Analysis on the Influencing Elements and Optimization Countermeasures of New Energy Vehicles’ Consumer Satisfaction——Analysis Based on Internet Online Review Mining

Suyang Wang¹, Qin Liu ²*
¹ School of Management, Wuhan University of Technology, Wuhan, Hubei, China
² School of Entrepreneurship, Wuhan University of Technology, Wuhan, Hubei, China
*Corresponding author’s e-mail: isunnygirl@qq.com

Abstract: With the rapid development of the Internet, more and more consumers are used to posting their own purchase reviews on online platforms, and automotive consumers are no exception. New energy vehicles have been regarded as one of the methods to improve energy and environmental issues, and have received attention from countries around the world. The Chinese government has also vigorously supported the new energy vehicle industry. However, in the process of green innovation, new energy vehicle manufacturers do not understand consumers' satisfaction with new energy vehicles. Therefore, in this paper, the online review of new energy vehicles is used as the research object, and the text mining method is used to analyse and research the online reviews of new energy vehicles to provide effective suggestions for new energy vehicle manufacturers in terms of green innovation.

1. Introduction
With the rapid development of the Internet, data information is also growing in an explosive manner. More and more scholars have begun to study data information on the Internet, trying to transform disordered information into orderly information from cluttered data information through some algorithms. Text mining is a method of data mining. At present, the development of text mining in foreign countries is relatively mature. Maron (1960) [1] published the first article on automatic classification. Rostek (2015) [2] believed that text mining refers to analysing and predicting texts through statistical modeling and other means, and obtaining information in the text. The most commonly used text mining methods include cluster analysis and sentiment analysis. Hobbs et al. (1982) [3] Task text mining was first applied in life science research and government intelligence research. At present, some related scholars in China also use text mining for research and learning. Hu Jing (2009) [4] designed a web page effective information extraction method based on the characteristics of web pages. Lin Jie and Miao Runsheng (2020) [5] realized the topic map by using Skip-Gram text training model and LDA topic model to extract topics and keywords from professional social media text. Park et al. (2009) [6] believe that consumers can post positive or negative evaluations of goods through e-commerce platforms. Gu et al. (2013) [7] found that the management response had a significant effect on customers with low satisfaction through text mining of online reviews on Ctrip's hotel pages.

Content from this work may be used under the terms of the Creative Commons Attribution 3.0 licence. Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.
Published under licence by IOP Publishing Ltd
Chunhua (2012) [8] proposed an extraction method based on product feature words and sentiment words to analyse sentiment of online reviews and analyse the sentiment orientation of online reviews.

Through the collation and analysis of domestic and foreign texts, most scholars use text mining to conduct research on online reviews and consumer satisfaction and have achieved many results. However, most of the scholars’ research objects are focused on Weibo public opinion monitoring, hotel industry, and tourism industry. Few people notice new energy vehicles’ online reviews. Therefore, this article takes new energy vehicle online reviews as the research object, and uses machine learning to study the new energy vehicle consumers’ satisfaction elements. Based on these elements, it proposes constructive significance for new energy vehicle manufacturers in green innovation.

2. Research framework and related technology analysis

2.1. Research Framework

The research on online reviews of new energy in this research is divided into the following steps: Collect new energy vehicle industry vertical domain websites, such as the online reviews of new energy vehicles, such as Autohome, Tencent Automotive, Pacific Automotive Network, etc. Data pre-processing, text mining and sentiment classification for new energy vehicles’ online reviews, and mining elements which affected consumer satisfaction of new energy vehicles’ online reviews. As shown in Figure 1 below.

Figure 1. Research Framework

2.2. Related technical analysis

2.2.1. TF-IDF algorithm. This paper uses the TF-IDF weighting algorithm for the weight calculation of online review features of new energy vehicles. TF is defined as word frequency. TF can be expressed by formula (1):

$$tf_{ij}=\frac{n_{ij}}{\sum_k n_{i,j}}$$  \hspace{1cm} (1)

The number of times the word appears in the file is represented by $n_{i,j}$, and the denominator represents the cumulative sum of the number of times the word appears in all files. IDF is defined as the inverse text frequency, indicating how often a word is used throughout the document. IDF can be expressed by equation (2):

$$idf_{i}=\log(\frac{N}{1+n_i})$$  \hspace{1cm} (2)

The total number of file texts is $N$, and the total number of files containing specific words is $n_i$. A specific word’s weight in a text file can be obtained by TF and IDF. $W_i$ represents the weight of a specific word in the text file set, which can be expressed by formula (3):

$$W_i= tf_{i,j}* idf_i$$  \hspace{1cm} (3)

2.2.2. Grey relational analysis. The grey correlation analysis method is used to analyze the new energy vehicle consumer satisfaction’s influencing elements in this study, and the weights of the influencing
elements of new energy vehicle consumer satisfaction are determined and ranked according to the weight. The specific calculation steps of gray correlation analysis are shown below.

1. Determine the reference and comparison columns. The reference sequence in this article refers to the new energy vehicle online review’s score after sentiment polarity analysis. The comparison series in this article refers to the scores after the emotional polarity analysis of each factor in the online review of new energy vehicles.

2. Dimensionless processing of the system's reference and comparison columns.

3. Find the gray correlation coefficient of the reference and comparison columns. The degree of correlation essentially refers to the degree of difference in geometric shapes between the curves. Therefore, the gray correlation coefficient can be expressed by equation (4), where Δmin refers to the second-stage minimum difference, Δmax refers to the two-stage maximum difference, and ρ is the resolution coefficient, which is generally between 0 and 1, usually 0.5

\[ \zeta_{oi} = (\Delta(\min) + \rho\Delta(\max)) / (\Delta_{oi}(k) + \rho\Delta(\max)) \]  

4. Find the degree of correlation \( r_i \).

\[ r_i = \frac{1/n}{\sum k \text{ for } i} \]  

3. Data acquisition and collation

The data source of this article is an online review of different brands of different energy vehicles of the Auto House, including BYD, Audi, Volvo, Weilai and other brands. This paper uses the Houyi collector to collect, received a total of 3,815 online reviews of new energy vehicles. Because the original data contains some invalid data, the data needs to be cleaned, and the review data irrelevant to the topic and duplicate online reviews are finally obtained 3367 reviews.

After the data is cleaned, the data needs to be segmented. Because Chinese text and English text are very different, we need to use the word segmentation tool to segment the online reviews of new energy vehicles. Jieba word segmentation package is used to segment the online reviews of new energy vehicles in this article, because jieba segmentation can achieve efficient word map scanning based on the prefix dictionary, and cut the sentence most accurately.

To use python for Chinese word segmentation, this research first need to load the jieba word segmentation package, then remove the numbers and English from online reviews, add words that are irrelevant to this research and have no practical meaning into the stop word library, and remove stop words from online reviews. The word segmentation results are shown in Table 1 below.

| number | Text after segmentation |
|--------|------------------------|
| 1      | Originally look on handsome fuel consumption start with late out hybrid…… |
| 2      | Long-distance cars no longer comfortable do not lose the level of cars heavy cars power trunk space small…… |

4. Analysis of elements affecting new energy vehicle’ consumer satisfaction

4.1. New energy vehicle online review sentiment analysis

This article analyses the sentiment of online reviews of new energy vehicles based on the sentiment dictionary, using the Taiwan University sentiment dictionary (NTUSD) as the basic dictionary. Taiwan University Affective Dictionary (NTUSD) is currently the most authoritative sentiment dictionary in domestic emotional dictionaries and is widely used. At the same time, denial words can change the emotional tendency of a sentence, so denial words have a significant impact in the new energy vehicle online review sentiment analysis. Based on the sentiment dictionary of Taiwan University and combining the characteristics of online reviews of new energy vehicles, this study created a proprietary sentiment dictionary to divide the sentiment dictionary into positive sentiment dictionary and negative
sentiment dictionary. And according to the degree adverbs, denial words, emotional words to the new energy vehicle online reviews assigned, the assignment table is shown in Table 2 below.

| Degree word | Score | Denial Words | Score | Positive words | Score | Negative words | Score |
|-------------|-------|--------------|-------|----------------|-------|----------------|-------|
| extreme     | 4     | not          | -1    | good           | 1     | dislike        | -1    |
| very        | 3     | no           | -1    | credible       | 1     | dissatisfied   | -1    |
| more        | 2     | don’t have   | -1    | safety         | 1     | disappointed   | -1    |
| little      | 0.5   | never        | -1    | like           | 1     | bad            | -1    |

This article uses python to program according to the above sentiment scoring method and gets sentiment orientation and sentiment scores for each online review. Some results are shown in Table 3 below.

| New Energy Vehicles Online Review | Sentiment score | Affective tendency |
|-----------------------------------|-----------------|--------------------|
| It's very easy to drive long distances. The electronic assist function is what I need. The sound effect is very good. Even chasing dramas in the car can make long-distance cars no longer boring…….. | 19.5 | positive |
| I don’t have started because the fuel consumption is a bit high, and now there are hybrids that can solve this biggest problem. The trunk is small, the capacity is not enough, and the new car has a very bad smell, which is a little bad…….. | -1.5 | Negative |

Through the sentiment analysis of new energy vehicle online reviews, this article divides sentiment tendencies into positive, general and bad reviews, and counts all 3367 online reviews. The results are shown in Table 4 below.

| Affective tendency | Number of online reviews | Percentage |
|--------------------|--------------------------|------------|
| Positive reviews   | 2076                     | 61.76%     |
| General reviews    | 201                      | 5.95%      |
| Bad review         | 1090                     | 32.29%     |

According to the above statistical results, it is known that among the 3376 new online reviews of new energy vehicles, they have the most positive positions. This indicates that most consumers are relatively satisfied with new energy vehicles, but 32.29% of consumers are still not satisfied with new energy vehicles. Therefore, new energy vehicle manufacturers need to continue to improve the quality of new energy vehicles and improve the consumer experience.

4.2. Elements affecting consumer satisfaction

After using python to program a series of pre-processing such as word segmentation, sentence segmentation, and removal of stop words for new energy automobile online reviews, high-frequency words are filtered out and drawn into a word cloud. The results are shown in Figure 2. In Figure 2, space, power, appearance, power consumption, energy, and comfort are the most words that consumers mentioned. This can further show that consumers' focus on new energy vehicles is in these aspects. In order to further clarify which aspects of new energy vehicles have an impact on consumer satisfaction, this article uses the TF-IDF algorithm to perform a weight analysis on the review text to obtain a vector model of new energy vehicle online review feature word weights. The 30 high-frequency nouns in Figure 2 are selected as feature vectors, and the hierarchical clustering results are shown in Figure 3. Therefore, the elements affecting consumer satisfaction of new energy vehicles are set as follows: space, power, handling, energy consumption, interior decoration, appearance, cost performance, and comfort.
This article uses python to implement the calculation of grey correlation analysis, using the sentiment scores of online reviews of new energy vehicles as a reference column. This paper uses the segmented data as the initial data to select the words associated with the 8 elements, and used the new energy vehicle online review sentiment analysis algorithm to calculate the sentiment value of each influencing factor. The sentiment value of 8 elements is used as the comparison column, and then the reference and comparison columns are processed and brought into the grey correlation coefficient formula for calculation. Through analysis, the relationship between the elements affecting consumer satisfaction of new energy vehicles is obtained. The results are shown in Figure 4.

### Table 1. Grey Correlation Analysis Results

|                | Satisfaction | Energy | Appearance | Power | Space | Interior | Cost Performance | Comfort | Manipulation |
|----------------|--------------|--------|------------|-------|-------|----------|------------------|---------|-------------|
| Satisfaction   | 1            |        |            |       |       |          |                  |         |             |
| Energy         | 0.99431      | 1      |            |       |       |          |                  |         |             |
| Appearance     | 0.53959      | 0.53416| 1          |       |       |          |                  |         |             |
| Power          | 0.67862      | 0.6796 | 0.71966    | 1     |       |          |                  |         |             |
| Space          | 0.90391      | 0.90214| 0.50641    | 0.63303| 1     |          |                  |         |             |
| Interior       | 0.62389      | 0.62477| 0.78529    | 0.86440| 0.61098| 1        |                  |         |             |
| Cost Performance| 0.99618      | 0.99258| 0.53313    | 0.67777| 0.91358| 0.64756  | 1                |         |             |
| Comfort        | 0.54282      | 0.54341| 0.96342    | 0.75001| 0.53985| 0.82121  | 0.55136         | 1       |             |
| Manipulation   | 0.68195      | 0.68303| 0.70990    | 0.98660| 0.66055| 0.88899  | 0.68962         | 0.72632| 1           |

### Figure 4. Grey Correlation Analysis Results

After conducting grey correlation analysis on 8 elements that affect the satisfaction of new energy vehicles, we can get the correlation ranking of the elements that affect the satisfaction of new energy vehicles as follows: cost performance > energy consumption > space > manipulation > power > interior > comfort > appearance. Among them, the impact of cost performance, energy consumption, and space on the satisfaction of new energy vehicles is more significant than the other five items, which indicates that when buying new energy vehicles, consumers first value the cost performance of a car. Second, consumers compare the energy consumption of re-energy vehicles. However, the two factors of comfort and appearance have the least impact on the satisfaction of new energy consumers.

### 5. Conclusion and suggestion

This study combines machine learning and python to analyse the word of mouth of new energy vehicles, identify the elements of new energy vehicle consumer satisfaction and evaluate the elements. This article first uses python to realize the sentiment analysis of new energy vehicles. According to the sentiment analysis results of 3367 word of mouth, most consumers are satisfied with new energy vehicles, but there are still about 32% of consumers who are not new energy vehicles very satisfied. In order to
further explore the elements that affect consumers' satisfaction with new energy vehicles, the TF-IDF algorithm is used to analyse the reputation of pre-processed new energy vehicles and select the top 30 nouns in the text of new energy vehicles as features. This paper performed hierarchical clustering analysis on these 30 feature words. This article summarized 8 influencing factors of new energy vehicle consumer satisfaction, and conducted an emotional analysis of the 8 influencing elements. The grey relational analysis was used to rank the elements of new energy vehicle consumer satisfaction. The ranking results show that new energy vehicle consumers pay more attention to the cost performance, energy consumption and space of new energy vehicles. Therefore, new energy vehicle manufacturers need to consider the cost-effectiveness of new energy vehicles after manufacturing, and the life and power consumption of rechargeable batteries for new energy vehicles.

Acknowledgments
This work is supported by the National Social Science Foundation of China (No. 19BSH105).

References
[1] Maron, M., et al. On Relevance Probabilistic Indexing and Information Retrieval [J]. Journal of the ACM, 1960(7):216-244.
[2] Rostek K. New Approaches in Supporting to SMEs Competitiveness [M]. Benchmarking Collaborative Networks. Springer International Publishing, 2015.
[3] Hobbs J R, Walker D E, Amsler R A. Natural Language Access to Structured Text [C]. Conference on Computational Linguistics, 1982.
[4] Hu Jing. Research and implementation of Web-based Chinese text mining technology [D]. Central South University, 2009.
[5] Lin Jie, Miao Runsheng. Research on the Construction of Theme Atlas in Professional Social Media—Taking Auto Forum as an Example
[6] Park C, Lee T M. Information direction, website reputation and e WOM effect: A moderating role of product type [J]. Journal of Business Research, 2009, 62(1):61-67.
[7] Gu B., Ye Q. First Step in Social Media: Measuring the Influence of Online Management Responses on Customer Satisfaction [J]. Production and Operations Management Society, 2013.
[8] Sun Chunhua. Research on the Influence of Emotional Expression on the Perceptual Usefulness of Online Reviews [D]. Hefei University of Technology, 2012.