WEP Warwick Evans Publishing

International Journal of Social Sciences and Public Administration

ISSN: 3005-9836 (Print), ISSN: 3005-9585 (Online) | Volume 4, Number 1, Year 2024
DOI: https://doi.org/10.62051/ijsspa.v4n1.21
Journal homepage: https://wepub.org/index.php/IJSSPA/index

Research on Automobile Product User Experience based on Semantic Analysis

Hongjing He[1,2,*], Li Yang[1,2], Wei Shang[1,2]

[1]China Automotive Technology And Research Center Co., Ltd., China
[2]China Auto Information Technology (Tianjin) Co., Ltd., China
[*]Corresponding author: hehongjing@catarc.ac.cn

ABSTRACT

The advantages and disadvantages of automobile products are related to the fate of automobile enterprises, in which user experience is very important in the design of automobile products. At present, most of the research on the user experience of automotive products focuses on the analysis of user behavior information, and mining the user experience characteristics from online reviews is a new direction. The first is to determine the data source and preprocess the data source; The second is to use word segmentation technology, word vector technology and Word2Vec model to build the automobile product feature word database; The third is to use factor analysis method to screen out the most representative product feature words; The fourth is to use syntactic dependency to extract the views of user comments; The fifth is the user experience analysis based on the sentiment dictionary in the automobile industry. To a certain extent, the method proposed in this paper satisfies the user experience method to provide certain reference opinions for automotive product design.

KEYWORDS

User Experience; Word Segmentation; Word Vector; Word2Vec; Factor Analysis.

1. BACKGROUND INTRODUCTION

In the highly competitive automobile market, good user experience can become an important means of brand differentiation competition. Through in-depth understanding of user needs, design automobile products that meet user habits and improve driving convenience and comfort, so as to stand out among many competitors. Excellent user experience can stimulate potential users' desire to buy and improve the market appeal of the product. When buying a car, users not only pay attention to the performance parameters of the vehicle, but also pay more attention to the actual use experience, including the ease of driving and the comfort of the interior. For example, Xie Xuemei[1] and others focus on in-depth analysis of user behavior information, especially the relatively new research direction of mining user experience characteristics from online reviews. By manually setting specific extraction rules, they can effectively extract product features and the corresponding user emotional views from reviews, and use affective computing technology to quantify these views, so as to obtain detailed user experience features. The results show that the method successfully solves the problem that product features and emotional views cannot be extracted synchronously. In addition, the study also found that users showed a high preference for the intelligent voice interaction function of the product, but their overall satisfaction with the intelligent function was relatively low. Based on these findings, they put forward specific optimization design suggestions for the intelligent system, appearance design and hardware configuration of the product. Zhan Beiling[2] et al. systematically sorted out and analyzed the existing research results and development track from the two dimensions of subjective emotion measurement and objective data recognition of user experience, and summarized the measurement methods and reasonable evaluation methods suitable for the field of product design. This paper summarizes the current measurement characteristics of user experience and a variety of measurement methods, including PAD affective scale based on subjective emotion, Positive and negative Affective Scale (PANAS) and SAM scale, as well as objective data based on eye tracking detection, facial expression recognition, skin electrical response, EEG and ECG emotion recognition methods. This paper further analyzed the advantages and limitations of various measurement methods, summarized the current problems, and prospected the future research direction. The conclusion of the study is that user experience measurement can reveal the "cognitive friction" between products and users, that is, the place where products fail to meet user expectations, and then provide a basis for improving product quality. This suggests that user experience plays a crucial role in product design.

The design, function, appearance and other characteristics of automotive products have a direct impact on user experience. Meanwhile, users' emotional needs and preferences will in turn affect the product design and development process. Therefore, in-depth analysis of users' emotional views and obtaining user experience information about product attributes can provide an important basis for product optimization design. In addition, dependency parsing can reveal the dependency relationship between the components of a sentence and show the semantic modification relationship between the words in the text. It is not limited by the physical location of the words, and can reflect the semantic collocation relationship of the words. Therefore, compared with window-constrained extraction methods, dependency parsing is more flexible in mining product features and their related emotional perspectives from users' online comments.

As can be seen from the above, user experience plays a crucial role in automotive product design. It is not only related to the market competitiveness of the product, user satisfaction and brand loyalty, but also an important driving force to promote product innovation and upgrading. Therefore, car manufacturers should attach great importance to user experience design and constantly optimize product performance and service levels to meet users' growing personalized needs. To sum up, this paper proposes a method of user experience research of intelligent human-computer interaction products based on semantic analysis.

2. RESEARCH METHODS

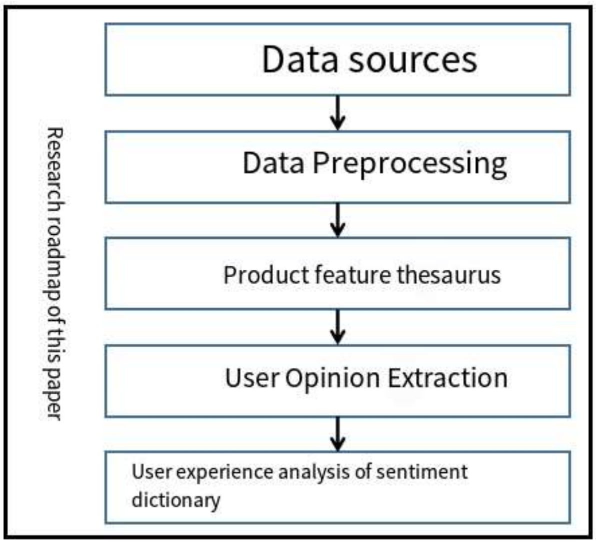

Figure 1. Research roadmap of this paper

The research roadmap of this paper is divided into 5 steps. The first step is the data source, mainly explaining the data source; The second step is data preprocessing, which preprocesses data based on text comment statements; The third step is the product feature lexicon, which classifies the automobile product feature lexicon based on the data after data preprocessing; The fourth step is to extract the user's view; The fifth step is the user experience analysis based on the sentiment dictionary in the automobile industry. The research roadmap of this paper is shown in Figure 1.

2.1. Data Sources

Intelligent connected vehicles are products that integrate various advanced technologies such as artificial intelligence, Internet of Things and big data, and their development helps to promote technological innovation and application and enhance national scientific and technological strength. Therefore, this paper studies the user experience of intelligent vehicles. Auto portal website is an online platform that provides auto news, evaluation, shopping guide, quotation, community communication and other services, which has important reference value for car lovers and car buyers. Therefore, the data sources of this paper mainly include Auto Home, Easy car network, Aika Car and Pacific Auto network. The web addresses of each automobile website are shown in Table 1 below.

Table 1. The urls of different automotive websites

Serial number	Car website name	website
1	Autohome	http://www.autohome.com.cn
2	Bitauto	http://www.bitauto.com
3	Aika Car	http://www.xcar.com.cn
4	Pacific Motors	http://www.pcauto.com.cn

2.2. Data Preprocessing

There are many non-standard expressions in the original text data of the above website, which will have a great impact on the subsequent grammatical analysis. The original text data contains "very good", "very great" and other meaningless comments for this study, so it is necessary to preprocess the original data, including deleting meaningless comments, deleting duplicate values, standardizing comment symbols and other operations. After the pre-processing, a total of more than 1000 comments were obtained as the data set for the follow-up research.

Crawling technology was used to crawl the relevant automobile website portal, and the comments after data preprocessing were shown in Table 2 below (example).

Table 2. Inter-car comment statements after data preprocessing

Serial number	Statements
1	The interior space is very spacious and the front seats are very comfortable.
2	Driving on crowded roads in urban areas, the power performance is very good, and can easily cope with traffic congestion and vehicle congestion.

2.3. Product Feature Lexicon

This paper is built on Word2Vec's product feature lexicon, which is an unsupervised, predictive deep learning-based model primarily used to compute and generate high-quality, distributed and continuous dense vector representations of words to capture contextual and semantic similarity. The model can absorb a large corpus of text, create a possible vocabulary, and generate a dense word Embedding for each word in the vector space representing that vocabulary.

2.3.1. Introduction to Word Vectors

Word vector is the transformation of the human language into the language that the computer can understand. This process is to transform the symbolic information of language into numerical information in the form of vector, that is, to transform the natural language problem into a machine learning problem. The mainstream word vectors mainly include the single heat coding model and the distributed representation model.

(1) unique heat coding

Unique heat coding: is to use a long vector to represent a word, the vector length is the size of the dictionary N. The value of each vector is 1, representing the position of the word in the dictionary, and all the other dimensions of the vector are 0.

(2) Distributed representation model

Distributed representation: its core idea is that the semantics of words are determined by context information, that is, words in the same context have similar semantics. In other words, the biggest contribution of this model is to bring related or similar words closer in distance.

2.3.2. Word2Vec Model

Word2Vec means exactly what its name suggests: turning words into vectors. The process is roughly to map word into a new space and represent it as a multi-dimensional continuous vector of real numbers, which is called word embedding.

The samples trained in a machine learning or deep learning model generally have an input value x and a target value y, and the mapping of f(x) to y is mainly constructed when the model is trained. However, the input value and target value of the language model are different from the traditional model. The language model treats x as a word in a sentence, and y is the context word of this word. The model f(x) is constructed, and then the sample (x,y) is judged whether it conforms to the laws of natural language.

Word2vec as a language model, although his ultimate goal is not to train a perfect, but he cares about the model after the training of the parameters, and these parameters as the input x of some vectorized representation, which is the word vector we mentioned above.

2.3.3. Product feature lexicon

First, in order to ensure the accuracy and completeness of the segmentation, it is necessary to add a user-defined dictionary and use the Jieba segmentation tool to segment and partof speech tagging the review text. Next, words that can represent the features of the product need to be screened from the reviews. This paper uses factorization method to select the most representative product feature words. The steps of factor analysis are shown below[3].

(1) The basic equations are established by calculating the observed data $\overline{x_k}, s_k$;

(2) The eigenvalue j is obtained from the correlation coefficient matrix R, j=1,2,... m and the variance contribution, contribution rate and cumulative contribution rate of each common factor, and according to the cumulative contribution rate, the number of common factor retention N is determined;

(3) The factor load matrix A was determined by principal component analysis;

(4) Maximum orthogonal rotation of variance, the variable coefficient is institutionalized (try to approach 0 or 1);

(5) Get the factor score function and calculate the sample factor score.

In this paper, it is only necessary to determine the variance contribution, contribution rate and cumulative contribution rate of each common factor, and determine the number N of common factor retention according to the cumulative contribution rate, that is, only need to go to the second step.

2.3.4. Specific Operation Steps of Factorization

Suppose that there are m short comment statements for factor analysis of the interior indicators, and the maximum number of words in each sentence is n, that is, the size of the word matrix of the interior indicators can be m×n, because the matrix is a word matrix, and the word matrix needs to be numerized here.

Unique heat coding is also known as one-bit effective coding, mainly using n-bit status register to encode N states, each state by his independent register bit, and at any time only one effective. Unique thermal coding is the representation of categorical variables as binary vectors. This first requires mapping categorical values to integer values. Each integer value is then represented as a binary vector with a zero value except for the index of the integer, which is labeled 1[3-4].

So for a word matrix of size m×n, the size of the matrix that is thermally encoded is m×n,

$$X = \begin{bmatrix} x_{1,1}, x_{1,2}, \cdots, x_{1,mn} \\ x_{2,1}, x_{2,2}, \cdots, x_{2,mn} \\ \vdots \quad \vdots \quad \vdots \quad \vdots \\ x_{m,1}, x_{m,2}, \cdots, x_{m,mn} \end{bmatrix}$$

For each element $x_{ik}, i = 1, \cdots, m; k = 1, \cdots, mn$, which in this case is a sparse matrix. After making a standardized transformation here, the transformation of x_{ik} is x_{ik}^*, that is, the transformation formula is: $x_{ik}^* = \frac{x_{ik} - \overline{x_{ik}}}{s_{ik}}$, where the mean of x_{ik} is $\overline{x_{ik}}$, the standard deviation of x_{ik} is s_{ik} [23], where $\overline{x_k} = \frac{1}{m}\sum_{i=1}^{m} x_{ik}$, $s_{ik} = \sqrt{\frac{1}{m}\sum_{i=1}^{m}(x_{ik})^2}$.

Common factors in factor analysis are common factors that cannot be directly observed but exist objectively. Each variable can be expressed as the sum of linear functions of common factors and special factors, namely:

$$X_i = \mu_i + \alpha_{i1}F_1 + \alpha_{i2}F_2 + \cdots + \alpha_{im}F_m + \varepsilon_i, (i = 1, 2, \cdots, p)$$

Among them, $F_1, F_2, \cdots, F_m$ is called common factor, ε_i is called special factor, the model can be expressed as a matrix:

$$X - \mu = AF + \varepsilon$$

Here:

$$X=\begin{bmatrix} x_1^* \\ x_2^* \\ M \\ x_3^* \end{bmatrix}, A=\begin{bmatrix} a_{1,1} & a_{1,2} & A & a_{1,m} \\ a_{2,1} & a_{2,2} & A & a_{2,m} \\ A & A & A & A \\ a_{p,1} & a_{p,2} & A & a_{p,m} \end{bmatrix}, F=\begin{bmatrix} F_1 \\ F_2 \\ M \\ F_m \end{bmatrix}, \varepsilon=\begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ M \\ \varepsilon_3 \end{bmatrix}$$

And to meet and several points:

(1) $m \le p$;

(2) $Cov(F,\varepsilon)=0$,that is, common factors are not related to special factors.

(3) $D_F = D(F) = \begin{bmatrix} 1 & & 0 \\ & 1 & \\ & o & \\ 0 & & 1 \end{bmatrix} = I_m$, that is, the common factors are not correlated and the variance is 1.

(4) $D_t = D(\varepsilon) = \begin{bmatrix} \sigma_1^2 & & 0 \\ & \sigma_2^2 & \\ & & \\ 0 & & \sigma_F^2 \end{bmatrix}$, that is, each special factor is not correlated, and the variance is not required to be equal.

Matrix A in the model is called factor load matrix, a_{ij} is called factor load, which represents the load of the i th variable on the j th factor, if the variable X_i is regarded as a point in m-dimensional space, a_{ij} represents its projection on the coordinate axis F_j.

2.4. Extraction of User's View

The syntax analysis based on the statement extracts the user's point of view, extract the user's point of view is divided into two steps. The first step is to formulate specific syntactic dependencies and their combinations as extraction criteria. These dependencies include subject-predicate relationship (SBV), verb-object relationship (VOB), deterministic neutral relationship (ATT), in-form structure (ADV), and active-complement structure (CMP), where nouns, adjectives, and verbs are denoted by n, a, and v respectively, and → symbol is used to indicate dependencies. In the second step, based on these extraction criteria and the constructed product feature lexicon, product features and their related user emotional perspectives are extracted from the review data. In doing so, consider not only adjectives as opinion words, but also verbs such as "fall in love" and "like" as opinion words that express the user's emotions.

In order to improve the accuracy and comprehensiveness of the extraction results as much as possible, we combine partial dependencies to form new extraction rules. For example, the comment "the space inside the car is good" expresses the user's opinion as "the space is good".

2.5. User Experience Analysis based on the Emotional Dictionary in the Automotive Industry

Quantitative processing of users' views can make it more clear and intuitive to understand users' emotional tendency towards product attributes. To this end, the sentiment dictionary can be used to calculate the sentiment of users' opinions, and the user's experience satisfaction can be reflected by calculating the emotion score of product features.

Aiming at the specific field of automobile products, and recognizing the basic role of emotion dictionary in emotion calculation, we construct a negative word dictionary containing 70 words and a degree adverb dictionary containing 214 words with reference to the HowNet dictionary, and mark each degree adverb in detail according to 6 intensity levels[1].

The emotional polarity and intensity of the user's point of view can be calculated, and the specific calculation process is as follows. The emotional polarity and intensity of the extracted user's view wi (i is a positive integer) in the sentiment dictionary in the field of automotive product seats[1] are denoted as pwi and swi, respectively, and the emotion score of each wi is denoted as Si.

Given the differences in the number of feature words included under each product feature class, simply summing up the emotion scores to represent the final scores of each class may lead to inaccurate results. Therefore, we adopted the method of averaging to calculate the mean emotional scores of various product features, as shown in formula (1).

$$Ej=sum(Si)/x \quad i=1,...,x \tag{1}$$

Analysis of formula (1) : j represents the JTH product feature class; x represents the number of feature words contained in the JTH product feature class, and x and j are positive integers.

3. RESULTS AND ANALYSIS

The method proposed in this paper to extract product features and matching emotional views from user online reviews can effectively solve the problem that product features and emotional views cannot be extracted at the same time. In addition, this paper obtains user experience information by mining online reviews, enriching relevant research on automotive product user experience. However, this paper also has limitations: the research only takes the car seat product as the representative, and does not involve the characteristic indicators of other automotive products, so the application scope of the proposed method is limited to a certain extent. In the future, more research objects will be added to carry out more extensive research. In addition, since the implicit product feature words in the review text are not considered, the robustness of the proposed method is low, and the extraction of implicit product feature words will also be considered in the future.

REFERENCES

[1] Xie Xuemei, Wang Minghui. Research on User Experience of intelligent Human-Computer Interaction products based on Dependency syntax analysis. Digital Library Forum, 2024, 20 (06): 44-53.

[2] Zhan Beiling, Lu Jing. A review of research on emotion measurement and emotion Recognition in User Experience [J]. Packaging Engineering, 2024, 45 (12): 12-21. DOI:10.19554/j.cnki.1001-3563.2024.12.002.

[3] https://blog.csdn.net/qq_29831163/article/details/88908811

[4] Xiangguang Dai,Nian Zhang,Keke Zhang,Jiang Xiong. Weighted Nonnegative Matrix Factorization for Image Inpainting and Clustering[J]. International Journal of Computational Intelligence Systems,2020.

[5] Ding Yulian,Wang Fei,Lei Xiujuan,Liao Bo,Wu Fang-Xiang. Deep belief network-Based Matrix Factorization Model for MicroRNA-Disease Associations Prediction.[J]. Evolutionary bioinformatics online, 2020,16.

[6] Guangfu Chen,Chen Xu,Jingyi Wang,Jianwen Feng,Jiqiang Feng. Nonnegative matrix factorization for link prediction in directed complex networks using PageRank and asymmetric link clustering information[J]. Expert Systems With Applications, 2020,148.