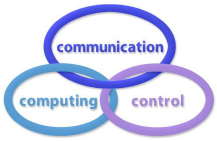
INTERNATIONAL JOURNAL OF COMPUTERS COMMUNICATIONS & CONTROL
Online ISSN 1841-9844, ISSN-L 1841-9836, Volume: 20, Issue: 6, Month: December, Year: 2025
Article Number: 6934, https://doi.org/10.15837/ijccc.2025.6.6934



CCC Publications

An Approach of Sentiment-Topic Mining Based on User Online Comments on New Energy Vehicles

Feng Hu, Zhaohan You, Junyuan Cai

Feng Hu
1. School of Management
Guangdong University of Technology, China
2. Institute of Big Data Strategic Research
Guangdong University of Technology, China
fenghu@gdut.edu.cn

Zhaohan You*
School of Communication
Hong Kong Baptist University, China
947873052@qq.com
*Corresponding author: 947873052@qq.com

Junyuan Cai
School of Management
Guangdong University of Technology, China
294968336@qq.com

Abstract

The automobile industry is undergoing revolutionary changes. New energy vehicles, instead of fuel vehicles, gradually occupy the automobile sales market and have become an irresistible trend. With the development of social networks and automobile forums, consumers are increasingly using Internet channels to post online comments on automobile products or services after purchasing automobile products or enjoying services. These online comments contain a large number of user opinions and sentiment information and reflect the description of products or services and user experience from the user perspective. To discover potential user demands and sentiment orientations, this paper proposed a sentiment-topic mining model based on online comments on new energy vehicles. Taking the Pacific Automobile website as an example, we used Python to collect 46,530 online comment data, analyzed positive and negative user sentiment orientations based on the Word2Vec-SVM model, analyzed the reasons for negative comments based on the LDA model, and revealed potential user demands and product inadequacies. The research results revealed that users were satisfied with "appearance", "power", "configuration" and "handling" of new energy vehicles but not with "space", "interior", "energy consumption" (battery life) and "comfort". The reasons why users were dissatisfied with new energy vehicles vary in different dimensions, which maybe reflect different user demands and current inadequacies of new energy vehicles. This sentiment-topic mining model can be generalized to discover potential user demands and analyze user sentiments in various industries, help enterprises gain insight into market trends, guide product improvement, and enhance product competitiveness.

Keywords: New energy vehicles, Online comments, Sentiment analysis, Topic mining.

1 Introduction

The excess of greenhouse gas emissions, caused by fossil fuels such as coal, oil and gas, has seriously damaged the global ecological environment, and countries around the world have set targets to limit carbon emissions. The transportation industry, as one of the major contributors to global greenhouse gas emissions, regards the research and development of new energy vehicles as the focus of its efforts to achieve decarbonization [21]. Countries around the world have formulated blueprints for the development of new energy vehicles according to their respective national conditions and have issued a series of incentive policies to promote technological improvement and the market development of new energy vehicles [18].

China is currently the world's largest automobiles consumer and producer. Since Tesla first entered China's new energy vehicle market in 2014, China has vigorously expanded the scale of the new energy vehicle market with the emergence of new car-making brands such as NIO, Xpeng, Li Auto and WM Motor. Under the dual roles of policy and market, new energy vehicles continued to grow rapidly in 2023, the annual production and sales scale reached 9.5 million, and the market share was approximately 31.6%, which ranks first in the world for nine consecutive years. However, China's new energy vehicles are still in a disadvantageous position in the market competition with traditional vehicles. How to break the current situation, get the recognition of users and rapidly increase the market penetration rate of new energy vehicles is an important issue that the government and electric vehicle manufacturers have to take into consideration.

With the popularity of the Internet and the rapid development of social networks, more and more consumers are choosing to share their shopping experience on social networks. In particular, due to travel restrictions during the epidemic, users gathered more on social networks and shopping platforms to form more topics and discussions. Major professional automotive websites, For example, Pcauto, Autohome and Dongchedi et al., also effort to create an interactive space for them to communicate and share their driving experiences and personal opinions. These online comments not only directly reflect the perception of user experience and real-time feedback on automobile products or services, but also affect the choice and decision of other potential users[5]. Enterprises can also obtain user demands and product optimization direction to better serve users. At the same time, compared to traditional user demand research, online review data has the advantages of large data volume and low collection cost [19]. Therefore, it has become an important topic to mine potential user demands based on online reviews, identify the direction of product optimization, and then improve user satisfaction and loyalty, and enhance enterprise competitiveness. Under limited budget and resource allocation, to maintain sustainable market competitiveness, enterprises achieve the optimal synergistic state of user experience and corporate interests through comprehensive consideration of user demand priorities.

However, online comment data are huge, and there may be some problems like a large number of repeated comments, irregular statement structure, etc. How to quickly and accurately obtain the core content of user comments and effectively assist potential users in making purchase decisions is very important. Therefore, this paper adopts advanced data mining, machine learning, and artificial intelligence technology to realize automatic and rapid processing of large-scale comment datasets, reduce time cost, and improve the efficiency of information screening and intelligence acquisition. Moreover, text sentiment analysis and topic extraction can discover user demands and sentiment orientations from relevant comments, so as to achieve multidimensional and in-depth user preference acquisition and help new energy vehicle manufacturers improve business activities and government formulate public policy.

2 Related work

2.1 Text analysis of online comments

The rapid development of natural language processing technology provides the convenience of mining valuable intelligence and knowledge hidden in unstructured text data. In particular, with the wide application of artificial intelligence and deep learning methods, more and more scholars are paying attention to online user comments to seek the potential intelligence and knowledge behind

texts to provide decision support for managers[1][10].

Sentiment analysis and topic extraction are two types of text analysis methods commonly used in online comments. The main goal of sentiment analysis is to classify text into positive, negative, or neutral categories. The current research hotspots focus on sentiment classification, sentiment orientation analysis, aspect level sentiment analysis, etc. Currently, sentiment analysis has a wide range of applications in brand management, market research, public opinion analysis, social media monitoring, and customer service. Through sentiment analysis, businesses and organizations can get a better understanding of the public's attitudes and perceptions and make more informed decisions and improvements. Online sentiment analysis is mainly based on sentiment dictionary, machine learning, and deep learning. Early research on text sentiment recognition adopts mainly the dictionary-based method[16]. This method first uses a pre-established sentiment word dictionary or vocabulary to mark the sentiment polarity of the words in the text, and then calculates the sentiment score of the whole text through a series of rules or algorithms to complete the sentiment classification task. With the rapid development and technological progress in the field of machine learning, text sentiment recognition research has gradually shifted to machine learning methods. These methods enable sentiment analysis to better understand and identify emotional content in text using massive amounts of data and powerful computing power[12]. In recent years, since deep learning has made great progress in computer science, this technology has gradually become the dominant method in the field of text sentiment recognition[20]. Compared to traditional dictionary-based or machine learning methods, deep learning can capture semantic and emotional information in text more effectively, so it has achieved more significant results in practical applications[7][15].

Topic mining is a method of extracting keywords that are related to each topic in the text. It can be used by manual method, word frequency method, unsupervised model, etc. Latent Dirichlet Allocation (LDA) is a common topic generation model and is an unsupervised Bayesian topic model to discover potential topics in a document dataset, which is especially suitable for automatic processing of large-scale corpora and has unique advantages in text similarity processing [2]. Subsequently, some scholars combined machine learning and deep learning models to enhance text semantics, and some scholars focused on improving text similarity[13]. In addition to direct LDA(D-LDA), PD-LDA[14], BDPCA-LDA[22] and so on have also emerged. In this paper, sentiment analysis and topic mining reveal different semantic information in online comments, but there is a strong correlation between sentiment and topic in online comments, and more potential intelligence and knowledge can be discovered through sentiment-topic coupling analysis.

2.2 User Comments mining of new energy vehicles

New energy vehicles use advanced technology to control power and drive, with new technologies and structures, and their technological upgrades are mainly concentrated in battery, electric drive, electronic control, vehicle networking, autonomous driving and other aspects, which are the blind spot of consumer cognition. With the rapid development of the Internet and social networks, User Generated Content (UGC) has become a new way for users to use the Internet, and the online comments of new energy vehicle users have become an important reference for consumers to purchase new energy vehicles. At the same time, it is also an important source of information for automobile manufacturers to capture the product and service demands of online users, to improve the product design, and the sales promotion of new energy vehicles.

In recent years, with the rapid development of the new generation of information technology, domestic and foreign research in this field has made great progress, mainly focused on sentiment analysis, demand acquisition, preference analysis, sales forecasting, and other aspects of new energy vehicles. For example, there are scholars who integrate sentiment analysis of online comments to build a new energy vehicle sales forecasting model [6], automatically extract, classify, and arrange product requirements[3], evaluate new energy vehicles[8], propose a multi-standard product recommendation and estimated risk psychology[4][9], or adopt a channel attention mechanism to amplify the importance of key information in comments, to facilitate the classification of comment relationships and sentiment prediction [17].

Although sentiment analysis methods have played an important role in online comments on new

energy vehicles, challenges remain with the development of new generation information technology and artificial intelligence. Firstly, the contextual information in the comments may be incomplete, which will lead to emotionally ambiguous analysis results. Secondly, the field of new energy vehicles covers numerous specialized terms that require a better understanding of domain-specific knowledge. In addition, it is necessary to investigate in depth the positive and negative evaluations obtained by sentiment classification, find the shortcomings in product design, and improve the design and operation activities of research and development of new energy vehicle manufacturing products.

In order to meet these challenges, this paper proposed a WSL (Word2Vec-SVM-LDA) model to improve the multidimensional classification method of online comments on new energy vehicles, with the aim of extracting users' attitudes towards products from user-generated online text reviews and providing more accurate sentiment classification and topic analysis results. Therefore, this paper collected online comment data from professional automobile websites, obtained user cognition, preference, and sentiment orientation toward various performance indicators and details of new energy vehicles through sentiment classification analysis, and mined user demands for new energy vehicles of different dimensions, to provide optimization strategies for the design, research and development, and promotion of new energy vehicles, and promote the vigorous development of the new energy vehicle industry.

3 Methodology

3.1 Overview of our method

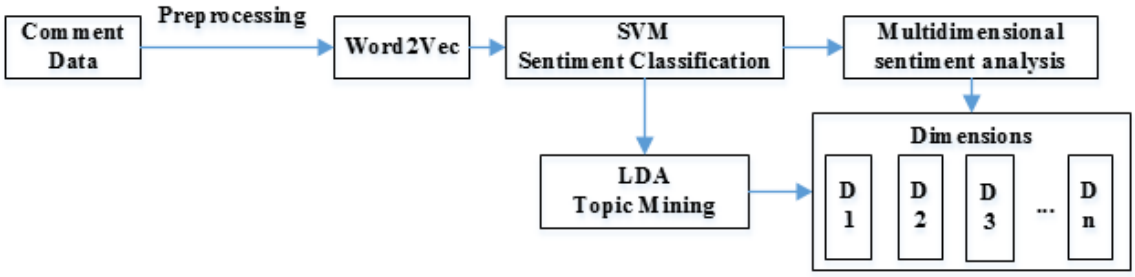


Figure 1: Sentiment-topic analysis framework of online comments

Fig.1 shows a sentiment-topic analysis framework based on online comments on new energy vehicles. The analysis process mainly includes three stages: user comment data acquisition and preprocessing, text vectorization based on the Word2Vec model, and sentiment-topic mining based on the SVM-LDA model.

In the first stage, user comment data is crawled by a Python tool from the target website and preprocessed. In this paper, we use the octopus crawling algorithm to gather user comments. Data preprocessing includes data cleaning and deduplication, splitting words, removing stopwords, etc.

In the second stage, text vectorization is carried out using the Word2Vec model. Word2Vec is a word vector model trained by Mikolov from more than one billion Google news words [11]. In essence, the Word2Vec model is a two-layer neural network that can transform words into vector representations. After the Word2Vec model is trained to construct a word vector, the results of each comment are also introduced into the Word2Vec model to obtain a text vector. Each comment is not the same length, so the average of all the term vectors in each comment is used as the text vector for a single comment, which is used as the input to the model.

In the third stage, based on the SVM-LDA model, we perform sentiment classification based on the SVM model and topic mining based on the LDA model. The SVM model is a supervised learning model that requires annotated data to be prepared in advance. The online comment data used in this paper have two dimensions (advantage and disadvantage dimensions). Therefore, the advantage dimension can be regarded as positive comments and the disadvantage dimension as negative comments, which can be supplemented and improved by manual annotation. After sentiment annotation and sentiment

Table 1: The environment configuration

Experimental Environment	Environment Configuration
Hardware	Notebook: DELL G15 5530-3746B CPU: Intel Core i7-13650HX Memory: 16GB Graphics Card: NVIDIA Geforce RTX 4060 8G
Software	OS: Windows 10 Programming Language: Python 3.8 IDE: Jupyter Notebook

classification based on the SVM model, we perform multidimensional sentiment analysis to identify comment topics and related topic keywords based on the LDA model.

3.2 Experiment

3.2.1 Experimental Environment

This experiment uses a DELL notebook configured with a CPU Intel Core i7-13650HX, memory 16G, and an 8G independent graphics card NVIDIA Geforce RTX 4060. The operating system is Windows11, the programming language is Python 3.8, and integrated development environment is Jupyter Notebook. The specific configuration of the experimental environment is shown in Table 1. This configuration provides powerful computing resources for the training and optimization of deep learning models and is especially suitable for large-scale data processing and the construction of the Word2Vec-SVM-LDA model.

3.2.2 Experimental Setup and Details

Step 1: Data Acquisition

Currently, there are many automobile information apps and websites in China, such as Pcauto, Autohome, Dongchedi, Yiche, Youjia, etc. Each website has different quality user comments. Through a comparative survey, Pcauto has an advantage in data acquisition in terms of users, comments, and comment systems. Therefore, we select the Pcauto website to collect user comments.

Pcauto user comments include 10 dimensions, such as advantages, disadvantages, space, interior, energy consumption, configuration, comfort, appearance, handling, and power. In this paper, a Python crawler tool Octopus is used to collect online user comments on new energy vehicles, and a total of 46530 online comments are obtained.

Step 2: Data Preprocessing

(1) Data cleaning

Original online comments contain some invalid data, such as repeated comments, meaningless URL information, emoticons, and special symbols. This invalid data information will affect the results of subsequent text analysis, so this paper needs to clean these original comments and remove duplicate records. After the above data cleaning and deduplication, 32,839 valid comments were obtained, including 3426 positive comments (advantages), 3374 negative comments (disadvantages), 3317 space comments, 3312 interior comments, 3267 energy consumption comments, 3302 configuration comments, 2786 comfort comments, 3365 appearance comments, 3332 control comments, and 3358 power comments. We saved these comment data in the local Excel table.

(2) Splitting words

Jieba library in Python is used to split user comments. Jieba library is an open source Chinese word segmentation tool, which can split a Chinese text according to the granularity of the word and tag each word with a part of speech. Jieba word segmentation algorithm based on prefix dictionary and suffix rule has high accuracy and speed, and is widely used in natural language processing, search engine, information retrieval and other fields.

The default thesaurus for stuttering participles is already very comprehensive, but some specific domain or industry terms may not be in the default thesaurus, in which case you can add a custom

Table 2: Parameter settings

Parameters	Settings	Description
sg	0	sg=1(skip-gram algorithm); sg=0 (CBOW algorithm)
vector_size	300	Term vector dimension
window	5	Indicating how many context words to take
alpha	0.03	Learning rate

Table 3: Partial synonyms

Observation	Synonyms
Power	('Speed up', 0.7438), ('Explosive power', 0.7385), ('Muddle along', 0.7259), ('Accelerate', 0.7243), ('Smoothy', 0.7118)
Control	('Precision', 0.7053), ('Nimble', 0.6790), ('Easy', 0.6406), ('Convenient', 0.6104), ('handy', 0.5963)
Appearance	('Appearance', 0.8732),('High appearance level', 0.7073),('Conquer', 0.7024),('High grade', 0.6901),('High Level', 0.6847)

dictionary to make up for the deficiency. The online user comments contain many specific auto industry terms, so we built a custom dictionary. The custom dictionary adds the Sogou lexicon of automotive terms and new energy vehicle brands and terms, such as "Nezha V", "Benben ESTAR", "Song Pro", "charging pile", "increase program", "front face", "side face", and so on. You can load the custom dictionary by calling the jieba.load_userdict() method to improve the accuracy of high words and get closer to the real data analysis results.

(3) Removing stopwords

Stopwords refer to some common, meaningless, or useless words. They do not contribute to the text analysis results but maybe interfere with the text subsequent processing and affect the accuracy of analysis results. Therefore, the Chinese HIT stopword table is used to remove stopwords, optimize the jieba word segmentation results, and facilitate subsequent data mining analysis.

Step 3: Model Training and Evaluation

(1) Word2Vec Model

After splitting words and removing stopwords, we use the Word2Vec model to build word vector training based on the results of word segmentation. Finally, the trained Word2Vec model is saved as "clean_pos_neg_vec.model". The parameter settings of the Word2Vec model are shown in Table 2.

After training the Word2Vec model, we use the function w2vmodel2.wv.most_similar() to see the cosine similarity between words and then get the synonyms of the words. The training results of the Word2Vec model can be evaluated. we select the three words "Power", "Control", and "Appearance" for observation, and show them in Table 3.

Through the synonyms of the entries in Table 3, it is found that the synonyms related to "power" are "speed up", "explosive power", "dragging water", etc.; the synonyms related to "control" include "precise", "flexible", "easy", "convenient", "handy", etc.; the synonyms related to "appearance" are "appearance", "high appearance level", "conquest", "grade", "high-end", etc.. The results of similar terms obtained by training distributed word vectors are in accordance with Chinese logic, which show that the Word2Vec algorithm can accurately capture the semantic relationship between different terms when processing Chinese text, thus providing a reliable basis for subsequent text analysis tasks.

(2) SVM Model

Sentiment analysis of online comments can discover user sentiment orientation through the feature vector analysis contained in natural language text, which can help enterprises gain insight into real user demands and experiences of their products or services, make more effective decisions in time, and improve product competitiveness and user satisfaction. Based on the Word2Vec model, we get a word vector, and the SVM model is used to classify the sentiment tendencies of user comments.

Comments sentiment tag: SVM model is a supervised learning model, and annotated data must be prepared in advance. As mentioned above, the data used in this paper are structured, with both a strength dimension and a weakness dimension of the review data. Therefore, the advantage dimension

Table 4: Training results of Word2Vec-SVM model

Results	accuracy	precision	recall	f1-score	support
Negative		0.92	0.93	0.93	669
Positive		0.93	0.92	0.93	691
avg/total	0.93	0.93	0.93	0.93	1360

can be regarded as positive comments and the disadvantage dimension as negative comments, which can be supplemented and improved by manual annotation. After sentiment annotation, we obtain 3,426 positive comments and 3,374 negative comments. All comment data are divided into test data (20%) and training data (80%) for model training.

Text vector construction: After training in the Word2Vec model and building word vectors, the results of each comment are also introduced into the Word2Vec model to obtain text vectors. Each comment is not the same length, so the average of all the term vectors in each comment is used as the text vector for a single comment, which is used as input to the model.

Tuning of model hyperparameters: The hyperparameters in the Support Vector Machine (SVM) model are C and gamma. The parameter C represents a weight that adjusts the size of the interval, the preference for classification accuracy, and therefore the tolerance for error. The higher the value of C, the higher the tolerance for misclassification. Overfitting is more likely to occur. In contrast, the smaller the value of C, the less tolerance for error. As long as the value of C is abnormal, the adaptability of the model to new data will decrease. Another hyperparameter gamma reflects the distribution of the data mapped to the new feature space. Similarly to the case with the value of the hyperparameter C, there is an inverse relationship between the value of the hyperparameter gamma and the number of support vectors. The number of support vectors has a big impact on the speed of training and prediction.

In this paper, sklearn's GridSearchCV grid search is used to search for hyperparameters. In the grid search, the parameters are adjusted in sequence according to the step size within a given parameter range, and the adjusted parameters are used to train the classifier. Among these parameters, the most accurate group should be found, which is actually a process of training and comparison. Grid Search is an exhaustive search that loops through all options and tries every possible one until the best performance parameters are obtained as the final result. It is like trying to find the largest number in an array. But the biggest drawback of this method is that it takes a long time.

We use gamma = [0.001, 0.01, 0.05, 0.1, 1,10,100] and C = [0.01, 0.1, 1,10,100] as candidate hyperparameters and validate the F1 value of each model. When the value of F1 is maximum, the model hyperparameter is the best. Finally, we obtain the optimal hyperparameters C=100, gamma=0.01 and F1=0.9321.

Model Evaluation: C-Support Vector Classification (SVC) of sklearn is used for the training of the SVM model, and the training results of the Word2Vec-SVM model are shown in Table 4.

According to the classification_report (Table 4) and confusion_matrix (Figure 2) provided by sklearn, the performance of the SVM based on the average sentence vector in the test set shows that the prediction accuracy rate is 0.92 for negative comments and 0.93 for positive comments. The average accuracy of the model is 0.93. The model has a recall rate of 0.93 in both positive and negative reviews. In addition, the F1 value of the model in both positive and negative comments is 0.93.

Through confusion matrix analysis, 1360 test data were obtained, but 1265 test data were correctly separated and 95 test data were incorrectly classified. The value of auc obtained from the ROC curve (Figure 3) is 0.978. The prediction accuracy using the SVM model reaches 0.978.

3.3 Experimental Results and Discussion

3.3.1 Results of sentiment analysis

In this section, we use the above classifier to classify and predict eight-dimensional comment data(space, appearance, power, configuration, interior, comfort, energy consumption, and control) and get users' positive and negative sentiment orientations. Figure 4 is a percentage comparison chart

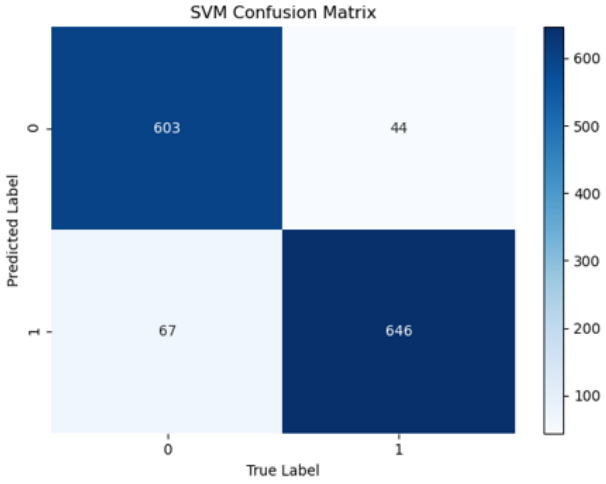


Figure 2: Confusion_matrix

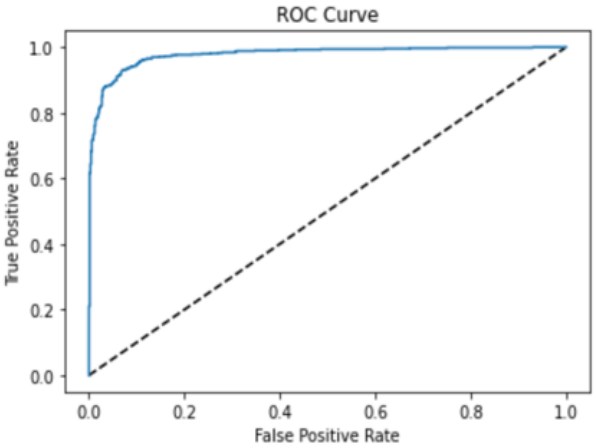


Figure 3: The ROC curve

of eight-dimensional positive comments.

According to the above results, user satisfaction with the four dimensions of appearance, power, configuration and control of new energy vehicles is relatively high, and the proportion of praise is greater than 58%. Among them, appearance represented the highest proportion of praise, reaching 88.11%, followed by power 67.42%. In addition, user satisfaction with space, interior, energy consumption (endurance), and comfort of new energy vehicles is not ideal, and the proportion of praise in these four dimensions is only about 40%. Among them, users are least satisfied with the energy consumption (battery life) of new energy vehicles, followed by comfort. In particular, manufacturers need to work on energy consumption and comfort to improve customer satisfaction in these areas.

3.3.2 Topic mining results analysis of negative comments

The unsupervised LDA model can extract the topics and topic words of negative comments, implement topic classification, and visually present the topic mining results to interpret and summarize in multiple different evaluation dimensions of new energy vehicles. According to the results of the above analysis of negative comments, the LDA model is used to classify the topics of negative comments on space, interior, energy consumption (battery life) and comfort.

(1) LDA analysis of comfort

The number of topics (n_topics) should be set in the first place for the analysis of the topic words obtained by LDA for the negative evaluation. After repeated trials and confusion analysis, it is found that the output quality of the model is best when six topics are selected. Therefore, six topic words are selected in the comfort dimension. Finally, the six negative evaluation topics in the comfort dimension

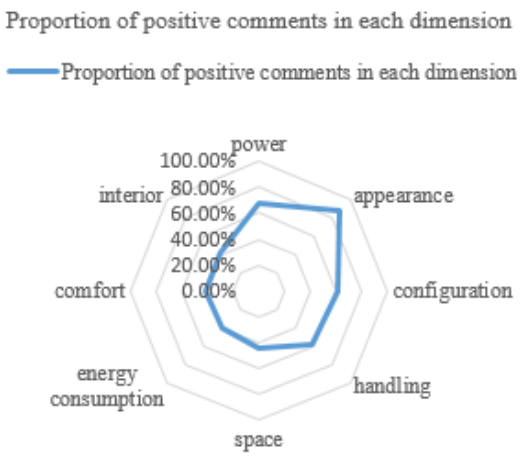


Figure 4: Percentage comparison chart of positive comments

Table 5: Topic extraction results of comfort dimension

Topic	Keywords
Topic 0	feel comfortable speed bumps cushioned chassis vehicle overall body problem adjustment
Topic 1	noise affects engine odor new vehicle vehicle basically taste aspect electric cars
Topic 2	rear space comfort adjustment overall aspect price place angle seats
Topic 3	effect soundproof air conditioning vehicle waist comfort leather car adjustable design
Topic 4	some suspension rear chassis road speed bumps adjust vehicle time speed
Topic 5	sound insulation sound engine basic effect mute process tyre window level

and the top 10 keywords corresponding to each topic are revealed as shown in Table 5.

By extracting the topic of negative comfort evaluation and its corresponding keywords from the LDA topic model, it can be found that users of new energy vehicles have more negative comments on comfort. Topic 0 and Topic 4 show that users are concerned about the chassis adjustment of new energy vehicles, and the shock absorber effect is not very good when passing the speed bump, and the comfort level is relatively low in this respect. Topic 1 and Topic 5 mainly reveal that the problem of vehicle noise and vehicle odor also affect the user's comfort satisfaction. Topic 2 and Topic 3 discover the space problem of the rear and seat design by some keywords like "rear", "seat", "angle", "waist", etc.

In short, users have the following negative comments on the comfort of new energy vehicles: chassis tuning, vehicle noise, vehicle odor, rear space and seat design. In general, the experience of using new energy vehicles is not ideal, and most of the current new energy models are not as good as the driving experience of traditional cars. Therefore, companies promoting new energy automobiles should open up a broader market, occupy more market share, and always focus on improving the comfort experience. Paying attention to the comfort of new energy vehicles can bring a better user experience, meet user demands, and improve users' driving pleasure.

(2) LDA analysis of interior

Similarly, ten topics are selected for the LDA model analysis of negative interior comments, and the quality of the model output is the best. Therefore, we identify the ten topics of the internal dimension and list the top five keywords for each topic as Table 6.

Through LDA topic extraction, we obtain the keywords of ten topics with negative comments on the interior dimension of new energy vehicles. From the analysis of the results, the main reasons users are dissatisfied with the interior dimension of the performance source car are the design and

Table 6: Topic extraction results of interior decoration dimension

Topic	Keywords
Topic 0	vehicle features design decorative elements
Topic 1	seats price leather steering wheel black
Topic 2	screen keys dashboard LCD phone
Topic 3	plastic prices some materials cheap
Topic 4	grade material space problem gap
Topic 5	feel smell a little new car
Topic 6	odor dashboard abnormal sound color new car
Topic 7	overall aspect detail design feel
Topic 8	place material feel large screen individual
Topic 9	material design style thick layout

Table 7: Topic extraction results of space dimension

Topic	Keywords
Topic 0	trunk seat children wife children friends overhead affect overall time
Topic 1	feel seat trunk storage basic wheelbase family outing family capacity
Topic 2	storage trunk car pressure place design storage grid basical wife instead of walking
Topic 3	Leg some head trunk distance the trunk sub problem adjust the backup
Topic 4	problem home trunk seat adjustment body convenient level reasonable design
Topic 5	personal adult adult problem some trunk center cross-legged floor sedan

use problems of the materials. From Topic 0, Topic 7 and Topic 9, it can be seen that "design", "decoration", "details", "style" and "layout" have become keywords, indicating that new energy vehicle users are not satisfied with the decorative style and layout details of the interior dimension. In addition, from Topic 1, Topic 2 "seat", "steering wheel", "LCD", "screen", etc., it can be seen that in the specific object, the customer is very concerned about the seat, the center screen, the LCD dashboard, and the steering wheel. These items play a very important role in the customer's evaluation of the interior. The key words of Topic 3, Topic 4 and Topic 9 are "plastic", "cheap", "grade", "materials", etc., mainly indicating that the material grade of the car is not high even cheap. Finally, from the "taste", "odor", and "abnormal sound" of Topic 5 and Topic 6, it can be seen that there is an overlapping problem with the comfort dimension, and the abnormal smell is also one of the important factors affecting consumers' satisfaction with the interior.

In short, consumers' negative evaluation of the interior of new energy vehicles has three aspects: the design details and layout style, material grade of seat leather, steering wheel, and center console, abnormal smell and movement of the car. Today, new energy vehicle manufacturers have focused mainly on battery safety, extending battery life, optimizing power systems, and other hardware, neglecting to improve the internal structure of the car. Therefore, after new energy vehicles have entered a new stage of development, how to make consumers feel fresh is a new problem faced by major automobile manufacturers.

(3) LDA analysis of space

In the LDA topic model analysis of spatial negative evaluation, six topics are selected, and the output quality of the model is the best. Therefore, six topic words are selected in the spatial dimension. The final determination of the six topics of the negative evaluation of the interior dimension and the top ten keywords corresponding to each topic is shown in Table 7.

Through LDA topic extraction, keywords of six topics with negative comments on the spatial dimension of new energy vehicles are obtained. From the results analysis, almost every topic has "trunk" in the keyword. It can be seen that users pay most attention to the size of the trunk in the spatial dimension, reflecting the user's relatively higher requirements for storage space when driving.

Table 8: Topic extraction results of consumption dimension

Topic	Keywords
Topic 0	battery air conditioning customary road condition capability actual individual integrated local situation
Topic 1	fuel consumption basic expected electric vehicle heavy fuel saving energy consumption new car displacement mixed
Topic 2	fuel consumption urban areas some range feel weight road new cars overall
Topic 3	power consumption run-in period seat feel some aspect sound insulation air conditioning effect noise distance
Topic 4	problem basic the official situation the car oil time mode data air conditioning

In addition, seats are also high-frequency keywords. It is not difficult to see that consumers have mentioned the seat in terms of interior, comfort, or spatial dimension. The leather material of the seat, the size of the seat space, and the space between the legs and the head are the main priorities of the users.

In summary, new energy vehicle manufacturers still need to reasonably design the storage space in the spatial dimension for users driving, whether it is the size of the trunk or the storage space in the car, such as the space for placing water cups and placing some small items. In addition, from the perspectives of interior decoration, comfort and space, the upgrade of seat seats should be strengthened, and it is also a special concern of users.

(4) LDA analysis of energy consumption

The output quality of the model is the best when five topics are selected for negative comments on energy consumption. Therefore, we choose five topic words in the energy consumption dimension. Finally, the five topics of negative evaluation of the energy consumption dimension and the top 10 keywords corresponding to each topic are determined as shown in Table 8.

New energy vehicles have both pure electric vehicles and hybrid electric vehicles, so fuel consumption will still be the keyword. It can be seen that, from the topic words of energy consumption, users pay the most attention to fuel consumption and power consumption, including battery life and data performance. New energy vehicles are products that have received people's attention in recent years, and they have gradually changed people's purchase standards and consumption concepts of cars. The reason why people choose new energy vehicles is mainly because some people believe that new energy vehicles are an environmentally friendly product, which can effectively improve the urban environment and improve people's environmental awareness.

However, the main problems of new energy vehicles are also energy consumption, long charging times, and poor endurance. In addition, the charging piles must also be equipped in a designated place to charge new energy vehicles. But charging piles are not yet fully popular in some cities in China. On the highway, the number of charging piles is much less than the number of the gas station. Due to constant charging and discharging, the battery will also suffer a capacity loss. After reaching a certain service life, the battery capacity was greatly reduced in comparison to the new battery. However, the cost of replacing a battery for consumers is the same as replacing a car. This leads to range and range anxiety problems. The future direction of new energy vehicle manufacturers in the dimension of energy consumption is mainly to further improve battery technology and improve battery endurance.

4 Conclusion

With the rapid development of mobile terminals, smart phones and information networks, the number of online comments by users has skyrocketed. Based on the Word2Vec-SVM model, this paper first performed sentiment analysis and classified topic extraction in 46530 online user comments collected and found that users were generally dissatisfied with the performance of new energy vehicles in four aspects: interior, comfort, space, and energy consumption. Then, based on the LDA topic

model, the reasons for the negative sentiment are deeply analyzed. It is found that in terms of comfort, the reasons for the negative emotion are concentrated in chassis adjustment, vehicle noise, vehicle odor, rear space and seat design. In the interior, the users are not satisfied with the design details and layout style of the car, the grade of material of seat leather, steering wheel and center console, abnormal movement and sound. In the spatial dimension, the sizes of the trunk and the seats are the main problems. In terms of energy consumption, long charging events, poor endurance and scarcity of charging piles lead to problems such as range anxiety.Therefore, two suggestions are given for the improvement of these four dimensions to new energy vehicle manufacturers:

(1) Enhancing the battery life capacity of new energy vehicles and improving the convenience of charging go hand in hand. Battery life and charging convenience are important reasons for consumers to choose new energy vehicles. On the one hand, new energy vehicle manufacturers should enhance battery life; On the other hand, to improve traffic efficiency and solve the mileage problem, it is necessary to increase charging speed and convenience, such as building more high-power charging piles and improving charging efficiency. At the same time, increasing the energy density and capacity of the battery can also reduce dependence on charging equipment and congestion during the peak period, and the two complement each other to solve the charging problem of new energy vehicles. In short, the larger the battery capacity, the faster the charging speed and the lower the battery cost, which is crucial for the development of new energy vehicles.

(2) Pay attention to user demands and improve the driving experience of new energy vehicles. Compared to traditional fuel vehicles, the ride experience of new energy vehicles still has a lot of room for improvement. Therefore, new energy vehicle manufacturers must pay attention to consumers' feedback on the driving experience of new energy vehicles, such as this paper found that consumers mentioned seats in comfort, space and interior. New energy vehicle manufacturers need to better meet customer requirements in seat design, leather materials, and space issues. Therefore, manufacturers of new energy vehicles must pay more attention to user driving feelings and make the corresponding adjustments in a timely manner to promote the healthy development of new energy vehicles.

Although this paper has achieved certain results in the study of online comments mining of new energy vehicles, there are still many shortcomings and areas that can be improved, which need to be perfected in future research work. For example, the classification of sentiment categories can be further refined, the specificity of research samples is still relatively strong, the accuracy of the experiment needs to be improved, and the use of information such as time and social relationships in user comments has not been fully considered, which needs to be further expanded and deepened in subsequent studies.

Funding

This work was supported by the National Natural Science Foundation of China under Grant 72272039.

Author contributions

The authors contributed equally to this work and should be considered co-first authors.

Conflict of interest

The authors declare no conflict of interest.

References

[1] Aci, M.; Yergök, D. (2023). Demand Forecasting for Food Production Using Machine Learning Algorithms: A Case Study of University Refectory, *Tehnicki vjesnik–Technical Gazette*, 30 (6), 1683–1691, 2023. https://doi.org/10.17559/TV-20230117000232

[2] Blei D. M.; Ng A. Y.; Jordan M. I. (2003). Latent Dirichlet Allocation, *Journal of Machine Learning Research*, 3, 993–1022, 2003.

[3] Cai, M.; Tan, Y.; Ge, B.; Dou, Y.; Huang, G.; Du, Y.(2022). PURA: A Product-and-User Oriented Approach for Requirement Analysis From Online Reviews, *IEEE Systems*, 16(1), 566–577, 2022.

[4] Chen, Y.; Wu, X.; Liao, H.; Kou, G.(2023). Consumer preference disaggregation based on online reviews to support new energy automobile purchase decision, *Procedia Computer Science*, 1485–1492,2023.

[5] Cheung, C.M.K.; Thadani D.R.(2012). Impact of electronic word-of-mouth communication: A literature analysis and integrative model, *Decision Support Systems*, 54(1), 461–470, 2012.

[6] Du, Y.; Wei, K.; Wang, Y.; Jia, J.(2022). New energy vehicles sales prediction model combining the online reviews sentiment analysis: A case study of Chinese new energy vehicles market, *Proceedings of the 3rd International Conference on Artificial Intelligence in China*, 424–431, 2022.

[7] Gao, R.; Yao, X.; Wang, Z.; Abedin, M.Z.(2024). Sentiment classification of time-sync comments: A semi-supervised hierarchical deep learning method, *European Journal of Operational Research*, 314(3), 1159–1173, 2024. https://doi.org/10.1016/j.ejor.2023.11.035

[8] He, S.; Wang, Y.(2023). Evaluating new energy vehicles by picture fuzzy sets based on sentiment analysis from online reviews, *Artificial Intelligence Review*, 56, 2171—2192, 2023.

[9] Jia, S.; Wu, B.(2018). Incorporating LDA based text mining method to explore new energy vehicles in China, *Artificial Intelligence Review*, 6, 64596–64602, 2018. https://doi.org/10.1109/ACCESS.2018.2877716

[10] Ma, Y.; Chen, G.; Wei, Q.(2017). Finding users preferences from large-scale online reviews for personalized recommendation, *Electronic Commerce Research*, 17, 3—29, 2017. https://doi.org/10.1007/s10660-016-9240-9

[11] Mikolov, T.; Chen, K.; Corrado, G.; Dean, J.(2013). Efficient estimation of word representations in vector space, *Computer Science*, 2013. https://doi.org/10.48550/arXiv.1301.3781

[12] Pang, B.; Lee, L.; Vaithyanathan, S.(2002). Thumbs up: sentiment classification using machine learning techniques, *Proceedings of the ACL-02 conference on Empirical methods in natural language processing*, 79–86, 2002.

[13] Rus V; Niraula N; Banjade R.(2013). Similarity measures based on latent dirichlet allocation, *Proceedings of the International Conference on Computational Linguistics and Intelligent Text Processing*, Heidelberg: Springer, 459–470, 2013.

[14] Song, Fengxi; Zhang David; Wang, Jizhong; Liu, Hang; Tao, Qing(2007). A parameterized direct LDA and its application to face recognition, *Neurocomputing*, 71(1-3), 191–196, 2007. https://doi.org/10.1016/j.neucom.2007.01.003

[15] Su, B.; Peng, J.(2023). Sentiment analysis of comment texts on online courses based on hierarchical attention mechanism, *Applied Science-Basel*, 13(7), 4204, 2023. https://doi.org/10.3390/app13074204

[16] Tong, R.M.(2001). An operational system for detecting and tracking opinions in on-line discussion, *Proceedings of the ACM SIGIR Workshop on Operational Text Classification*, 2001.

[17] Wang, M.; You, H.; Ma, H.; Sun, X.; Wang, Z.(2023). Sentiment analysis of online new energy vehicle reviews, *Applied Sciences*, 13, 8176, 2023.

[18] Wu, Yang Andrew; Ng, Artie W.; Yu, Zichao Yu; Huang Jie; Meng Ke(2020). A review of evolutionary policy incentives for sustainable development of electric vehicles in China: Strategic implications, *Energy Policy*, 148, 111983, 2021.

[19] Xiao, S., Wei, C.P., Dong, M.(2016). Crowd intelligence: analyzing online product reviews for preference measurement, *Information and Management*, 53(2), 169–182, 2016.

[20] Xuan, W., Deng, M. (2023). Logistics Service Quality Sentiment Analysis with Deeper Attention LSTM Model with Aspect Embedding, *Tehnicki vjesnik–Technical Gazette*, 30 (2), 634–641, 2023. https://doi.org/10.17559/TV-20221018031450

[21] Yuan, X.L.; Liu, X.; Zuo, J.(2015). The development of new energy vehicles for a sustainable future: A review, *Renewable and Sustainable Energy Reviews*, 42, 298–305, 2015.

[22] Zuo, Wangmeng, Wang Kuanquan, Zhang David, Zhang Hongzhi(2007). Combination of two novel LDA-based methods for face recognition, *Neurocomputing*, 70(4-6), 735–742, 2007. https://doi.org/10.1016/j.neucom.2006.10.008

Journal's webpage: http://univagora.ro/jour/index.php/ijccc/

This journal is a member of, and subscribes to the principles of,
the Committee on Publication Ethics (COPE).
https://publicationethics.org/members/international-journal-computers-communications-and-control

Cite this paper as:

Hu, Feng; You, Zhaohan; Cai, Junyuan. (2025). An Approach of Sentiment–Topic Mining Based on User Online Comments on New Energy Vehicles, *International Journal of Computers Communications & Control*, 20(6), 6934, 2025.
https://doi.org/10.15837/ijccc.2025.6.6934