

Analysis of China's New Energy Public Opinion

—Based on Weibo Comments⁽¹⁾—

孔 令建

Changzhou Vocational Institute of Mechatronic Technology

Kong Lingjian

Kanagawa University

[Abstract] Public opinion now plays a pivotal role in shaping consumer purchase decisions. To gain insight into the prevailing sentiment among the Chinese population regarding new energy, we have undertaken the collection of textual data from China's largest online social platform, Weibo. We intend to employ Python for the analysis of these data, encompassing tasks such as word frequency analysis, sentiment analysis, and thematic analysis. The research revealed that words associated with automobiles, new energy technology, and the strength of enterprises appeared with notable frequency. It was also observed that the Chinese populace exhibits a slightly higher inclination towards positive sentiment in relation to new energy compared to negative sentiment. The primary areas of interest within the domain of new energy predominantly revolve around ten key topics, including innovation within new energy enterprises, the development of the automobile industry, and advancements in new energy technology. To foster greater acceptance and endorsement of new energy solutions among the Chinese public, it is advisable to implement strategies that bolster awareness and support. These may encompass intensifying publicity and promotional efforts, extending support for enterprise growth, and enhancing infrastructure development. These initiatives are essential to promote the sustainable growth of the new energy industry.

[Keywords] New energy, Weibo, TF-IDF analysis, Sentiment analysis, LDA

Table

1. Introduction
2. literature review
3. Data acquisition and preprocessing
4. Word frequency analysis

(1) Weibo is one of the largest social media platforms in China, often referred to as “China's Twitter”

- 5. Sentiment analysis
- 6. Topic analysis
- Conclusion

1. Introduction

In light of the escalating gravity of global warming and environmental concerns, the development and utilization of new energy sources have garnered widespread global attention. As the world's largest developing country, the development of new energy industry is of great significance to the country's economic growth, environmental protection and energy security. In recent years, the Chinese government has attached great importance to the new energy industry and introduced a series of policies and measures to promote the rapid development of the new energy industry. To a certain extent, consumers' demands for new energy products and services have also actively driven the improvement of industrial production efficiency and the upgrading of product iteration. In this context, Weibo, as an important social media platform in China, has become an important channel for the public to obtain information, express their views and participate in discussions. This study discusses the demand for new energy from the perspective of consumers (public opinion), which is of great significance for understanding the cognition and demand of the Chinese people for new energy, and the research results can provide reference for China's new energy development planning.

2. literature review

By reviewing relevant literature, the research status of China's new energy public opinion is sorted out. For example, Christopher and Ruth (2012) analyzed the comments of solar user on Twitter in the United States and found that consumers' perception and purchase intention of solar products are influenced by social networks. Chen et al., (2019) found through public opinion analysis that the positive role of attaching importance to infrastructure construction and promoting the application of new energy vehicles can promote the sales market of new energy vehicles and activate the automotive industry. Xing and Liu (2022) based on Weibo data found that most of the attention paid to China's dual carbon goals is focused on topics such as green economy and new energy development. Chen et al. (2021) discussed the development status of new energy vehicles and related issues under the dual effects of positive factors and disadvantages such as policy support and public opinion support.

Existing literature studies have demonstrated the extensive influence of public opinion on new energy, underscoring its constructive impact on the advancement of the new energy industry. However, there is little research on China's new energy public opinion analysis based on Weibo comments. This paper entails the retrieval of up-to-date public opinion data from Chinese consumers concerning new energy. It subsequently encompasses word frequency analysis, sentiment analysis, and thematic analysis. The overarching goal is to gain a deeper understanding of the Chinese pub-

lic's perceptions and requirements regarding new energy. Ultimately, the analysis results will be leveraged to offer recommendations for the further development of new energy in China.

3. Data acquisition and preprocessing

3.1 Data acquisition

Weibo stands as one of the most widely embraced new media platforms in China. According to official Weibo data for December 2022, there were a total of 586 million monthly active users, marking a net increase of approximately 13 million users on a year-on-year basis. Additionally, the platform boasted an average of 252 million daily active users, with a net increase of around 3 million users year-on-year.

For this study, the keyword “new energy” was employed as the primary search term. This enabled the retrieval of consumer comments related to new energy on Weibo. The data collection process was facilitated using Python's requests library, resulting in the accumulation of a comprehensive dataset comprising 272,404 new energy review entries spanning the timeframe between January 2021 and March 2023.

3.2 Data Processing

In order to ensure the accuracy and completeness of the data, Python is used to process the crawled online comment data, including missing data cleaning, data deduplication, mechanical compression and word removal, phrase filtering processing, and removal of meaningless information. After data processing, 132,253 valid comment data related to new energy were obtained.

4. Word frequency analysis

4.1 TF-IDF

Word frequency analysis refers to counting the number of occurrences of each word in the text, analyzing which words in the text are used more frequently and which words are used less frequently, so as to understand the main characteristics and key content of the text, and it is also a precursor to text mining and natural language processing. Compared with the word frequency analysis method based on the frequency of occurrence of words in the text, TF-IDF (Term Frequency-Inverse Document Frequency) is a more effective analysis method, commonly used in information retrieval and text mining, and can also be used to measure the importance of words in a sentence or article.

TF-IDF analysis consists of two parts, word frequency (TF) and inverse document frequency (IDF) (Hai et al.). TF stands for Word Frequency, which counts the number of times a word appears in the text, in general, the more times a word appears in a document, the more important the word is in the document. The word frequency formula is: $TF(t, d) = \text{number of occurrences of word } t \text{ in document } d / \text{total number of words in document } d$. IDF stands for Inverse Document Frequen-

cy, which takes into account the frequency at which a word appears throughout the text collection and is used to measure the universality of a word and the importance of a word in the entire document set. If a word appears in too much text, then its IDF value will be smaller, the inverse document frequency formula is: $IDF(t) = \log(\text{total number of documents} / \text{number of documents containing word } t)$, using the logarithmic function \log to reduce the effect of the number of occurrences of words in the overall document set.

Since the IDF value reflects the importance of a word, TF-IDF calculates the weight of a word by multiplying the TF value by the IDF value. Therefore, the TF-IDF score of some words that are highly frequent but not characteristic of the entire text collection will be suppressed. The TF-IDF formula is: $TF-IDF(t, d) = TF(t, d) \times IDF(t)$. A higher TF-IDF score indicates that the word is relatively important in this document.

4.2 Analysis of results

The TF-IDF analysis results for the processed data are presented in Table 1. These results reveal that when discussing new energy in Weibo comments, the most frequently occurring words related to automobiles include terms like “car”, “battery”, “charging”, and “model.” This observation underscores that, in the context of new energy, consumers place a significant emphasis on automobiles. Specifically, the public’s attention centers on factors such as the driving range of new energy vehicles, which facilitates long-distance travel. There is also considerable interest in the charging speed and distribution of charging stations for new energy vehicles. Consumers express a desire for fast-charging capabilities and an expanded network of charging infrastructure.

Secondly, keywords such as “technology”, “technique” and “intelligence” represent consumers’ technical requirements for new energy vehicles, indicating that Volkswagen hopes that new energy vehicles will have more intelligent functions, such as automatic driving, voice interaction, intelligent navigation, etc., so as to improve the convenience and safety of driving. Consumers are also concerned about the energy efficiency of new energy vehicles and hope that automakers will adopt more advanced energy-saving technologies to reduce vehicle energy consumption and polluting emissions.

Lastly, in the TF-IDF analysis, it’s notable that words associated with the scale and strength of enterprises, such as “enterprise”, “company”, and “plate”, are prevalent. This signifies that consumers’ interest in new energy enterprises primarily revolves around the robustness and capability of

Table 1. High-frequency words and TF-IDF values in Weibo text data

Ranking	keyword	TF-IDF value	Ranking	keyword	TF-IDF value	Ranking	keyword	TF-IDF value
1	Car	0.09349	6	Market	0.027362	11	Energy	0.022496
2	Technology	0.031253	7	Charge	0.025706	12	Industry	0.020265
3	Battery	0.030742	8	Industry	0.024985	13	Technique	0.018967
4	Project	0.028877	9	Firm	0.024494	14	Intelligent	0.018601
5	Enterprise	0.02798	10	Plate	0.023435	15	Models	0.017691

these enterprises. Consumers aspire for companies to deliver high-quality, dependable, and high-performance new energy products. Furthermore, there's a strong emphasis on after-sales service capacities and service quality. Consumers highly value receiving attentive and satisfactory after-sales support from these enterprises.

5. Sentiment analysis

5.1 Sentiment classification

Sentiment analysis is a natural language processing technique designed to automatically identify and extract mood, emotions, or emotional polarities expressed in text (Ji et al.2018). Sentiment analysis can help companies understand customer attitudes and feedback, help decision makers better understand public perceptions of a particular event or product, and guide marketing strategies.

This study uses the sentiment analysis module from Python's Snow NLP library for sentiment classification. First, the processed comment text data is pre-trained and a pre-trained model is generated, and then an emotional dictionary is generated based on the words in the training set. Since the corpus trained by the default model of Snow NLP is e-commerce consumer reviews, the model is trained by combining the emotional vocabulary dictionary of Dalian University of Technology. The Chinese emotion vocabulary ontology database of Dalian University of Technology explains the existing domestic dictionaries from various aspects, such as establishing Chinese ontology library resources from the perspectives of the part of speech, emotional polarity, polarity strength and other aspects of the words themselves.

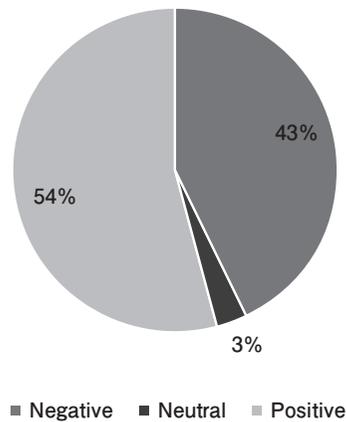
5.2 Analysis results

Based on the dataset's characteristics and the analysis results, we have established a sentiment scoring system. Sentiments scoring 0.6 and above are categorized as positive emotions, while

Table 2. Examples of sentiment analysis

Comment on the content	Sentiment score
New energy will replace real estate in the next decade	0.986
The rapid development of China's new energy vehicle industry has become a highlight of green consumption	0.98
The competition between new energy and fossil energy is an intricate one, making it challenging to predict a clear winner.	0.558
China strengthens new energy innovation	0.563
The new energy kinetic energy recovery model is too stupid to use at all	0.0075
How difficult it is for new energy vehicles to survive the winter	0.155
New energy is miserable	0.122
I'd like to express my dissatisfaction with the fact that the hardware of BYD Auto and BYD Tang's new energy vehicles does not align with the manufacturer's promotional claims.	0.0001

Figure 1 Proportion of the three types of emotions in the review



scores falling within the range of 0.6 to 0.4 are labeled as neutral emotions. Sentiments scoring 0.4 and below are categorized as negative emotions. To illustrate this categorization, Table 2 shows case examples of comments along with their corresponding emotional scores. Positive emotions encompass expressions of joy, happiness, love, satisfaction, and like. Negative emotions, on the other hand, encompass feelings of anger, sadness, disgust, disappointment, and so forth. Neutral sentiments are indicative of a lack of overt emotional inclination in the text, signifying neither clear positive nor negative emotions.

The analysis results showed that there were 71548 comments with positive emotional tendencies, 56345 comments with negative emotional tendencies, and 4360 comments with neutral emotional tendencies. The proportion of the three types of emotional categories is shown in Figure 1, the proportion of positive emotions, negative emotions and neutral emotions is 54%, 43% and 3%, respectively, and positive emotions are slightly higher than negative emotions, indicating that the number of people who have a positive attitude towards new energy in China is slightly higher.

6. Topic analysis

6.1 LDA Thematic Analysis

LDA (Latent Dirichlet Allocation) topic analysis is a probabilistic model for discovering the topic structure of text, the basic idea of which is to infer the topic distribution of each document as well as the word distribution of each topic by observing the word frequency distribution in the text (Wang & Li, 2019). LDA, as an unsupervised learning method, necessitates the pre-configuration of the number of topics. Experimentation with various topic quantities, result observation, and selection based on specific requirements is a common practice. In the context of the LDA topic model, perplexity is typically employed as an indicator to determine the optimal number of topics. The degree of perplexity can be understood as how uncertain the trained model is about which topic document D belongs to for an article D, and this degree of uncertainty is the degree of confusion, the

lower the degree of confusion, the better the effect of clustering.

6.2 Analysis Results

The comprehensive perplexity calculation of the cleansed Weibo data (Figure 2), combined with multiple attempts to analyze the results, sets the number of topics in this paper to 10.

Figure 2 Confusion calculation results

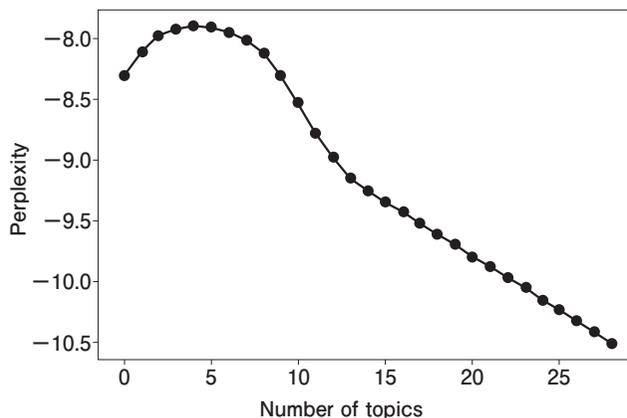


Table 3 Results of LDA thematic analysis

Subject name	Subject content (topic word frequency)					
Topic 1	Enterprise (0.022)	Industry (0.021)	Project (0.019)	New (0.017)	Innovation (0.014)	Economy (0.011)
Topic 2	Car (0.128)	New Energy (0.121)	link (0.036)	Sales (0.027)	Web page (0.023)	market (0.021)
Topic 3	Plate (0.033)	New Energy (0.022)	Rose (0.021)	Index (0.020)	Fund (0.017)	Individual stocks (0.016)
Topic 4	Technology (0.015)	Chip (0.013)	Enterprise (0.009)	Industry (0.009)	Market (0.008)	Time (0.008)
Topic 5	New energy (0.086)	Car (0.037)	Charge (0.037)	Cart (0.026)	Electric car (0.022)	Owners (0.017)
Topic 6	Battery (0.026)	Technology (0.018)	Intelligent (0.017)	Electricity (0.017)	Galaxy (0.016)	Products (0.016)
Topic 7	Vehicle (0.016)	Subsidy (0.011)	Car (0.007)	Inhale (0.007)	Public transit (0.007)	Epidemic situation (0.007)
Topic 8	Firm (0.052)	Share (0.029)	Increase (0.025)	New energy (0.019)	Technology (0.018)	Battery (0.017)
Topic 9	Energy (0.035)	Carbon (0.027)	New energy (0.019)	Green (0.015)	Country (0.013)	Nudge (0.011)
Topic10	Market (0.032)	Funds (0.020)	Industry (0.020)	Tap (0.013)	High (0.011)	Concern (0.010)

The outcomes of the LDA thematic analysis, conducted using the Python library Gensim, are depicted in Table 3. Within this analysis, Topic 1 is closely associated with new energy enterprises, highlighting consumer interest in enterprise innovation within the new energy sector, as well as the prominence of new energy projects as key focal points. Theme 2 is related to the development of the new energy vehicle industry, and sales volume, market demand, etc., which are hot spots for

consumers. Theme 3 is related to the stock market of new energy enterprises, reflecting consumers' concern for the future development of new energy enterprises. Theme 4 is related to new energy technology, battery and chip technology are particularly concerned; Theme 5 is related to consumers' use of new energy vehicles, and the convenience of charging is concerned; Theme 6 is related to new energy battery technology, and the storage technology of electric energy is a hot spot for consumers. Theme 7 is related to transportation, and consumers are concerned about the popularity of new energy vehicles. Theme 8 is related to scientific and technological innovation of new energy enterprises, especially battery technology. Theme 9 is related to the green and low-carbon development of the country, consumers pay attention to the green development brought by new energy; Theme 10 is related to the market, consumers pay attention to the future development of the new energy market.

Conclusion

Consumers use the Internet as a carrier to publish a large amount of information on new energy cognition, demand, evaluation and other aspects, these new energy review data reflect consumers' concerns, expectations and opinions on new energy, and analyzing these text data can guide the healthy development of new energy market and industry. By collecting 132,253 valid text data from Weibo with "new energy" as the keyword, this study finds through TF-IDF analysis that the frequency of words related to automobiles, new energy technology, and enterprise scale is relatively high; the positive sentiment of Chinese people towards new energy is slightly higher than the negative sentiment; the public's attention to new energy mainly focuses on new energy enterprise innovation. There are 10 aspects, including automobile industry development, new energy enterprise stock market, new energy technology, new energy vehicle use, new energy battery technology, transportation, scientific and technological innovation of new energy enterprises, national green and low-carbon development, and market development, focusing on the automotive field.

Based on the above analysis results, it is recommended that: First, Strengthen publicity and promotion, and increase publicity on new energy technology and new energy vehicle industry through social media platforms such as Weibo. Improve the public's awareness and understanding of new energy, and enhance their positive attitude towards new energy. Second, support the development of enterprises, encourage and support the development of new energy enterprises, provide policy support and financial support, and promote the expansion of enterprise scale and technological innovation. Then, improve infrastructure construction, accelerate the construction of new energy vehicle charging facilities, improve the coverage and convenience of charging facilities, increase the transformation and upgrading of transportation infrastructure, and promote the use and development of new energy vehicles. Third, promote green and low-carbon development, increase the development and utilization of renewable energy, and reduce dependence on traditional fossil energy. Finally, strengthen scientific research and innovation, increase support for new energy technology research and development, encourage scientific research institutions and enterprises to strengthen

cooperation, train more professionals in the field of new energy, and promote the development and application of new energy industry through scientific and technological innovation.

● References

- [1] Christopher, M. J., & Ruth, L. B. The impact of social networks on consumer adoption of renewable energy: A comparative study of Facebook and Twitter [J]. *Renewable and Sustainable Energy Reviews*, 2012, 37(6), pp. 594–604.
- [2] Chen Q., Deng H. Y., & Zhang R. S., New energy vehicle policy and situation analysis based on LDA thematic model [J]. *Journal of Guangzhou University (Natural Science Edition)*, 2019, 18(5), pp. 34–42.
- [3] Hao H. J., Zhang K. P., Wang W., & Gao G. A Tale of Two Countries: International Comparison of Online Doctor Reviews Between China and the United States [J]. *International Journal of Medical Informatics*, 2017, 99, pp. 129–134.
- [4] Ji P., Zhang H. Y., & Wang J. Q. A Fuzzy Decision Support Model with Sentiment Analysis for Items Comparison in e-Commerce: The Case Study of PConline. com [J]. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 2018, pp. 1–12.
- [5] Xing Y. C., & Liu S. Y. Topic Discovery and Emotion Research of “Double Carbon Goals” Short Text Based on Tabular Emotion Analysis Mathematics in Practice and Theory, 2022, 52(8), pp. 164–174.
- [6] Chen Z., Zhang K. H., & Jia S. W. Influence Factors and Risk Analysis of New Energy Vehicles from The Perspective of System. *Environmental Science and Pollution Research*. <https://doi.org/10.1007/s11356-021-15232-3>, 2021
- [7] Wang T., & Li M. Research on Comment Text Mining Based on LDA Model and Semantic Network, *Chongqing Technol & Business Univ (Nat Sci Ed.)*, 2019, 36(4), pp. 9–16.