

# How does consumer demand for new energy vehicles drive innovation? – based on social network and text mining analysis

**Jun Huang**

Southwest University

**chen cao** (✉ [1281802269@qq.com](mailto:1281802269@qq.com))

Southwest University

**Hancheng Huang**

University College London

**Zihan Xiao**

University of Ottawa

**Yunfan Huang**

University of Bristol

---

## Research Article

### Keywords:

**Posted Date:** March 14th, 2022

**DOI:** <https://doi.org/10.21203/rs.3.rs-1417400/v1>

**License:**   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

---

# How does consumer demand for new energy vehicles drive innovation? -- based on social network and text mining analysis

Jun Huang <sup>1</sup>, Chen Cao <sup>2</sup>✉, Hancheng Huang <sup>3</sup>✉, Zihan Xiao <sup>4</sup>, Yunfan Huang <sup>5</sup>

Market demand is the internal source of technological innovation. Automobile enterprises get various information about product improvement from consumers, and consumer demand points out the direction of technological innovation for new energy automobile enterprises. First, Python was used to write a crawler to crawl the text data of consumer comments on new energy vehicles on Autohome website. Then, big data analysis is carried out based on social network, word cloud map, LDA topic model and other methods. Finally, the empirical study found that commuting, picking up children, appearance, space, appearance level, and interior are among the key factors affecting ordinary consumers to buy new energy vehicles; consumers are not satisfied with the new energy vehicles mainly for the trunk, abnormal sound, endurance, noise, and interior. Among the 8 themes of the LDA theme model, theme 6 (noise problem), theme 1 (brake problem) and theme 3 (odor problem) show high theme intensity, and mass consumers are increasingly dissatisfied with the low financial incentives and high prices of new energy vehicles. Finally, based on the findings of the above influencing factors, relevant management suggestions on product innovation are provided for new energy vehicle enterprises.

## Introduction

The development of new energy vehicles is an important way to transform Chinese automobile industry, as well as a strategic measure to cope with climate change and promote green development. In the past few decades, China has become a major global automobile consumer. For China the emerging of new energy vehicles will provide a new opportunity for the automobile industry to reshuffle. Seizing the opportunity to change, China may overtake developed countries in the automobile industry and compete with them fairly to emerge as an automobile power. In addition, as global warming, energy crisis and other issues become increasingly serious, environmental problems have attracted more and more attention, and the public has begun to pay more and more attention to the development of new energy vehicles <sup>[1]</sup>. New energy vehicles are regarded as clean products and have been recognized worldwide <sup>[2-3]</sup>. In terms of production and sales, China has become the world's largest market for new energy vehicles <sup>[4]</sup>. Since 2009, the total volume of automobile production in China has surpassed that of the United States and become the world's largest automobile market <sup>[5]</sup>. However, the excessive oil consumption has also triggered social concerns on energy and environmental issues <sup>[6-8]</sup>.

As new energy vehicles are considered to be an effective way to solve environmental problems and energy crisis, sales of new energy vehicles have been increasing in recent years. At the end of December 2018, sales of new energy vehicles increased by more than 18 times compared with 2014. On October 20, 2020, The Office of the State

1 College of Economics and Management, Southwest University, Chongqing 400715, China; john0277@swu.edu.cn

2 College of Economics and Management, Southwest University, Chongqing 400715, China; 1281802269@qq.com

3 Department of physics and astronomy, University College London, London WC1E 6BT, UK; zcapuab@ucl.ac.uk

4 Department of Mathematics and Statistics, University of Ottawa, Ottawa K1N 6N5, Canada; zxiao054@uottawa.ca

5 College of Education, University of Bristol, Queens Road, Bristol, BS8 1QU, UK. xs20823@bristol.ac.uk

✉ Correspondence 1: College of Economics and Management, Southwest University, Chongqing 400715, China; 1281802269@qq.com; (86)17831122536;

✉ Correspondence 2: Department of physics and astronomy, University College London, London WC1E 6BT, UK; zcapuab@ucl.ac.uk; (44)0755-6716282

Council of China issued the Development Plan for New Energy Automobile Industry (2021-2035), which clearly points out that it is necessary to strengthen standard docking and data sharing, closely follow the pulse of the era of "deepening big data sharing, leading high-quality development", and constantly explore breakthroughs in the high-quality development of big data of new energy vehicles for the future of the automobile industry. Therefore, in order to understand consumers' demand for new energy vehicles, it is necessary to collect and identify the big data that can affect consumers' purchase of new energy vehicles. According to this big data information, the government can implement more effective policies to promote the development of the new energy vehicle industry.

Most of the existing research believes that consumer demand points out the direction of technological innovation for enterprises. In order to meet the diversified consumer demand, enterprises will increase R&D investment and carry out relevant technological innovation. Market demand is the internal source of technological innovation, and the classic industrial organization theory explains the demand-driven innovation theory as follows: innovation is a function of market demand, consumer demand will improve the R&D investment of enterprises, and innovation for the pursuit of market profit is the main cause of endogenous technological change. Schmookler proposed the "demand-driven hypothesis" that technological innovation is the result of enterprises' pursuit of new technologies in the fierce market competition to obtain final profits for enterprises<sup>[9]</sup>. Market demand usually leads to the emergence of new technologies, and market demand is the fundamental driving force for technological innovation of enterprises. Schmookler's demand-driven hypothesis suggests that expenditure technological innovation is a function of market demand, and demand drives technological innovation of enterprises. Therefore, by understanding the diverse needs of consumers, enterprises can obtain the direction and information of product improvement and technological innovation from consumers, and finally develop and consolidate new markets through differentiated products. Therefore, enterprises can also carry out more effective product innovation based on customer orientation. Most studies on the demands of ordinary consumers are based on questionnaire surveys or experimental studies, but these methods are susceptible to the influence of observers and subjects or time-space limitations<sup>[10-11]</sup>. Therefore, non-interventional online content analysis method is adopted in this study, which is based on big data analysis methods such as social network analysis and text mining method to gain insight into the consumption process of consumers from the perspective of their needs<sup>[12]</sup>. In addition, big data text analysis results provide product innovation related management suggestions for Chinese new energy vehicle enterprises.

## 1. Literature Review

In recent years, research in the field of new energy vehicles has attracted wide attention, and many scholars have studied the factors influencing consumers' purchase of new energy vehicles<sup>[13]</sup>. Scholars outside of China have found that the main factors affecting consumers' purchase of new energy vehicles include: energy consumption<sup>[14]</sup>, energy price<sup>[15]</sup>, number of energy supply stations<sup>[16]</sup>, national subsidies<sup>[17]</sup>, national policies<sup>[18]</sup>, carbon dioxide emissions<sup>[19]</sup>, etc. In addition, some individual micro factors are also founded<sup>[20]</sup>, such as driving skills<sup>[21]</sup>, personal needs<sup>[22]</sup>, personal behavior<sup>[23]</sup>, psychological needs<sup>[24]</sup> and personal characteristics<sup>[25]</sup>. Chinese scholars have also studied the main factors affecting consumers' purchase of new energy vehicles, mostly using questionnaire data to analyze the factors affecting consumers' purchase of new energy vehicles. Among them, Wang Yuehui and others conducted a questionnaire survey on Beijing citizens. Regression analysis and factor analysis were used to find that buying attitude, subjective norms and perceived behavioral control are the key factors affecting consumers' purchase of new energy vehicles<sup>[26]</sup>. Xu Guohu et al. conducted a questionnaire analysis on customers in automobile sales service shop. Their findings revealed that after-sales service, purchase cost, automobile quality, use of energy consumption and the surrounding influence are the main factors that influence consumers' purchase of new energy vehicles<sup>[27]</sup>. By investigating the private car owners via internet, Wang Ning found that age, education level and the average family's annual income are the main factors influencing consumers to purchase new energy vehicles<sup>[28]</sup>. Through SOR theory analysis, Li Chuang et al. found that charging policy had the greatest influence on consumers'

purchase of new energy vehicles<sup>[29]</sup>. Through analysis of survey data, Liu Xueyuan et al. found that consumers have greater enthusiasm to buy new energy vehicles because of their energy saving and contribution to environmental protection<sup>[30]</sup>. Sun Xiaohua et al. found that price is the main factor affecting consumers' purchase of new energy vehicles<sup>[31]</sup>.

Literature review found that most of the existing research is carried out by questionnaire survey. However, in the traditional method of manual extraction of questionnaire survey, the sample selection is easily affected by subjective factors, and due to the limitation of workload, the sample size is also small. In recent years, more and more research uses the technology of big data and text mining to analyze the massive consumer comment data, through the big data analysis of the online behavior data of a large number of consumers, so as to obtain some findings that are difficult to obtain by questionnaire survey or experimental research. In addition, through big data analysis of online behavior data of consumers of new energy vehicles, factors affecting consumers' purchasing decisions can be comprehensively observed from the perspective of consumers. At the same time, big data analysis can overcome some limitations of questionnaire survey or experimental research in terms of time and space attributes<sup>[32]</sup>. Based on the text mining technology of big data, it is possible to find potentially useful information of consumers or useful inspiration for new product innovation from massive user comment texts<sup>[33]</sup>, or find some information that may be ignored by manual analysis<sup>[34]</sup>. Moreover, the selection bias can be avoided because relatively complete samples are collected directly from the network. Therefore, on the basis of ensuring the objectivity and comprehensiveness of sample selection, this paper, combined with text mining technology, explores the factors influencing consumers' purchase of new energy vehicles by using the big data of Chinese consumers' online behavior, and excavates the main factors of consumers' dissatisfaction in the process of purchasing new energy vehicles. On the one hand, it can provide new energy vehicle enterprises with information on relevant research & development. On the other hand, it can provide relevant investment companies and government departments with market development trend, industrial technology innovation and other decision-making basis.

## 2. Research Design

### 2.1. Data sources and processing

In this paper, Python is used to write a crawler to crawl the text data of comments about new energy vehicles on Autohome website. Autohome is the largest automobile website in China and has strict regulations on user comment data. Therefore, the reserved comment data are relatively real data with high data quality. The crawl time range ends on May 12, 2021. A total of 27,064 pieces of data were climbed. Due to the small amount of data in 2012 and 2013, 27,054 pieces of data were retained after deleting the data in these two years, among which each piece of text data was one line, including "time", "most satisfied", "least satisfied", "space", "power", "control", "energy consumption", "comfort", "appearance", "interior", "cost performance", "why choose this car" and other 12 fields. Then, the text data is processed as follows: (1) The text data is read by using pandas, and then by using jieba library to perform word segmentation, custom dictionary and stop word filtering for the 11 fields except "time" in each text data. (2) The Python program was used to traverse the whole database to obtain the keyword co-occurrence matrix of the "Why do I choose this car" field comment text, and the Ucinet software was used to conduct the social network analysis of the "why do I choose this car" field comment text. (3) According to the "time" field, the content of the "least satisfied" field was summarized from 2014 to 2021 year by year, and jieba and WordCloud library were used to generate the WordCloud map of each year by year. (4) Gensim and Python were used to conduct LDA topic model analysis on the text of the "most unsatisfactory" field.

### 2.2. Research methods and framework

To provide comprehensive insight into the new energy automobile consumer demand, and to explore the

direction of the "subjunctive-mood" product improvement information and technological innovation on the basis of in-depth analysis of consumer demand, this article uses the Ucinet software to build text field "why choose this car" social network diagram and analyse new energy automobile main impacting factors behind consumers' decisions on purchase of new energy vehicles. Then, Python and Jieba libraries are used to draw the word cloud topic evolution diagram of the text data of the "least satisfied" field of ordinary consumers to study the change of the least satisfied aspect of ordinary consumers for new energy vehicles over time. Finally, Python and Gensim package are used to analyze the LDA topic model of the "least satisfied" dictionary text data and study the topic division and topic intensity evolution of the most dissatisfied text data.

### 3. Empirical Analysis

#### 3.1. Social network analysis of "Why choose this car" field

In this paper, Ucinet and Python are used for social network analysis. Keywords in the "Why do I choose this car" field are used as nodes, and the number of co-occurrence of keywords is used as the weight between nodes. The more co-occurrence of two keywords, the closer the relationship between the two keywords is. Among them, the social network analysis results of the comment text of "Why choose this car" field are shown in Figure 3.

##### 3.1.1. Point centrality

The calculation formula of absolute dot read centrality is as follows:

$$C_D(n_i) = d(n_i) = \sum_j X_{ij} = \sum_j X_{ji} \quad (1)$$

Where,  $n$  is the total number of nodes,  $X_{ij} = 0$  (when node  $n_i$  is not directly adjacent to  $n_j$ ),  $X_{ij} = 1$

(when node  $n_i$  is directly adjacent to  $n_j$ )

Calculation formula of relative point centrality:

$$C_D(n_i) = \frac{d(n_i)}{(N-1)} \quad (2)$$

Where  $n$  is the total number of nodes,  $n \geq 2$ ,  $d(n_i)$  is the absolute centrality of node  $n_i$ .

Point degree centrality is a key index to reveal the importance of nodes in social networks. If a node is at the center of the network, its point centrality value is the highest. Table 1 shows the ranking of node centrality in the social network field "Why choose this car", where Degree represents absolute centrality, NrmDegree represents standardized centrality, and Share represents the ratio of centrality of each node to total centrality. According to the value of Degree, "new energy" node ranks first, indicating that the keyword "new energy" is in the core position in the network. Nodes such as "appearance", "model", "cost performance ratio", "space" and "price" rank higher, indicating that these keywords are more important in the network and are the main factors for consumers to consider when buying new energy vehicles.

##### 3.1.2. Social network graph of "Why choose this car"

In Figure 1, the node size represents the importance of the node. The larger the node, the more important the keyword. The thicker the line between the two nodes, the closer the connection between the two keywords. As can be seen from Figure 1, nodes such as "new energy", "appearance", "cost performance", "appearance level" and "space" are larger, indicating that these keywords are more important in the network. According to the connection between the keywords can be found in the social network, the keyword nodes associated with the "new energy"



**Table 1.** Point centrality score

keywords	Degree	NrmDegree	Share	keywords	Degree	NrmDegree	Share
New energy	62286	6.915	0.022	The budget	16342	1.814	0.006
appearance	48980	5.438	0.018	A friend	15515	1.722	0.006
byd	34893	3.874	0.013	To save money	14426	1.602	0.005
configuration	32166	3.571	0.012	good-looking	14264	1.584	0.005
Cost performance	30903	3.431	0.011	The battery	13974	1.551	0.005
life	30111	3.343	0.011	Suitable for	13691	1.52	0.005
space	29538	3.279	0.011	reason	13675	1.518	0.005
The price	28992	3.219	0.01	Conform to the	13293	1.476	0.005
consider	28084	3.118	0.01	The tram	13255	1.472	0.005
A test drive	27485	3.051	0.01	price	13167	1.462	0.005
models	26368	2.927	0.009	travel	12092	1.342	0.004
brand	23386	2.596	0.008	driving	11909	1.322	0.004
power	23004	2.554	0.008	hybrid	11896	1.321	0.004
To and from work	22974	2.55	0.008	Fuel consumption	11865	1.317	0.004
The car	22069	2.45	0.008	The cost of	11424	1.268	0.004
fuel	21306	2.365	0.008	To attract	11097	1.232	0.004
interior	20315	2.255	0.007	meet	11069	1.229	0.004
charging	18948	2.104	0.007	The mileage	10814	1.201	0.004
Level of appearance	18785	2.085	0.007	experience	10773	1.196	0.004
Instead of walking	16971	1.884	0.006	performance	10436	1.159	0.004

### 3.2. Analysis on the topic evolution of the word cloud of the "most unsatisfactory" field

This paper firstly conducts word segmentation for the text data of the "most unsatisfactory" field, extracts keywords from the text according to the principle of  $TF - IDF$ .

$TF - IDF$  formula is as follows:

$$IDF(x) = \log \frac{N}{N(x)+1} + 1 \quad (3)$$

Where,  $N$  represents the total number of texts, and  $N(x)$  represents the number of texts containing words in the whole text.

$$TF - IDF(x) = TF * IDF(x) \quad (4)$$

Where  $TF$  stands for word frequency,  $IDF(x)$  stands for the measurement of the importance of a word  $x$ .



\* Note:后备箱(Trunk),空间(Space),异响(Abnormal sound),内饰(Interior),续航(Battery life), 座椅(Seat), 刹车(The brake),发动机(The engine)

**Figure 2.** Topic evolution of "most unsatisfactory" field word cloud from 2014 to 2021

The WordCloud map is mainly to highlight the words with high word frequency in the text, and the higher the word frequency is, the bigger the font will be. This paper uses jieba and WordCloud library to comment on the "why you choose this car" field and draws the WordCloud topic evolution map of the "least satisfactory" field comment text in years to analyze the topic evolution of the field comments, as shown in Figure 2. As can be seen from Figure 2, with the passage of time, from 2014 to 2016, ordinary consumers were mainly dissatisfied with the "trunk", "space", "abnormal sound", "interior" and other aspects of new energy vehicles. The trunk includes, for example, "there is no open button in the trunk, so it is not practical to open the trunk by keys or central control every time"; "There are no physical buttons in the trunk"; "The trunk is hard to close," and other comments. "Space" includes things like "Could the trunk be even smaller?" "Too little trunk space"; "Low co-pilot legroom"; "Wheelbase is a bit short, space is a bit small" and other comments. "Abnormal sound" includes, for example, "Serious, abnormal sound in the car"; "Central control abnormal sound"; "Engine sound is too loud when power is lost"; "The engine can be very noisy at times" etc. Interior aspects include such as "rough interior work"; "The interior is too bad, as there's too much hard plastic and it feels cheap."

In 2017, according to the word cloud map, mass consumers are not satisfied with driving space and comfort. From 2018 to 2019, mass consumers were mainly dissatisfied with the "range", "seat", "space" and other aspects of new energy vehicles. The battery life includes, for example, "The battery life in winter is a headache"; "Battery life is affected by weather temperature"; "Battery life is not very good"; "Charging is a hassle"; "Charging is too slow"; "Charging piles in the city are few, and charging piles are expensive" and other comments. The seats include "the back of the seat is a little too straight"; "The seat angle adjustment is not large enough"; "Poor seat comfort, poor wrapping, short seat, no waist support"; "Seats are not soft enough"; "The seat cannot be adjusted up or down"; "The seat can only be adjusted forward and backward, not up and down"; "The seats are not soft enough and the suspension is not comfortable."

From 2020 to 2021, mass consumers are mainly not satisfied with the "interior", "space", "brake", "endurance" and other aspects of new energy vehicles. Interior decoration includes, for example, "the interior has a strong odor,

and smelling for a long time will make me feel a little dizzy"; "Interior is monotonous, and it's easy to get aesthetic fatigue"; "The interior is too simple and there are no leather seats"; "The interior has a strong smell"; "Interior odor problem"; "The interior layout is not good, the interior color is too dark, and there is no feeling of dynamic"; "Interior technology sense is not enough" and other comments. Mass consumers have higher requirements for interior decoration. From 2014 to 2017, although interior materials have been improved, ordinary consumers have put forward higher requirements for the "dynamic" and "technological sense" of interior decoration. In addition, the odor of interior decoration is the most common source of consumer dissatisfaction. Braking aspects include "Abnormal brake sound"; "The brakes are too stiff"; "Soft brake"; "Brake feeling is not linear enough"; "The brakes are a bit high and often hit the foot at the start"; "Front brake pads wear out too fast."

### 3.3. LDA topic model analysis

Blei et al. proposed the LDA model in 2004. The LDA topic model can be regarded as a three-layer Bayesian model containing feature words, documents and document sets, and each layer has random variables or parameters belonging to this layer [35].

#### 3.3.1. Topic Number Selection

In order to further mine the "least satisfactory" dimension review text, this paper analyzes the LDA topic model of the "least satisfactory" dimension review text, and uses Python language, Gensim and pyLDAvis libraries to generate and visualize the topic. Topic confusion and topic consistency score provide two methods to judge whether a given topic model is good or bad. Topic confusion refers to the degree of uncertainty when using the model to divide documents into topics. The lower the topic confusion is, the better the model is. The theme consistency score refers to whether the words with high probability corresponding to each theme generated by the model are semantically consistent. The higher the theme consistency score is, the better the model effect is. Theme consistency score can better judge if a model is good or not [36-37], so this article uses theme consistency score to determine the number of topics, as shown in figure 3. The consistency score increases as the number of topics increases until it reaches a certain value and levels off. The number of topics corresponding to the consistency of the highest score ahead of the leveling-off is decided as the optimal number of themes.

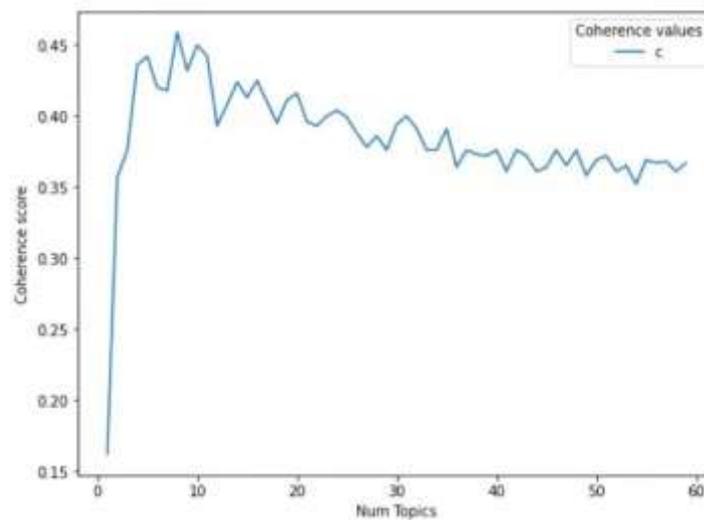


Figure 3. Topic consistency score chart

As you can see from Figure 3, the topic consistency score increases gradually until it reaches its maximum value when the number of topics reaches 8. Therefore, according to the score of theme consistency, this paper finally generates the text content into 8 themes, as shown in Table 2. Then, this paper draws a pie chart based on the theme intensity.

#### 3.3.2. Topic Description

As can be seen from Table 2, topics 2, 4 and 5 are popular topics, among which topic 6 (noise problem) has the highest intensity, accounting for 16.65%, indicating that the attention of battery life energy consumption occupies a dominant position in the comments from the "energy consumption" dimension. Theme 7 (brake problem) and theme 8 (odor problem) also had high theme strength, accounting for 16.18% and 13.34%, respectively.

Table 2 shows the theme-word distribution of the topic, and the probability value represents the importance of the topic in the corresponding topic. It can also be seen from Table 1 that consumers' most unsatisfactory comments on new energy vehicles cover a wide range of topics, including, for example, "battery life in winter is very troublesome", "battery life is not very good..." Such comments. For dissatisfaction of new energy vehicles' interior decoration, it can be found from "interior", "design", "plastic", "seat", "appearance" and other theme keywords that the interior design of new energy vehicles is not good; there is not enough sense of technology; there is too many elements of plastic; it shows a rough workmanship; seats are not soft; the seat cannot be adjusted up and down, etc. Consumers' dissatisfaction with the keys of new energy vehicles includes comments such as "keyless entry reaction is slow occasionally" and "intelligent remote-control keys do not support window opening and window closing functions". From issues on dissatisfaction with the price of new energy vehicles as well as "the price", "discount" subject keywords, it can be found that the average consumers complain about low discounts, and its theme has the highest intensity, including "the most unsatisfactory part is no discount, and the price is so fixed", "I can't believe there is no discount", etc. For the theme of the space of new energy vehicles, keywords such as "space", "back", "trunk", "seat", etc. revealed that mass consumers are not satisfied with the space of the trunk, back row, seat, and other aspects of new energy vehicles. For the aspect of noise, keywords include "brake", "sound", "function", "system", "image", "I can't stand the permeability of the seat; it is not very good and gives you stuffy feeling" and "reverse video camera resolution is not very good, especially at night. It is even a little unclear", "there is no radar probe in front of the car. Sometimes it actually enters the parking slot directly, and it is more practical to have a front probe", "I hear abnormal sounds from low-speed brake in winter" and other comments. It can be seen that consumers are not satisfied with the braking system and reversing image. For topics on new energy vehicles odor, the subject keywords include "taste", "new" and "new car smell is a bit too strong", "new car smells bad, although later I got a refill of ozone disinfection, but still, it feels a bit smelling". Apparently, mass consumers found the odor of new energy vehicles unacceptable.

### **3.3.3. Theme-word distribution**

Table 3 is the theme-word distribution, where the probability value represents the importance of the topic word in the corresponding topic. As can be seen from Table 3, in topic 6 (noise problem), the probability of the theme words "noise", "sound insulation", "engine", "sound" and "effect" is large, indicating that consumers are not satisfied with the noise of the engine and the sound insulation effect of new energy vehicles. In theme 7 (brake problem), the probability of the theme words "problem", "brake", "automatic", "abnormal sound" and "system" is large, indicating that consumers are not satisfied with abnormal sound when braking, automatic brake system is too sensitive, etc. In theme 8 (odor problem), the probability of "a little", "taste", "new car" and other keywords is relatively large, indicating that consumers are dissatisfied with the smell of the new car they buy.

**Table 2.** LDA topic table

The theme number	Topic keywords	The theme describes	Intensity of theme
Topic1	Life,The mileage, The battery, Model, A little bit, Air conditioner, Place, The problem, Electricity, Ability to	Life problem	0.06000303
Topic2	Interior, Design, Feeling, A little bit,The seat, The smell, Plastic, Appearance, Place, In terms of	Interior problems	0.11009662
Topic3	The key, Trouble, Function, Mobile phone, Km, Hours, Feed, A friend, Car body, A little bit	The key problem	0.11909799
Topic4	The car, Place feeling, A little bit, The price,preferential, In terms of, The problem, Chassis, Time	Price preference problem	0.13525714
Topic5	Space, The back, A little bit, Trunk things, The seat, Feeling, Place, The spare tire,Too small	Space problem	0.13563000
Topic6	The noise, Sound insulation, A little bit, The engine, Voice, The effect, Tire, Feeling,Speed,The city	The noise problem	0.14238442
Topic7	Problem, Brake,Automatic,Sound,Function,System,Upgrade, Image,Manufacturer	Brake problems	0.14281882
Topic8	A little bit, Feeling,Power,Taste,Place,Car,A new car,Habit,Disadvantages,The problem	Odor problems	0.15471198

**Table 3.**Subject-word distribution table

Topic 6 Noise problem		Topic 7 Brake problems		Topic 8 Odor problems	
Subject headings	probability	Subject headings	probability	Subject headings	probability
The noise	0.05947109	The problem	0.04524544	A little bit	0.06126317
Sound insulation	0.05227583	The brake	0.03015113	feeling	0.05324196
A little bit	0.05013231	automatic	0.023587	power	0.04767775
The engine	0.04859921	sound	0.02285136	taste	0.04696631
voice	0.04600891	function	0.02235553	place	0.04499832
The effect	0.04177654	system	0.02047497	models	0.0403829
tire	0.03838537	upgrade	0.01843351	A new car	0.03761671
feeling	0.03242158	vehicle	0.01767632	habit	0.02495613
speed	0.02525566	image	0.013852	disadvantages	0.02163444
The city	0.02176506	manufacturer	0.01339527	The problem	0.02133282

**3.3.4. Visualization of hot Topics**

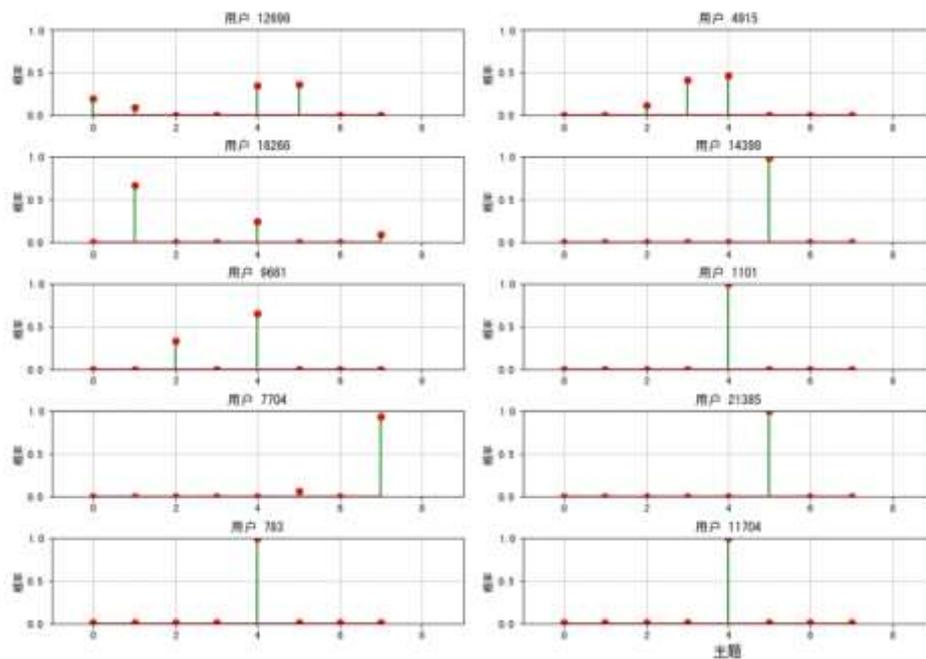
In this paper, 8 topics generated by LDA model were visualized by using pyLDAvis library, as shown in Figure 4. Each circle in the figure represents a topic, the number marked in the circle indicates the serial number of the corresponding topic, and the size of the circle indicates the number of documents containing the topic. The larger the circle, the more documents contained that topic. The distance between the circles indicates how closely related the topics are, and the closer the circles are, the more related the topics are. The bar chart on the right of the graph represents the topic with the highest distribution probability ranking, the light-colored bar represents the probability that the topic belongs to the corresponding topic, and the dark bar represents the degree of correlation between the topic and the corresponding topic.

Figure 4 shows the visualizations of three hot topics, namely, noise, braking and odor. It can be seen from the visualizations that the keywords corresponding to the three hot topics in the figure are consistent with those in Table 3. Theme 6 (noise problem) has a strong correlation with theme 7 (braking problem) and theme 8 (endurance problem); There is a strong correlation between theme 8 (odor problem) and theme 5 (space problem).

### 3.3.5. Documentation -- Visualization of topic distribution

The topic division of each text can be obtained by studying the distribution of document-topic. In this paper, 10 documents are randomly selected from the analysed text to draw the document-topic distribution map, as shown in Figure 5. The higher the vertical line, the higher the probability that the text belongs to the corresponding topic of this vertical line.

As you can see from Figure 5, among the randomly selected documents, a document is more likely to belong to one or two topics. The results of random sampling show that most texts have obvious topic division, indicating that LDA topic model can effectively mine the topic tendency of each text.

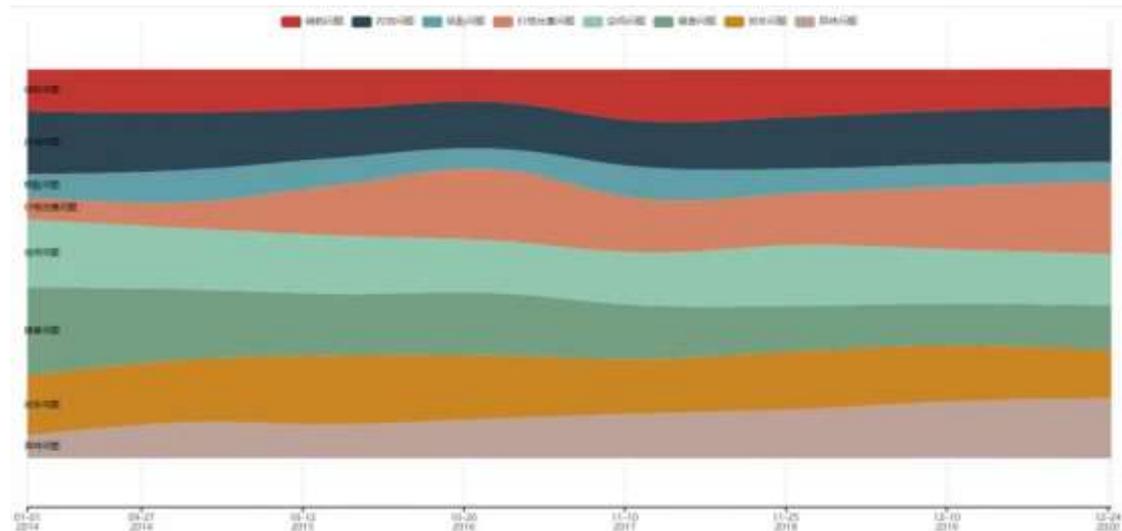


\* Note:用户(The user),概率(The probability),主题(The theme)

Figure 5. Document -- Topic distribution diagram



theme intensity is drawn (Figure 6).



\* Note:续航问题(Life problem),内饰问题(Interior problems),钥匙问题(The key problem),价格优惠问题(Price preference problem),空间问题(Space problem), 噪音问题(The noise problem), 刹车问题(Brake problems),异味问题(Odor problems)

**Figure 6.** River map of thematic intensity evolution

The wider the theme intensity evolution of rivers is, the more intense the corresponding theme is. According to the theme of the intensity of river figure and the evolutionary line chart, it can be found that "battery life" river width first narrows as time elapses, then widen again, followed by another narrow pattern. Namely, the themes intensity drops first, and then increases before decreasing again. On the whole, it shows a downward trend, and among the eight themes, the theme of "battery life" has a low intensity, indicating that mass consumers are not satisfied with the battery life of new energy vehicles. Over time, the theme intensity of "interior problems" decreases first and then rises. After 2018, the theme intensity has been rising, indicating that ordinary consumers are paying more and more attention to the interior of new energy vehicles. "Keys issue" theme intensity fluctuated as time goes by: it rose at first and lowered; it then rose again followed by another fall. As a whole it presents a downward trend, and in the eight themes, the strength of the "keys issue" theme has been ranking the lowest, showing that the average consumer attention to new energy vehicles key aspects is the weakest among the eight themes, and overall, it presents the downward trend. "Preferential price" theme intensity rose first; it dropped before another rise. After 2019, the strength of the issues of "preferential price" on the whole presents a rising trend. As the government put an end to purchase subsidy policy of new energy vehicles, mass consumers are increasingly paying attention to automotive price and promotions. This shows that mass consumers are paying more and more attention to the small financial incentives and relatively expensive prices of new energy vehicles. "Space problem" theme strength dropped first. It then rose before it fell again. Overall, it has a downward trend, but the intensity of the theme of the space problem ranked high among eight themes. This revealed that although the mass consumers are not satisfied with the new energy automobile rear and trunk space, on the whole, their dissatisfaction is declining while the consumers' attention to the theme has always been strong. The intensity of the theme of "noise problem" decreased first and then increased with time, and showed a significant decline overall, indicating that the attention of ordinary consumers dissatisfying with the noise of new energy vehicles gradually decreased with time. As time elapsed, the theme intensity of "brake problem" first increased, then decreased, then increased again and finally decreased. On the whole it shows a downward trend, indicating that the attention of ordinary consumers dissatisfying with the braking of new energy vehicles gradually weakened over time. In the time span, the theme intensity of "odor problem" has a pattern of rising first followed by a slight decline, and there was a

significant rise at the end. This shows an obvious upward trend on the whole, indicating that mass consumers are increasingly dissatisfied with the odor of new energy vehicles.

## 4. Conclusions and management suggestions

### 4.1. Research Conclusions

(1) "appearance", "space", "interior" and "the price" is the main factor affecting the ordinary consumers to purchase new energy vehicles. It can be found, according to the social network graph (figure 1), that the "new energy" and "appearance" keywords are positioned at the heart of the network, indicating that the factor most affecting ordinary consumers to purchase new energy vehicles is the "appearance". This conclusion is consistent with the China New Energy Vehicle Experience Index (NEVXI) Research report released by JD Power in 2019, which shows that in terms of the overall user experience of vehicles, new energy vehicle owners have a high degree of satisfaction with the overall user experience of vehicles, and the highest score attributes to the appearance design. Relatively low ratings were given to battery performance and on-board high-tech features. In addition to the exterior factors mentioned by NEVXI study, this paper added that "work", "pick up" and "children" keywords are closely intercorrelated. This reveals that consumers who purchase new energy vehicles are primarily concerned with commuting to and from work, picking up children and other purposes. The "price", "economic", "affordable" and "budget" keywords are also closely related, showing that consumers still care more about whether the vehicle is economical when buying new energy vehicles.

(2) According to the word cloud theme evolution chart, it can be concluded that ordinary consumers are not satisfied with new energy vehicles mainly including "interior", "trunk", "space", "engine", "abnormal sound", "battery life", "seat", "brake", etc.

(3) According to the LDA thematic model, ordinary consumers pay more attention to "noise problem", "brake problem", "odor problem" and other unsatisfactory aspects, while they pay less attention to "battery life problem", "brake problem" and other unsatisfactory aspects. This conclusion is also consistent with the China New Energy Vehicle Experience Index (NEVXI) Research report released by JD Power in 2019, which shows that the most complained problem of new energy vehicle owners is the body interior, with a PP100 score of 16.3. Only a small number of owners complained about the battery and electric motor, the core components of neVs, with scores of 7.4 and 3.4 respectively in PP100. In addition to "interior decoration problems", this paper adds to the NEVXI research report that "noise problems" and "odor problems" are also the main factors leading to the dissatisfaction of mass consumers.

### 4.2. Management Implications

(1) New energy automobile enterprises should take advantage of big data, artificial intelligence, and other advanced technology to access accurate information that affects consumer demand and to improve and innovate in terms of "appearance", "interior", "price" and "space" concerned by consumers. Enterprises should also continue to improve and make good use of tax breaks, the purchase of preferential policies, such as subsidies to better adapt to market trends and to meet the needs of Chinese consumers.

(2) New energy vehicle enterprises should speed up research and development and product innovation in terms of "abnormal sound", "seat", "trunk", "space", "brake" and other aspects that ordinary consumers are not satisfied with, and constantly improve these dimensions, so as to improve consumers' experience in these unsatisfactory aspects.

(3) The mass consumers of new energy vehicles pay close attention to shallow tech fields such as "interior problem", "preferential price" and "smell", while they pay less attention to advanced tech fields including "battery life" and "brake problem". This shows that Chinese consumers focus on technical problems at a shallow level, and

their focus on cutting-edge technical aspects shows a declining trend. However, as J.D. Power, vice President of new energy vehicle service solutions in China, said: "Chinese new energy vehicle owners not only want their vehicles to have a sense of technology and digital attributes, but also have certain requirements and expectations for the craftsmanship and quality of the vehicle itself." Therefore, new energy vehicles Chinese enterprises cannot excessively cater to mass consumers' demand on shallow technology such as "appearance", "interior" and "good value for price", and "space". It also needs to improve and innovate vehicles in terms of "battery life", "brake problems" and other aspects relatively deep in the field of technology innovation. Namely, Chinese manufacturers need to polish their products to ensure the good quality inside and out.

(4) The factors concerned by mass consumers who purchase new energy vehicles show dynamic and periodic characteristics. This reflects the developing and changing trend of the Chinese new energy automotive consumer. New energy automobile enterprises should strengthen technological innovation and research & development in the aspects of "noise problems", "brake problems" and "smell problems" to resolve consumers' dissatisfaction. New energy vehicle enterprises should take into account both the demands of special consumers and the demands of general consumers in a scientific way, and give full play to their guiding and reference roles in product innovation.

## 5. Deficiencies and prospects

This paper uses the social network, word cloud, and LDA theme model to study the online average consumer behavior data. Based on the comment text analysis of new energy vehicles general consumers, the paper explores relevant, main factors impacting on new energy vehicles purchase demand and product innovation. However, due to the lack of new energy vehicles Chinese product innovation data, the causal inference by quantitative empirical research of product innovation cannot be realized. If in the future there is a professional organization to collect and sort out new energy vehicles in the Chinese industry to establish database of product innovation, the methods used in this paper can then be combined with econometrics causal inference method to analyze the effect of new energy vehicles average consumer demand factors on the automobile enterprise product innovation.

## Reference

- [1] Yang D X, Qiu L S, Yan J J, et al. The government regulation and market behavior of the new energy automotive industry [J]. *Journal of Cleaner Production*, 2019, 210:1281-1288.
- [2] Y Xiao, Zuo X, Kaku I, et al. Development of energy consumption optimization model for the electric vehicle routing problem with time windows [J]. *Journal of Cleaner Production*, 2019, 225(JUL.10):647-663.
- [3] Ma Y, Shi T, Zhang W, et al. Comprehensive policy evaluation of NEV development in China, Japan, the United States, and Germany based on the AHP-EW model [J]. *Journal of Cleaner Production*, 2018, 214(MAR.20):389-402.
- [4] Huang Y, Qian L. Consumer preferences for electric vehicles in lower tier cities of China: Evidences from south Jiangsu region [J]. *Transportation Research Part D Transport and Environment*, 2018, 63(AUG.):482-497.
- [5] Liu Y, Yang Y, Klampfl E. Energy Challenges the Automobile Industry Faces in [J]. *Iaee Energy Forum*, 2010.
- [6] Du J, Ouyang D. Progress of Chinese electric vehicles industrialization in 2015: A review [J]. *Applied Energy*, 2017, 188:529-546.
- [7] Yin, Xiang, Chen, et al. China's transportation energy consumption and CO2 emissions from a global perspective.[J]. *Energy Policy*, 2015, 82(3), 233-248.
- [8] Wang Y F, Li K P, Xu X M , et al. Transport energy consumption and saving in China[J]. *Renewable and Sustainable Energy Reviews*, 2014, 29(7):641-655.
- [9] SCHMOOKLER Economic sources of inventive activity [J]. *The Journal of Economic History*, 1962, 22(1):1-20.
- [10] Wang Z, Xue M, Y Wang, et al. Big data: New tend to sustainable consumption research [J]. *Journal of Cleaner Production*, 2019, 236(5):117-499.
- [11] Buschken J, Allenby G M. Sentence-based text analysis for customer reviews [J]. *Operations Research*, 2017, 57(5-6):537-540.
- [12] Zarrinkalam F, Kahani M, Bagheri E. Mining user interests over active topics on social networks[J]. *Information Processing &*

- Management, 2018, 54(2):339-357.
- [13] Rezvani Z, Jansson J, Bodin J. Advances in consumer electric vehicle adoption research: A review and research agenda [J]. *Transportation Research Part D*, 2015, 34(34):122-136.
- [14] Zhang L, Zhao Z, Xin H , et al. Charge pricing model for electric vehicle charging infrastructure public-private partnership projects in China: A system dynamics analysis[J]. *Journal of Cleaner Production*, 2018, 199(PT.1-1130):321-333.
- [15] He H, Fan J, Li Y, et al. When to switch to a hybrid electric vehicle: A replacement optimisation decision [J]. *Journal of Cleaner Production*, 2017, 148(APR.1):295-303.
- [16] Hardman S, Jenn A, Tal G, et al. A review of consumer preferences of and interactions with electric vehicle charging infrastructure [J].*Transportation Research Part D: Transport and Environment*, 2018, 62(JUL.):508-523.
- [17] Boqiang, Lin, et al. Why people want to buy electric vehicle: An empirical study in first-tier cities of China [J]. *Energy Policy*, 2018.112(1):233-241.
- [18] Zhang X, Bai X, Shang J. Is subsidized electric vehicles adoption sustainable: Consumers' perceptions and motivation toward incentive policies, environmental benefits, and risks [J]. *Journal of Cleaner Production*, 2018, 192(AUG.10):71-79.
- [19] Li W, Long R, Chen H, et al. Effects of personal carbon trading on the decision to adopt battery electric vehicles: Analysis based on a choice experiment in Jiangsu, China [J]. *Applied Energy*, 2018, 209(JAN.1):478-488.
- [20] Wenbo, Li, Ruyin, et al. A review of factors influencing consumer intentions to adopt battery electric vehicles [J]. *Renewable & Sustainable Energy Reviews*, 2017.
- [21] Jensen A F, Mabit S L. The use of electric vehicles: A case study on adding an electric car to a household [J]. *Transportation Research Part A Policy & Practice*, 2017, 106:89-99.
- [22] Xian Z, Wang K, Hao Y, et al. The impact of government policy on preference for NEVs: The evidence from China [J]. *Energy Policy*, 2013, 61(Oct.):382-393.
- [23] AJA, AJB, BMAC. Preference and lifestyle heterogeneity among potential plug-in electric vehicle buyers-Science Direct [J]. *Energy Economics*, 2015, 50:190-201.
- [24] Morton C, Anable J, Nelson J D. Exploring consumer preferences towards electric vehicles: The influence of consumer innovativeness [J]. *Research in Transportation Business & Management*, 2016, 18:18-28.
- [25] He X, Zhan W, Hu Y. Consumer purchase intention of electric vehicles in China: The roles of perception and personality [J]. *Journal of Cleaner Production*, 2018, 204(PT.1-1178):1060-1069.
- [26] Wang Yuehui, WANG Qing. Influence factors of purchasing intention of new energy vehicles in Beijing: Based on TAM and TPB Integrated Model [J]. *Management science*, 2013, 21(52):691-698.
- [27] Xu Guohu, Xu Fang. Research on Influencing Factors of New Energy Vehicle Purchase Decision [J]. *China Population, Resources and Environment*, 2010, 20(11): 91-95.
- [28] Wang Ning, Yan Runlin, Liu Yafei. Research on the feature identification and market acceptance of potential consumers of electric vehicles [J]. *China Soft Science*, 2015(10):70-84.
- [29] Li Chuang, Ye Lulu, Wang Liping. The Influence of New Energy Vehicle Consumption Promotion Policy on Potential Consumers' Purchase Intention [J/OL].*China Management Science*:1-11[2021-10-15].<https://doi.org/10.16381/j.cnki.issn1003-207x.2019.1845>.
- [30] Liu Xueyuan, Sun Min, Wujin Chu. Research on innovative development countermeasures of my country's new energy vehicle industry:based on a questionnaire survey of China, Germany, Japan, South Korea and the United States [J]. *Science and Technology Progress and Countermeasures*, 2015, 32(21):77 -83.
- [31] Sun Xiaohua, Xu Shuai. Research on the Influence of Government Subsidies on New Energy Vehicle Purchase Intention [J]. *Journal of Dalian University of Technology (Social Science Edition)*, 2018, 39(03):8-16.
- [32] Buschken J, Allenby G M. Sentence-based text analysis for customer reviews [J]. *Operations Research*, 2017, 57(5-6):537-540.
- [33] Zhang Ning, Zhao Wenfei, Pang Zhiliang, Zhang Xin. Research on the influencing factors of creative adoption in the open innovation community of enterprises: the perspective of value co-creation [J]. *Science and Technology Progress and Countermeasures*, 2021, 38(16): 91-100.

- [34] Zarrinkalam, Fattane, Kahani, et al. mining user interests over active topics on social networks [J]. Information Processing & Management, 2018, (4): 99-132.
- [35] Blei D, Ng A, Jordan M. Latent Dirichlet Allocation [J]. Journal of Machine Learning Research, 2003, (3): 993-1022.
- [36] Roeder, M. and A. Both, et al. Hinneburg, Exploring the Space of Topic Coherence Measures [J]. Association for Computational Linguistics, 2015, 2(13), 399-408.
- [37] Mimno D M, Wallach H M, Talley E M, et al. Optimizing Semantic Coherence in Topic Models[C].//Proceedings of the Conference on Empirical Methods in Natural Language Processing. Association for Computational Linguistics, Association for Computational Linguistics. 2011, 27(6):727- 733.