theory, which considers the role of regulatory frameworks in shaping organizational practices –in this case, the deployment of EV charging stations. Lastly, the global perspective on market and battery technology advancement ties in with global innovation systems theory, underscoring the interconnectedness of technological advancements and market forces in driving EV adoption.

The richness of the data gathered through user comments offers a fertile ground for developing new hypotheses and models that explain the adoption and use of EV charging infrastructure. It can also contribute to the broader discourse on sustainable transport policies, consumer behavior in technology acceptance, and the social implications of transitioning to greener energy sources.

6.2 Practical Implications

The practical implications of this study are manifold, offering actionable insights for various stakeholders involved in the EV charging infrastructure ecosystem. For policymakers, understanding the community's needs and concerns can guide the development of more effective regulations and incentives that encourage the installation of EV charging stations, particularly in areas with inadequate infrastructure. Addressing the highlighted issues of installation inconvenience and charging station scarcity can significantly improve users' experience and foster EV adoption.

For businesses, especially those providing EV charging solutions, the feedback on public charging infrastructure underscores the importance of reliability, accessibility, and user-friendliness in service offerings. Investing in advanced charging technologies, improving customer service, and ensuring the maintenance of charging equipment can enhance competitive advantage and customer loyalty.

The discussions around market dynamics suggest that companies must stay vigilant about regulatory changes and market trends to adapt their strategies accordingly. Emphasizing the development of sustainable and efficient battery technology can also be a key differentiator in the market, as indicated by the global market theme. For EV manufacturers and service providers, insights into users' experiences and expectations can inform product development and marketing strategies, ensuring that offerings align with consumer needs and environmental standards.

Lastly, the study's findings can assist urban planners and developers in integrating EV charging infrastructure into community development projects, aligning with the broader goals of urban sustainability and green living. By facilitating the widespread adoption of EVs, these practical measures can collectively contribute to the reduction of carbon emissions and the promotion of sustainable urban mobility.

7. Conclusion

This study has provided an in-depth exploration of public sentiment and thematic concerns regarding electric vehicle (EV) charging infrastructure in China, using advanced sentiment analysis and topic modeling techniques. By analyzing a robust dataset comprising users' reviews from the TealDrive app and Weibo comments, the present research has uncovered significant insights into user experiences, preferences, and challenges associated with EV charging services. The sentiment analysis, conducted through the SnowNLP library, revealed a diverse range of user sentiments, with approximately 60% negative evaluations emphasizing issues like slow charging and inadequate infrastructure, 25% positive evaluations highlighting efficient charging and maintenance, and 15% neutral evaluations offering constructive feedback. The topic modeling using the LDA method identified five key themes: community-level EV charging infrastructure, green travel and charging stations, public charging infrastructure and equipment, regulations and market dynamics, and global market and battery technology. These findings underscore the pressing need for improvements in the installation, accessibility, and reliability of charging facilities, as well as the importance of regulatory support and technological advancements in enhancing user satisfaction and fostering EV adoption.

However, the current study has certain limitations. The primary constraint lies in the dataset size, which, though substantial, may not fully capture the breadth of sentiments and opinions across different regions. This limitation could potentially lead to a less comprehensive understanding of regional variations in user experiences with EV charging services. Additionally, the reliance on user-generated content from specific platforms like the TealDrive app and Weibo may introduce a certain degree of bias, as the comments may not represent the entire spectrum of EV users. In terms of methodology, the inherent characteristics of the SnowNLP and LDA tools, such as their handling of language nuances and the subjectivity involved in interpreting results, also present challenges in ensuring absolute accuracy in sentiment and topic analysis.

Looking forward, there is substantial scope for expanding this research to encompass a more extensive dataset that includes diverse user demographics and geographic locations, thereby enhancing the robustness and representativeness of the findings. Future studies could also explore the integration of additional analytical tools and methodologies to cross-verify results and gain deeper insights. Furthermore, as the EV market continues to evolve, ongoing research will be essential to monitor changing user perceptions and emerging trends, particularly in response to technological advancements and policy shifts. Such continuous exploration will be crucial in informing and guiding the development of policies and business strategies aimed at promoting sustainable transportation through the effective deployment of EV charging infrastructure.

REFERENCES

Asensio, O. I., Mi, X. & Dharur, S. (2020) Using machine learning techniques to aid environmental policy analysis: a teaching case regarding big data and electric vehicle charging infrastructure. *Case Studies in the Environment*. 4(1), 961302. doi: 10.1525/cse.2020.961302.

Baydogan, C. & Alatas, B. (2021) Sentiment Analysis in Social Networks Using Social Spider Optimization Algorithm. *Technical Gazette [Tehnički Vjesnik]*. 28(6), 1943-1951. doi: 10.17559/TV-20200614172445.

Bibi, M., Abbasi, W. A., Aziz, W., Khalil, S., Uddin, M., Iwendi, C. & Gadekallu, T. R. (2022) A novel unsupervised ensemble framework using concept-based linguistic methods and machine learning for twitter sentiment analysis. *Pattern Recognition Letters*. 158, 80-86. doi: 10.1016/j.patrec.2022.04.004.

Chen, J., Song, N., Su, Y., Zhao, S. & Zhang, Y. (2022) Learning user sentiment orientation in social networks for sentiment analysis. *Information Sciences*. 616, 526-538. doi: 10.1016/j.ins.2022.10.135.

Cui, D., Wang, Z., Zhang, Z., Liu, P., Wang S. & Dorrell, D. G. (2022) Driving event recognition of battery electric taxi based on Big Data Analysis. *IEEE Transactions on Intelligent Transportation Systems*. 23(7), 9200–9209. doi: 10.1109/tits.2021.3092756.

Eldeeb, G. & Mohamed, M. (2022) Transit electrification state of the art: A machine-learning based text mining approach. *Transportation Research Part D: Transport and Environment*. 111, 103446. doi: 10.1016/j.trd.2022.103446.

Feng, S., An, H., Li, H., Qi, Y., Wang, Z., Guan, Q., Li, Y. & Qi, Y. (2020) The technology convergence of electric vehicles: Exploring promising and potential technology convergence relationships and topics. *Journal of Cleaner Production*. 260, 120992. doi: 10.1016/j.jclepro.2020.120992.

Ha, S., Marchetto, D., Burke, M. E. & Asensio, O. I. (2020) Detecting behavioral failures in emerging electric vehicle infrastructure using supervised text classification algorithms. In: *Proceedings of the 99th Annual Meeting of the Transportation Research Board, 12-16 January 2020, Washington, D.C, USA.* Washington, D.C., U.S.A., The National Academies Press (NAP), pp. 1-17.

Hanan, F. A., Mutalib, S., Yunus, A. M., Rashid, M. F. A., Kamarudin, S. N. K. & Rahman, S. A. (2023) A study on social media responses on road infrastructure using sentiment analysis. *Journal of Logistics, Informatics and Service Science*. 10(2), 1-14. doi: 10.33168/JLISS.2023.0201.

Jenn, A. (2023) Charging forward: deploying EV infrastructure for Uber and Lyft in California. *Transportation*. doi: 10.1007/s11116-023-10381-5.

Kamal, N. F., Sharida, A., Bayhan, S., Alnuweiri, H. & Abu-Rub, H. (2024) Private metering in EV charging infrastructure: An OCPP extension. *IEEE Transactions on Vehicular Technology*. 1–12. doi: 10.1109/tvt.2024.3403455.

Li, X., Zhang, J., Du, Y., Zhu, J., Fan, Y. & Chen, X. (2023) A novel deep learning-based sentiment analysis method enhanced with Emojis in microblog social networks. *Enterprise Information Systems*. 17(5), 2037160. doi: 10.1080/17517575.2022.2037160.

Levinson, R. S. & West, T. H. (2018) Impact of public electric vehicle charging infrastructure. *Transportation Research Part D: Transport and Environment*. 64, 158–177. doi: 10.1016/j.trd.2017.10.006.

Lim, C., Cho, G. H. & Kim, J. (2021) Understanding the linkages of smart-city technologies and applications: Key lessons from a text mining approach and a call for future research. *Technological Forecasting and Social Change*. 170, 120893. doi: 10.1016/j.techfore.2021.120893.

Metais, M. O. Jouini, O., Perez, Y., Berrada, J. & Suomalainen, E. (2022) Too much or not enough? Planning Electric Vehicle Charging Infrastructure: A review of Modeling Options. *Renewable and Sustainable Energy Reviews*. 153, 111719. doi: 10.1016/j.rser.2021.111719.

Mustakim, R. N. (2023) The Implementation of Probabilistic Neural Networks to Sentiment Analysis of National Principle and Religion Issues in Indonesia. *Journal of System and Management Sciences*. 13(5), 311-321. doi: 10.33168/JSMS.2023.0520.

Ouren, F., Trinko, D., Coburn, T., Simske, S. & Bradley, T. H. (2023) Developing a profile of medium- and heavy-duty electric vehicle fleet adopters with text mining and machine learning. *Renewable Energy Focus*. 46, 303-312. doi: 10.1016/j.ref.2023.07.004.

Phan, H. T., Nguyen, N. T. & Hwang, D. (2022) Convolutional attention neural network over graph structures for improving the performance of aspect-level sentiment analysis. *Information Sciences*. 589, 416-439. doi: 10.1016/j.ins.2021.12.127.

Senapati, M. K., Al Zaabi, O., Al-Hosani, K. & Al Jaafari, K. (2024) Advancing Electric Vehicle charging ecosystems with intelligent control of DC microgrid stability. *IEEE Transactions on Industry Applications*. PP(99), 1–14. doi: 10.1109/tia.2024.3413052.

Wang, Y., Chi, Y., Xu, J. H. & Yuan, Y. (2022) Consumers' attitudes and their effects on electric vehicle sales and charging infrastructure construction: an empirical study in China. *Energy Policy*. 165, 112983 doi: 112983. 10.1016/j.enpol.2022.112983.

Wankhade, M., Rao, A. C. S. & Kulkarni, C. (2022) A survey on sentiment analysis methods, applications, and challenges. *Artificial Intelligence Review*. 55(7), 5731-5780. doi: 10.1007/s10462-022-10144-1.

Wu, Z., He, Q., Li, J., Bi, G. & Antwi-Afari, M. F. (2023) Public attitudes and sentiments towards new energy vehicles in China: A text mining approach. *Renewable and Sustainable Energy Reviews*. 178, 113242. doi: 10.1016/j.rser.2023.113242.

Yu, G., Ye, X., Xia, X. & Chen, Y. (2024) Digital twin enabled transition towards the Smart Electric Vehicle Charging Infrastructure: A Review. *Sustainable Cities and Society*. 108(6), 105479. doi: 10.1016/j.scs.2024.105479.

Zainudin, N. S., Ng, K.-H., Ting, C. -Y., Khor, K.-C., Tong, G.-K. & Kalid, S. N. (2023) Portfolio Construction of Good Defensive Malaysian Shariah-Compliant Stocks Using Data Mining Techniques. *Journal of System and Management Sciences*. 13(5), 260-275. doi: 10.33168/JSMS.2023.0517.