J. COMBIN. MATH. COMBIN. COMPUT. 127b (2025) 2417--2437

Journal of Combinatorial Mathematics
and Combinatorial Computing
www.combinatorialpress.com/jcmcc

Combinatorial Press

Design of Computational Identification Model and Optimization Strategy for Consumer Demand Based on Multimodal Data Fusion

Tianyi Yu[1, ✉]

[1] *Zhejiang Business College, Hangzhou, Zhejiang, 310000, China*

ABSTRACT

This paper draws a framework for constructing user demand modal information, uses crawler technology to obtain online review text information, processes the text information, and mines relevant consumer demand information. The LDA topic model is used to extract the topics of consumer concern from the online comments, identify the topics of consumer demand and clarify the concern degree of each demand. The KANO model is proposed to establish a consumer demand classification method based on the KANO model by combining product characteristic attributes and consumer demand information. Examine the theme discrimination performance of the LDA model on the hotel category, footwear category, and food category datasets. Combine the preprocessed user demand data to statistically quantify user demand for quantitative Kano transformation. Classify user demands into Kano categories and calculate the priority order of user demands to get the product optimization strategy. The weighted order of consumers' demands for automobiles is footrest, cigarette lighter, antenna, window, low beam, key, etc. in order. It can be found that automobile consumers pay more attention to the needs of antenna, cigarette lighter, pedals, and enhancement of accessory functions. As a result, automobile manufacturers should increase the seat comfort, improve power, enhance the flexibility of shifting such aspects of the whole vehicle handling experience, in addition to improving the lights, keys and other car quality related needs.

Keywords: LDA model, KANO model, consumer demand, online reviews

1. Introduction

A series of new business forms, new models, new platforms and new organizations have arisen in the context of the digital economy. Different regions and subjects continue to accumulate data production elements, different groups continue to obtain data elements, massive super data contains huge value, and the effective use of data elements will bring strong impetus to the digital development of the

✉ Corresponding author.
E-mail address: lynn118211@163.com (T. Yu).
Received 30 May 2024; Revised 15 August 2024; Accepted 10 January 2025; Published Online 16 April 2025.
DOI: 10.61091/jcmcc127b-137

economy and society [1-3]. The depth of the application of digital technology to enhance the competitiveness of enterprises, enterprises can use the real-time perception of big data to do accurate market analysis, so that the decision-making layer to fully grasp the market, to improve the scientific and accurate decision-making, enterprises to face the impact of the sudden epidemic [4-5]. Make full use of digital thinking to achieve cost reduction and efficiency, to establish accurate personalized customized marketing programs, to determine the appropriate high-quality customer base, avoid risk, adjust and optimize the brand marketing decision-making for their own empowerment, so as to achieve high-quality development of enterprises [6-8].

Emerging digital industry empowers all industries to digitalization, networking and intelligent transformation, digital industrialization promotes the deep integration of the digital economy and the real economy, new industries, new models continue to give rise to richer social data resources in the process of technological innovation [9-10]. Whether it is commercial data, industrial data, or collected data, computed and derived data, and application data in social behavior, they all present dynamic and diverse types of data sources around the full-cycle flow of information [11-13]. In the context of massive data in diverse dimensions, digital resources are inevitably an important strategic resource for enterprises. Digital social life enriches the data production elements, and when data is integrated into social life, it also changes the reality [14-15].

In recent years, digital consumption has opened up a new pattern of intelligent life for residents, which is defined as the practice of promoting and selling products and services through digital communities [16-17]. The national development plan clearly puts forward the digital China strategy, which aims to grasp the opportunities of the times, accelerate the development of the digital economy, give rise to new modes of industry and grow the new engine of the economy with digital transformation, and digital consumption has become an important productive force in today's innovation and leadership [18-20]. Research oriented towards the new features presented by the digital community helps to grasp the winds of consumer behavior and psychological changes, so that enterprises and related units are more fully integrated into the digital queue [21-22]. However, the development path of digital consumer activities that emphasizes channel costs, generalizes self-media publicity and neglects community maintenance also brings consumers the same negative impacts that cannot be ignored, such as homogenization of product content, false information and monotonization of retail models [23-24]. Affected by the same kind of problems, consumers show certain potential resistance, specifically focusing on the lack of group personalization, shallow level of content acceptance and lack of asset awareness [25]. The reason for this is that digital product suppliers do not really understand the consumer group, and the lack of a fine operation mode that is highly compatible with preferences makes it more difficult to meet the actual demand through digital community consumption activities, and how to carry out consumer-centric service optimization has gradually become a research hotspot [26-28].

Consumer demand has been the focus of research in the socio-economic field board, and the factors that affect consumer demand, and consumer demand for the economy is the impact of many concerns. Literature [29] carried out a survey practice for consumer consumption behavior, mindset research, found that the perceived value of consumer goods in the non-expression of the virtual image is highly correlated. Literature [30] describes the behavior of consumers using digital devices, promoting cultural connections between devices, between devices and markets, and new practices, which inadvertently creates new consumer demands and norms. Literature [31] explores the impact and influence of future uncertainty on consumption and investment, which exerts a volatile influence on economic activity mainly through price stickiness. Literature [32] based on the autoregressive integrated moving average (ARIMA) model to analyze and forecast the consumer demand for hydropower, and the introduction of the national GDP and population data for a comprehensive analysis, that the hydropower development planning for the country's economic development and progress in the significance.

Scholars have explored how to better analyze and predict consumer demand and consumer behavior, and then realize the reasonable allocation of marketing and production resources, focusing on multimodal data fusion analysis, mathematical functions, and intelligent algorithms. Literature [33] reviews and summarizes the research used for commodity consumption modeling, and analyzes the trend of commodity consumption analysis, and puts forward specific recommendations for commodity consumption modeling and consumption analysis. Literature [34] proposed a simple and efficient prediction algorithm based on the additive and divisible utility function and random utility strategy, and used it for residential energy consumption demand prediction, and verified the feasibility based on simulation experiments. Literature [35] conceived a consumer demand forecasting model that integrates multimodal data such as user geographic location, behavioral data, and product attributes, which has better performance in terms of accuracy and loss rate compared to traditional consumer demand forecasting methods. Literature [36] envisioned a multimodal probabilistic consumption prediction system based on relevant textual and visual data to predict the demand of consumer preferences, which was tested on the Amazon platform, and the prediction system can assist stakeholders in predicting consumer personalized consumption behaviors at a fine-grained level and from a multimodal perspective. Literature [37] segmented the multimodal continuous heterogeneous distribution of consumer preferences based on sparse learning theory, which improved the performance of the model in consumer heterogeneity analysis and assisted marketing. Literature [38] demonstrated and analyzed the positive significance of the integration and analysis of different datasets in aiding the understanding of the logic of consumer behavior of electronic goods, with video data being used to predict consumers of electronic products with the highest accuracy. Literature [39] utilizes a deep learning approach in order to integrate customer attention metrics and thus predict consumer hotel demand and consumption behavior and logic, which is also examined based on a real data set, confirming the superiority of the proposed approach.

This paper analyzes multimodal data sources, obtains online review text information through crawling techniques, and processes the text information to mine relevant consumer demand information. Preprocessing online review data, calculating the correlation between neighboring words identifies all fixed multi-word expressions and extracts product attribute features. Consumer attention topics are captured from online reviews using LDA topic modeling. A user demand classification method based on KANO model is proposed to calculate product feature weights using online review data as an example. The classification efficiency of LDA topic model on different datasets is examined. Combining word frequency analysis and LDA topic analysis, the main influencing factors of user satisfaction are extracted and summarized. And carry out satisfaction conversion to obtain the Kano categories of attributes corresponding to key needs, sort consumer demand priorities, and get product adjustment strategies.

2. Multimodal data fusion for consumer demand capture

In the field of information data, multimodal is often used to represent different forms of data forms. And multimodal data mainly refers to the data obtained through different domains or perspectives for the same described object, and each domain or perspective describing the data related to this object is called a modality. This paper argues that multi-source, multi-modal data compared to single-modal data, multi-modal data can provide more accurate information for the same description of the object through the mutual support, supplementation, and modification of the information of each module. Compared with the traditional single-modal information collection method, multimodal information recognition can analyze and judge the information of multiple dimensions of the same description object, so that the information fusion is more accurate [40].

Applying the analysis method of multimodal information data enables better understanding and interpretation of related events or phenomena. This integrated analysis can provide more accurate and comprehensive information support for decision-making, as well as stronger support for practical problem solving. By comprehensively analyzing multimodal data of the same descriptive object, it is possible to gain an in-depth understanding of the object's characteristics and reveal the correlations and trends hidden in each perceptual perspective, thus enhancing the grasp of the essence of the problem.

2.1. *Data sources and collection methods*

Users can express their opinions and suggestions at any time and any place through online comments without the limitations of geography and time, providing a variety of different perspectives and needs. By analyzing the text information of consumers' online comments is an effective way to obtain users' demands, which can help designers better grasp the market demand and users' needs, and provide powerful support for the optimization and innovation of products and services.

Based on the preliminary process of consumer demand information acquisition, crawler technology is used to obtain online comment text information, and the text information is processed to mine relevant consumer demand information. The preliminary preparation is to collect and organize and merge the user review text of the product, sift out the inefficient and useless review data, so as to ensure the validity and reasonableness of the review text. K-Means clustering analysis is used to process the cleaned text information to get the initial user demand text set. At the same time, considering the diversity of Chinese semantics, the processed text information is clustered and refined again through the focus group discussion method to determine the end-user demand information, and the user demand modal information construction process is shown in Figure 1.

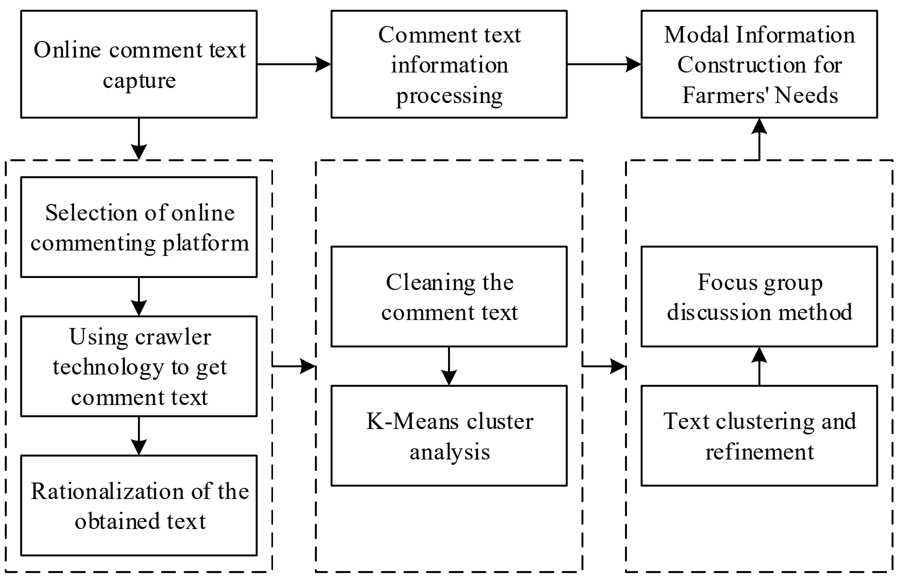

Fig. 1. User demand modal information build process

2.2. *Online comment preprocessing*

Unlike the English text of the words separated by spaces, the Chinese text is seamless articulation, the need for lexical processing. The current Chinese word separation tools mainly include “jieba”, “thulac”, “pkuseg”, “ltp” Different lexical tools may have different effects on the text in different fields.

In this paper, we select the research domain lexical segmentation package provided by “pkuseg” to split all the comment data into several phrases. In order to ensure the accuracy of word segmentation, the process of word segmentation was carried out to remove the deactivated words, including pronouns, auxiliaries and so on. Subsequently, a lexical tagging tool was used to label the lexical properties of each word after the splitting.

2.3. Feature extraction of product attributes based on online reviews

Product attribute feature extraction mainly includes the following steps: fixed multi-word expression recognition and screening, candidate product attribute feature thesaurus construction, online product attribute recognition, candidate product attribute feature word categorization and product attribute thesaurus construction.

In addition to a single noun, product attribute words may also contain some fixed multi-word expressions. Fixed multi-word expressions contain several words after word segmentation, but they can be regarded as a complete word, and such fixed multiword expressions cannot be accurately identified in the word segmentation process, including some attractions, hotel names or Chinese idioms, for example, "no choice" usually means that online consumers are very satisfied with a certain attribute, and the word will be split into four separate words in the word segmentation process. Looking at these individual words in isolation does not reveal their true semantics and misleads subsequent analysis. Given the implicit meaning of fixed multi-word expressions, it is valuable to construct a product attribute thesaurus containing fixed multi-word expressions in this study. Therefore, this chapter identifies all fixed multi-word expressions by calculating the degree of association between adjacent words. Specifically, the Information Entropy Algorithm (IE) and the Point Mutual Information Algorithm (PMI) are used to identify fixed multi-word expressions, which are often used to measure the correlation between words.

The Point Mutual Information algorithm quantifies whether a set of words can constitute a fixed multi-word expression by measuring the probability of the joint distribution of the words against the probability of their separate occurrences, a method that has been widely used in previous research. The definition of point mutual information is shown in Equation (1):

$$PMI\left(w_i, w_j\right) = \log_2 \frac{P\left(w_i, w_j\right)}{P\left(w_i\right)P\left(w_j\right)} \tag{1}$$

where $PMI\left(w_i, w_j\right)$ denotes the probability of words w_i and w_j occurring together, while $P\left(w_i\right)$ and $P\left(w_j\right)$ denote the probability of words w_i and w_j occurring alone, respectively. The larger the value of mutual information, the greater the correlation between the two.

The left-right information entropy algorithm is derived from the concept of “information entropy”. “Information entropy” describes the average amount of information contained in the received information after eliminating redundancy, and is usually used to represent the uncertainty of things. The definition of left and right information entropy is shown in equations (2) and (3):

$$IE_{left}(X) = \sum \left[P(PREX \mid X) * \log_2 P(PREX \mid X) \right] \tag{2}$$

$$IE_{right}(X) = \sum \left[P(XPOST \mid X) * \log_2 P(XPOST \mid X) \right] \tag{3}$$

where X denotes the analyzed phrase, and PRE and $POST$ denote the individual words adjacent to the left and right of X, respectively. $P(PREX \mid X)$ denotes the conditional probability that the left neighboring word is PRE when X occurs, and $P(XPOST \mid X)$ is the same. The left-right information entropy can be used to determine whether a word is a boundary word or not, and the larger the value of left-right information is, the more likely it is that the word is an independent word. In this chapter, the smaller value between the left information entropy and the right information entropy is selected and added to the point mutual information value to characterize the correlation between words. The higher the final value, the more likely it is to be a fixed multi-word expression. Both of these metrics have their own advantages and disadvantages, and combining the two scores mentioned above (i.e., PMI, IE) reduces the likelihood of extracting spurious collocations and better measures the degree of correlation between words.

2.4. Consumer demand information mining based on online reviews

LDA topic model is a generative probabilistic topic model that can be used to extract topics from a large number of documents [41-42]. In this paper, we use LDA topic model to extract topics of consumer interest from online reviews. A specific description of the LDA-based topic extraction method is given below. In LDA topic model, the perplexity value reflects the merit of the topic model, the smaller the perplexity value, the better the performance of topic extraction, the perplexity value can be calculated by equation (4), i.e.:

$$Perplexity(R) = \exp\left\{-\frac{\sum_{n=1}^{N}\log P\left(WD_n\right)}{\sum_{n=1}^{N} D_n}\right\} \tag{4}$$

where R denotes the online comment corpus with a total of N customer online comments. D_n is the number of words in online review r_n, WD_n is the sequence of words corresponding to review r_n, and $P(WD_n)$ is the probability of word generation in the review.

In LDA topic modeling, the consistency value can be used to assess the quality of the topic. A larger consistency value indicates a higher semantic consistency of the topic words, implying that the topic is more meaningful and semantically coherent. Specifically, given topic f_j, the V most relevant words to the topic are denoted as $W^{f_j} = \left\{w_1^{f_1}, w_2^{f_2}, \cdots, w_1^{f_j}\right\}$, the consistency value of the topic is calculated as:

$$C\left(f_j; V^{f_j}\right) = \sum_{v=2}^{v}\sum_{v=1}^{v}\log\frac{U\left(w_v^f, w_v'^f\right)+1}{U\left(w_v^f\right)} \tag{5}$$

Where $U\left(w_v^{f_j}\right)$ denotes the word frequency of word $w_v^{f_j}$ in the document (review) and $U\left(w_v^f, w_v'^f\right)$ denotes the number of times w_v^f and $w_v'^f$ co-occur in the document (review). The implementation flow of the LDA based topic extraction method is shown in Fig. 2.

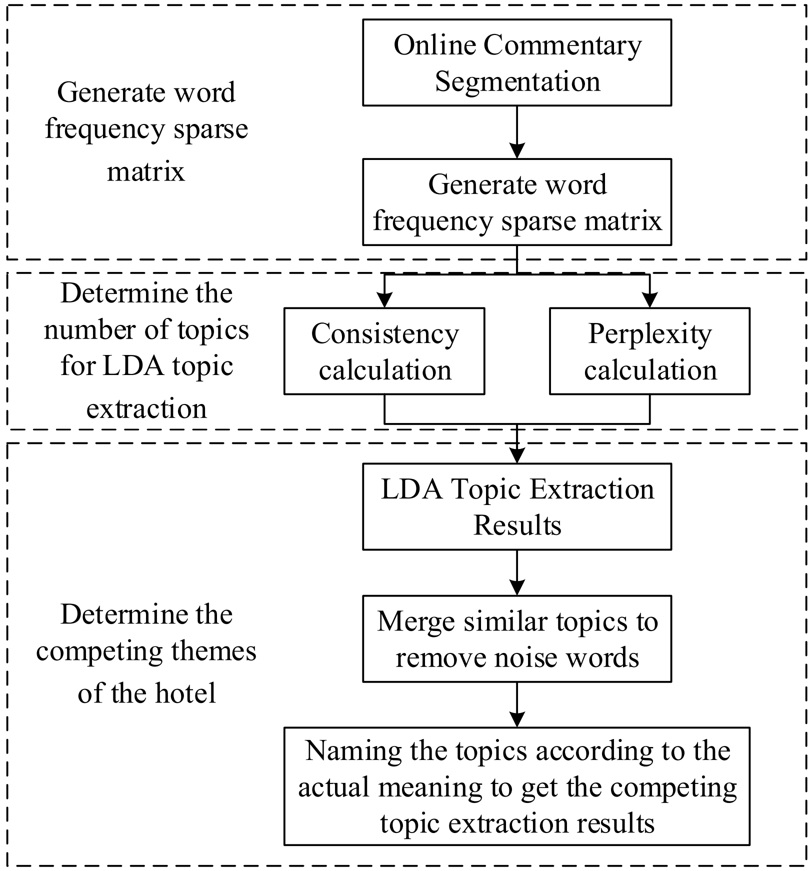

Fig. 2. The implementation process of the topic extraction method based on LDA

In this paper, the gensim library of Python is used to realize LDA-based online review topic extraction, and the consumer demand topic extraction results are obtained after the above steps. Finally, summarize the meaning of the corresponding topic words for each topic and name each topic according to the actual meaning of demand topic extraction to get the set of topics that consumers pay attention to, denoted as $F=\{f_1,f_2,\cdots,f_J\}$, where f_j denotes the j rd demand topic, $j=1,2\ldots,J$. Since each topic is actually a collection of related words, demand topic f_j corresponds to a collection of topic words $W_j^F=\{w_j,w_{j2},\ldots,w_{jp_j}\}$, where w_{jp} denotes the p th topic word corresponding to demand topic f_j, and $j=1,2,\ldots,J$, $p=1,2,\ldots,P_j$, P_j are the number of topic words corresponding to demand topic f_j.

2.4.1. Consumer demand topic sentence identification. The specific steps for consumer demand topic sentence identification are as follows:

1) Sentence segmentation for consumer online reviews. According to the characteristics of online comments and Chinese notation habits, the use of ",", ".", "?", "!" and other punctuation marks are utilized to clause each online comment in the online comment collection $R_i=\{r_1,r_{i2},\cdots,r_{iQ}\}$ of the consumer H_i, i.e., the text content located between any two punctuation marks is taken as a clause of the online comment of the consumer.

The set of online comment sentences $S_i=\{s_{i1},s_{i2},\cdots,s_{iE_i}\}$ of consumer H_i is obtained by clause clauses, noting $SW_{ie}=\{SW_{ie}^1,SW_{ie}^2,\ldots,SW_{ie}^{Zi}\}$ as the set of words contained in sentence s_{ie}. Where SW_{ie}^2 denotes the z th word in the e th comment sentence of consumer H_i, $i=0,1,2,\ldots,I$, $e=1,2,\ldots,E_i$, $z=1,2,\ldots,Z_i^e$, Z_i^e are the number of words corresponding to the e th comment sentence of consumer H_i.

2) Identify the demand topic to which each comment sentence is directed. For each sentence s_{ie} in

the set of online comment sentences of the consumer H_i, a matching relationship is determined between the set of topic words $W_j^F = \{w_{j1}, w_{j2}, \cdots, w_{jf}\}$ corresponding to the demand topic f_j and the set of words $SW_{ie} = \{SW_{ie}^1, SW_{ke}^2, \ldots, SW_{ik}^{Zi}\}$ corresponding to the comment sentence s_{ie}, and if $W_i^F \cap SW_{ie} \neq \varnothing$, the sentence s_{ie} is 1 of the online comment sentences of the consumer H_i that addresses the demand topic f_j.

3) Obtain a collection of sentences for each demand topic. Looping step 2, iterating through each sentence in the collection of sentences S_i of the online comments of the consumer H_i, identifying all sentences of the consumer H_i on the demand topic f_j by judging the matching relationship between the collection of topic words and the collection of words corresponding to the sentences, and obtaining the collection of sentences $FS_{ij} = \{FS_{ij}^1, FS_{ij}^2, \cdots, FS_{ij}^{G_i^j}\}$, wherein FS_{ij}^g denotes the g th sentence of the online comments of the consumer H_i against the demand topic f_j, and $i=0,1,2,\ldots,I$, $j=1,2,\ldots,J$, $g = 1, 2, \ldots, G_i^j$, G_i^j are the number of sentences of the consumer H_i against the the number of sentences of the demand topic f_j.

2.4.2. Attention Determination of Consumer Needs. Attention to consumer demand themes refers to the level of customer attention and importance given to the demand theme. Multiple consumer demand themes may be mentioned in online reviews, and the more times a theme is mentioned, the higher customer attention it has. Customer attention is an important factor to consider when developing a competitiveness enhancement strategy, therefore, this section gives a methodology for determining customer attention.

First, the frequency with which demand topic f_i is mentioned in the online reviews of consumer H_i is determined $Freq_{ij}$. In this paper, the frequency with which demand topic f_j is mentioned in the online reviews of consumer H_i is obtained by counting the number of sentences G_i^j in the set of sentences $FS_{ij} = \{FS_{ij}^{\text{i}}, FS_{ij}^2, \cdots, FS_{ij}^{\sigma_f{}'}\}$ targeting demand topic f_j in the online reviews of consumer H_i, viz:

$$Freq_{ij} = G_i^j, i=0,1,2,\ldots,I, j=1,2,\ldots,J \tag{6}$$

Then, the ratio of the frequency of mentions of demand topic f_j in the online reviews of consumer H_i to the frequency of mentions of all topics is calculated, and this is used as the customer attention level of consumer H_i with respect to demand topic f_j. Consumer H_i customer attention level AT_{ij} for demand topic f_j can be determined by equation (7), i.e.:

$$AT_{ij} = \frac{Freq_{ij}}{\sum_{j=1}^{J} Freq_{ij}}, i = 0,1,2,\ldots,I, j = 1,2,\ldots,J \tag{7}$$

where $\sum_{j=1}^{J} Freq_{ij}$ denotes the sum of the frequency of mentions of the J demand topics in consumer H_i online reviews.

3. Classification of user needs based on the KANO model

3.1. Kano model

The Kano model, which can classify product functional attributes based on user feedback, is an evaluation method that makes the invisible attributes of service quality visible. The model categorizes product quality attributes into five categories: attractive quality (A), one-dimensional quality (O), must-have quality (M), undifferentiated quality (I) and reverse quality (R).

As a qualitative analysis method, Kano model is a crucial entry point for identifying users' key requirements and assisting designers in product function planning. The use of Kano model in interface design research can help designers and developers to deeply explore user needs and accurately identify the functions or services that users value and expect to realize. It also realizes reasonable resource allocation in the design and development process, which improves user satisfaction and avoids the waste of design and development resources.

One-dimensional quality is the quality attribute that has a very high degree of fit with user expectations. The higher the degree of realization of this type of quality attribute, the higher the user's satisfaction, and the lower the degree of realization, the lower the user's satisfaction. Typically, these are requirements that the user can understand and accurately describe.

Required quality, is the quality attribute that must be present to make users satisfied. The degree of optimization of the pair does not result in a significant increase in the level of satisfaction, similar to Herzberg's two-factor theory of health care factors. When this type of quality attribute is not provided, the user's satisfaction will plummet or even quit using the product.

Enchantment quality is a quality attribute that exceeds the user's expectations, and this type of quality attribute surprises the user by satisfying the user's latent needs and exceeding the user's expectations. This type of quality does not lead to dissatisfaction when it is not provided, but if it is provided, it leads to a significant increase in user satisfaction. Therefore, this type of attribute is easy to please the user and can give the user a better experience when using the product and increase user loyalty.

Undifferentiated quality is a quality attribute that is irrelevant to the user. This type of quality attribute does not have a large impact on user satisfaction, or refers to requirements that have little impact on satisfaction regardless of functional performance.

Reverse quality, are quality attributes that cause strong dissatisfaction and lead to low levels of satisfaction. Reverse quality is less important compared to the other four quality attributes, so the impact of reverse quality is usually not considered when applying the Kano model for classification.

3.2. Product Characterization

The prerequisite of user demand classification is to obtain the user's demand, and the user's demand for cell phones is usually responded by product features, i.e., consumers satisfy their own intrinsic needs by choosing cell phones that contain various product features. Generally, enterprises will determine the product features that can be clearly recognized by consumers based on internal and market research, and then develop corresponding questionnaires to collect consumers' opinions on the determined product features. This paper, however, proposes a user demand acquisition method based on online reviews, which not only improves the amount of data, the timeliness of data, etc., but also avoids the subjectivity of enterprises in selecting product features, because the mined product features are actively shared by consumers after using the products.

3.3. Feedback collection

The feedback information in this article refers to the opinions expressed by consumers on the feature set of mobile phone products, and the traditional market research usually adopts the five-point Likert scale method, that is, consumers are required to make "very dissatisfied", "not very satisfied", "average", "relatively satisfied" and "very satisfied" judgments on each product feature. However, it is difficult for ordinary consumers to have enough discernment to effectively distinguish these five grades, and the focus of this paper is on the degree of influence of whether the product features meet the consumer's utility, so it is only necessary to collect information about whether consumers are satisfied with the product features, that is, the consumers' emotional tendency to the product features.

3.4. Modeling Consumer Utility

The purpose of consumer utility modeling is to analyze the effect of product characteristics on the overall utility of the consumer, the basic idea is that the utility of the product obtained by the consumer is a function of the characteristics of the product (product attributes), i.e., the product utility $=\mathrm{f}$

(product attributes), where $X_1, X_2, X_3, \ldots, X_n$, is the n characteristics of the product, X is the product consisting of $X_1, X_2, X_3, \ldots, X_n$, and U is the utility, and the materialization formula is shown:

$$U(X) = U(X_1) + U(X_2) + U(X_3) + \cdots + U(X_n) \tag{8}$$

In general, there are three preference estimation models in the joint analysis method: vector model, ideal point model and score function model. The specific mathematical expression formulas are shown in Eqs. (9) to (11):

$$U = \sum_{j=1}^{n} \beta_j X_j \tag{9}$$

$$d^2 = \sum_{j=1}^{n} \beta_j \left(X_j - X_{ideal} \right)^2 \tag{10}$$

$$U = \sum_{j=1}^{n} f_j \left(X_j \right) \tag{11}$$

In the vector model, consumer utility increases linearly with the performance of product features. While in the ideal point model, consumer utility will first increase with the increase or improvement of product features, but it will not keep increasing, and once the ideal point is exceeded, consumer utility will decrease. The score function model is more flexible, can use the above two different utility mechanisms for different product features, and also conforms to the KANO model demand classification theory in this paper, so this paper adopts the score function model to model consumer utility. The specific model is:

$$U = \alpha + \sum_{j=1}^{n} \left(\beta_j^{pos} X_j^{pos} + \beta_j^{neg} X_j^{neg} \right) \tag{12}$$

U represents the utility of the consumer in using the cell phone. In previous online review mining studies, star ratings of reviews are generally used to refer to consumer utility. A rating of $X_j^{pos} = 1$ indicates that consumers have positive sentiments towards product feature j of the cell phone. A rating of $X_j^{neg} = 1$ indicates that the consumer has a negative sentiment towards product feature j of the cell phone. If consumers have neither positive nor negative emotions towards product feature j, it means that the product feature j basically meets consumers' expectations, in which case X_j^{pos} and X_j^{neg} are both ground, and β_j^{neg} represents the influence weight of product feature j on consumers' utility in the presence of negative emotions. Based on the formula, parameters α and β can be estimated by regression.

3.5. Product Characterization Weight Estimation

According to the parameter estimation of the consumer utility model and the coding of the evaluation of emotions are 0-1 variables, so the parameters in front of the evaluation of emotions have an intuitive meaning, i.e., β_j^{pos} represents the consumer utility resulting from obtaining a positive evaluation of the product features j by the consumer, while β_j^{neg} represents the consumer utility resulting from obtaining a negative evaluation of the product features j by the consumer. According to the formula, it is possible to construct the extreme variance Range of the variables:

$$Range_j = \left| \beta_j^{pos} - \beta_j^{neg} \right| \tag{13}$$

Range in the formula is the difference between the utility of the two, indicating that the product feature j from negative evaluation to positive evaluation on the consumption of large, the more important the enterprise should be product feature j innovation management. In summary, the normalization of Range can be obtained for each product feature weight:

$$Weight_j = Range_j \Big/ \sum_{1}^{n} Range_j \tag{14}$$

3.6. Classification of user requirements based on the KANO model

According to the different ways in which the fulfillment of consumer needs acts on product utility, needs can be classified as basic needs, expectation needs, charm needs, undifferentiated needs and reverse needs.

Undifferentiated needs have a limited degree of influence on consumer utility regardless of whether they are satisfied or not, so the judgment rule for undifferentiated needs is: $\left|\beta_j^{pos}\right| < \gamma$, and $\left|\beta_j^{neg}\right| < \gamma$, where γ is the threshold that can be regulated.

Expectation demand and reverse demand are both linearly varying, and this paper creates a new variable Backlog to measure linearly varying demand:

$$Back\log_j = \beta_j^{pos} + \beta_j^{neg} \tag{15}$$

Backlog is the sum of consumer utility from positive and negative evaluations of a product feature, and in general, positive evaluations lead to positive consumer utility and negative evaluations lead to negative consumer utility.

Both basic demand and glamor demand vary nonlinearly. When $\left|Back\log_j\right| > \delta * Range_j$ three reference points $\left(-1, \beta_j^{neg}\right)$, (0, 0) and $\left(1, \beta_j^{pos}\right)$ form a curve that will be either externally convex or internally concave. When $Back\log_j > 0$ it is a glamor demand, otherwise it is a basic demand.

4. Consumer demand mining and optimization strategy analysis

4.1. Comparative analysis of LDA algorithms on different datasets

In order to evaluate the effectiveness of online review text representation based on LDA topic model, a series of experiments are conducted and analyzed on three datasets, namely, hotel category, food category and shoes category. Firstly, we calculate the LDA perplexity under different number of topics, and compare and develop the optimal number of topics in each domain data.The hyperparameters of the LDA topic model training input are $a = \frac{1}{k}$, b =0.02, where k is the number of topics. To ensure that the model training converges, set the number of iterations $iters$ = 3000.

In order to analyze the influence of the number of topics on the degree of confusion, the minimum number of topics is 10, and the interval of 5 is increasing, and the maximum number of topics is 100.The changes of the degree of confusion under different numbers of topics are shown in Fig. 3.

With the increase of the number of themes, the confusion degree of hotel, food and shoes shows a decreasing trend, which also verifies the conclusion that the confusion degree is inversely proportional to the number of themes. Taken as a whole, under the same theme, the hotel category has the smallest perplexity and the food category has the largest value, which indicates that the training effect of the hotel category data is better than that of the food category and the footwear category. The overall perplexity of the hotel category data is at [-6.5, -7.5]. Considering the size of the number of topics and the confusion degree will be small, the optimal number of topics can be selected, in which the optimal number of topics for the hotel category, the food category, and the footwear category are 90, 70, and 70, respectively.

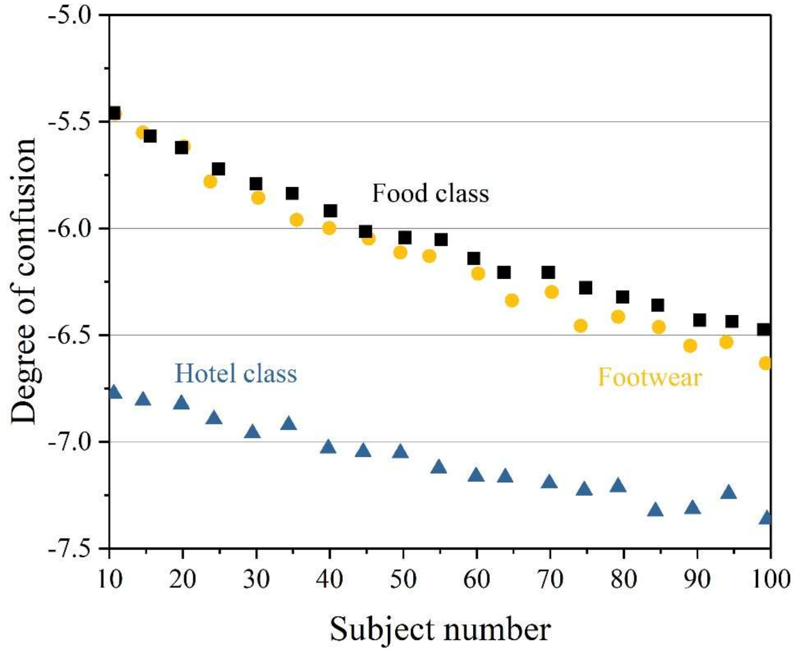

Fig. 3. The size of the confusion in different topic Numbers changes

The distribution of words under different themes is not the same, and the connotation of their themes is also different. The distribution of words under the theme in the field of footwear, limited to space, show $Topic\ 1$, $Topic\ 2$ two different themes, the right side of each theme lists the probability distribution of each word under the theme. The theme-word distribution of footwear data is shown in Table 1, and the two themes are centered around style and aesthetics, and size respectively, in which the distribution probability of "suitable" is 0.213 in $Topic\ 2$.

Table 1. Shoe data theme - word distribution

num	$w1$	$w2$	$w3$	$w4$	$w5$
$Topic\ 1$	Delicacy	Workmanship	Style	Appearance	Look good
	0.157	0.562	0.611	0.459	0.562
	$w6$	$w7$	$w8$	$w9$	$w10$
	Color	Size	Wear well	Aesthetics	Praise
	0.392	0.451	0.068	0.121	0.078
$Topic\ 2$	$w1$	$w2$	$w3$	$w4$	$w5$
	Suitable	Code number	Width	Size	Comfort
	0.213	0.021	0.321	0.021	0.251
	$w6$	$w7$	$w8$	$w9$	$w10$
	Husband	Shoes	Family	Like	Sole
	0.038	0.251	0.008	0.611	0.362

The theme-word distribution of the data in the hotel category is shown in Table 2, with most of the hotel's $Topic\ 1$ themes revolving around services, facilities, etc., and the $Topic\ 2$ themes being the hotel's location.

Table 2. Hotel class data theme - this distribution

num	$w1$	$w2$	$w3$	$w4$	$w5$
$Topic\ 1$	Service	Room	Quietness	Facilities	Window
	0.132	0.043	0.256	0.331	0.472

	w_6	w_7	w_8	w_9	w_{10}
	Toilet	Hotel	Bathing	Clean	Meal
	0.553	0.211	0.139	0.663	0.078
Topic 2	w_1	w_2	w_3	w_4	w_5
	Sense	Parking lot	Convenience	Traffic	Address
	0.0326	0.213	0.332	0.218	0.0421
	w_6	w_7	w_8	w_9	w_{10}
	Suburban	Incity	Pedestrian street	Taxi	Perimeter
	0.0063	0.0124	0.054	0.0758	0.095

The theme-word distribution of the food category data is shown in Table 3. The two themes of the food category describe logistics and flavor sensation, etc., respectively.

The synthesis of hotel, footwear and food category data themes shows that the differences between the individual themes trained by the LDA model are obvious, with strong differentiation, and thus the differences between the different document distributions obtained are also large. When the content between the documents is similar, the distribution of topics tends to be consistent, the similarity between the two document vectors calculated by the higher degree of similarity. The use of LDA model can be a good way to discover the potential semantic relationships within the document, making up for the shortcomings of the traditional vector space model that can only characterize shallow word relationships.

Table 3. Food class data theme - word distribution

num	w_1	w_2	w_3	w_4	w_5
Topic 1	Supermarket	Affordable	Cheap	Reassurance	Environment
	0.063	0.025	0.521	0.123	0.0213
	w_6	w_7	w_8	w_9	w_{10}
	Express delivery	Logistics service	Timely delivery	Clean and tidy	Shipping
	0.055	0.0324	0.266	0.0552	0.0219
Topic 2	w_1	w_2	w_3	w_4	w_5
	Like	Buy	Thing	Taste	Satisfaction
	0.021	0.0325	0.039	0.0213	0.0111
	w_6	w_7	w_8	w_9	w_{10}
	Taste	Yummy	Field production	Diversification	Quality assurance
	0.125	0.024	0.036	0.0875	0.062

4.2. Statistical analysis of user demand data

The three automobile information platforms of AutoNews, Car Knowledge and EasyCar are currently more authoritative automobile forums, and by comparing the downloads, ratings and content richness of these three websites, the AutoNews website, which is more representative in all aspects, was finally selected as the object of data analysis in this paper.

AutoHome.com is an automobile website that specializes in providing users with services such as choosing and buying cars. In terms of downloads and searches, AutoHome is ranked first among life search sites, and it is more suitable as a data source for samples than other platforms because of its higher ratings and the richer content it contains. Therefore, this paper crawls the review data about new energy vehicles on the website of Automotive Home to extract and summarize the users' needs and the factors affecting their satisfaction, in order to improve the design of the offline questionnaire related to the content.

This survey crawled the comments on the word-of-mouth page about new energy vehicles on the website of AutoHome from January 3, 2023 to December 15, 2023 by Octopus Collector, totaling 521,362 comments. Each review includes brand, series, model, purchase price, purchase time, purpose of purchasing the car, most and least satisfied, and evaluation of the overall and each factor, and then imported into Excel to form text information.

In this paper, the LDA model is used to identify the hidden user needs in the online review data, according to the specifics of the distribution of the review topics, the number of topics is taken as 4, and the analysis is carried out by using the LDA topic model, and the topic-word distribution obtained from the new energy vehicle user review data is shown in Table 4.

From the table, we know that theme 1 is user comments about sensory experience, theme 2 is user comments about new energy vehicle related function configuration, theme 3 is user comments about new energy vehicle economy, and theme 4 is user comments related to new energy vehicle service.

Combining the results of word frequency analysis and LDA theme analysis, the main influencing factors of new energy vehicle user satisfaction are extracted and summarized as follows:

Theme 1, the characteristic words include words such as space, appearance, interior, design, etc. According to the keyword analysis, it can be concluded that the main elements affecting user satisfaction include the sensory factors such as space, environment and appearance and interior of new energy vehicles. The distribution probabilities of space and appearance are 0.032 and 0.012 respectively.

Theme 2, feature words including driving (0.019), range (0.021), power (0.020) and other words, according to the keyword analysis can be concluded that the main content that affects user satisfaction includes new energy vehicles range, power and other kinetic configurations.

Theme 3, feature words including cost-effective, price, low, save money and other words, according to the keyword analysis can be concluded that the main content affecting user satisfaction includes the cost-effectiveness of new energy vehicles, fuel consumption and other economic factors. Among them, the distribution probability value of fuel is 0.021, and consumers are more concerned about the fuel consumption of the car.

Theme 4, the characteristic words include words such as service, experience, introduction, attitude, etc. According to the keyword analysis, it can be concluded that the main content affecting user satisfaction includes the service of new energy vehicles, experience and other service quality.

Table 4. User comment data theme - word distribution

Topic 1	P	Topic 2	P	Topic 3	P	Topic 4	P
Space	0.032	Driving	0.019	Performance ratio	0.008	Charging	0.011
Appearance	0.012	Endurance voyage	0.021	Price	0.009	Convenience	0.01
Design	0.011	Power	0.020	Low	0.009	Battery	0.008
Interior	0.009	Function	0.033	Save money	0.005	Service	0.006
Model	0.021	Maneuvering	0.002	Fuel oil	0.021	Not bad	0.005
Face value	0.006	Use	0.009	Satisfy	0.012	Raise	0.003
Body	0.001	Accelerate	0.006	Economy	0.013	Use	0.007
Seat	0.003	Power	0.002	Oil consumption	0.015	Everyday	0.008

4.3. Quantitative Carnot Transformation of Consumer Needs

The satisfaction rating criteria are expressed here and the satisfaction rating criteria are shown in Table 5. Satisfaction with product/service attributes is categorized into five ratings: liked, deserved, indifferent, tolerable, and very disliked.

Table 5. Satisfaction rating criteria

		Like	Rightly	Indifferent	Tolerable	Dislike
Product/service properties	Satisfy the satisfaction of the property	1	0.725	0.45	0.325	0
	Dissatisfaction with the satisfaction of the attribute	1	0.725	0.45	0.325	0
	Dissatisfaction with the attribute	0	0.325	0.45	0.725	1

So, based on the above rating criteria, using the dissatisfaction when an attribute is not satisfied as the X-axis coordinate for attribute i, and the satisfaction when an attribute is satisfied as the Y-coordinate for that attribute, the Kano Evaluation Results Classification table can be changed to the following table.The Kano Evaluation Results Classification is shown in Table 6.

Table 6. Kano evaluation results classification

	Like	Rightly	Indifferent	Tolerable	Dislike
Like	Q (0,1)	A (0.325,1)	A (0.450,1)	A (0.725,1)	O (1,1)
Rightly	R (0,0.725)	I (0.325,0.725)	I (0.450,0.725)	I (0.725,0.725)	M (1,0.725)
Indifferent	R (0,0.450)	I (0.325,0.450)	I (0.450,0.450)	I (0.725,0.450)	M (1,0.450)
Tolerable	R (0,0.325)	I (0.325,0.325)	I (0.450,0.325)	I (0.725,0.325)	M (1,0.325)
Dislike	R (0,0)	R (0.285,0)	R (0.450,0)	R (0.725,0)	Q (1,0)

According to the formula for attribute point coordinates, the representation of attribute i corresponding to multiple critical requirements on the satisfaction coordinates can be obtained. According to the distribution of attribute points on the coordinates, its Kano classification can be determined. In the actual calculation process, a table can be built to find out the sum of the sentiment scores of each of the multiple attributes in each word-of-mouth document in order to analyze its degree of satisfaction for a particular user. Combining the normalized results of sentiment calculation of a single word-of-mouth document, i.e., the satisfaction evaluation of a single user for a product/service, the average dissatisfaction of the product/service for users who think the attribute is not satisfied (average dissatisfaction when the attribute is not satisfied) and the average satisfaction of the product/service for users who think the attribute is satisfied (average satisfaction when the attribute is satisfied) is calculated, and the results of the satisfaction calculation are shown in Table 7, which shows the following results for the automobile The satisfaction value of air vents is -15.645.

Table 7. Satisfaction calculation results

Word of mouth	Word-of-mouth satisfaction	Assembly gap	Air outlet	Damping	Fetal noise	Air conditioning	Air noise
1	0.521	-23.362	-15.645	-9.632	-8.217	-1.352	-0.315
2	0.336	0	0	0	0	0	0
3	0.293	0	0	0	0	0	0
4	0.321	-20.31	0	-14.399	0	-3.221	0
5	0.475	0	-16.321	0	0	0	-4.663
6	0.201	0	0	0	0	0	0
7	0.364	0	0	0	0	0	0

Through the above table, the attribute point coordinates can be calculated and its Kano category can be determined based on its position on the satisfaction coordinates. The results of attribute point coordinates and Kano category calculation are shown in Table 8. The average satisfaction when buttons, brightness is a non-differentiating attribute, and satisfaction are 0.653 and 0.629, respectively. Cigarette lighter, keys, and sofa are all required attributes, with average dissatisfaction levels of 0.562, 0.711, and 837, respectively, when they are not met. Staccato is a charm attribute with a mean satisfaction of 0.806 when satisfied.

Table 8. Properties point coordinates and carnot category calculation results

Feature	Satisfied average satisfaction	The average dissatisfaction is the average	Carno classification
Button	0.653	0.569	Indifference attribute
Cigarette lighter	0.721	0.562	Essential attribute
Frustration	0.806	0.725	Charm attribute
Key	0.656	0.711	Essential attribute
Brightness	0.629	0.566	Indifference attribute
Sofa	0.536	0.837	Essential attribute

Thus, this study has obtained the Kano categories of the attributes corresponding to the key requirements, and prioritized and evaluated the requirements in order to further prioritize the requirements to be developed and responded to. The parameters affecting the efficiency of satisfaction enhancement are calculated by fitting the Kano function exclusive to the attribute through its Kano category and substituting the known points. Attributes belonging to Kano categories that are reverse attributes and undifferentiated attributes are removed. The satisfaction enhancement efficiency a and the parameter b of the obtained attribute are calculated, and the Kano fitting function of the attribute can be obtained, and the function image can be plotted for further analysis if needed. The values of parameters a and b corresponding to the calculated attributes are shown in Table 9. Taking low beam and key as an example, low beam and key are categorized as desired attribute and required attribute, respectively, with a and b values of 0.593 and 0.218, 0.613 and 0.191, respectively.

Table 9. Kano parameter calculation results

Feature	Carno classification	a	b

Near light	Expected attribute	0.593	0.218
Key	Essential attribute	0.613	0.191
Pedal	Expected attribute	0.721	0.045
Cigarette lighter	Essential attribute	0.612	0.833
Aerial	Essential attribute	0.433	0.475
Window	Essential attribute	0.509	0.869
Acoustics	Essential attribute	0.662	0.825
Sofa	Essential attribute	0.486	0.722
Frustration	Charm attribute	0.239	0.545
Oil box cover	Essential attribute	-0.425	-0.297
One key start	Essential attribute	-0.361	-0.126
Extinguishing	Essential attribute	-0.415	-0.163
Connecting rod	Essential attribute	-0.362	-0.158
Stern door	Essential attribute	-0.405	-0.173
Lock car	Essential attribute	-0.365	-0.081
Spare tire	Essential attribute	-0.123	-0.053
Oil tank	Essential attribute	-0.112	-0.046

4.4. Requirements sequencing and product optimization strategies

The purpose of this section is to obtain the higher satisfaction enhancement efficiency attribute and analyze the priority of the attribute corresponding to the requirement. The initial average satisfaction of the feature attributes corresponding to the requirement is analyzed in combination with the initial average satisfaction of the feature attributes corresponding to the requirement, and the attributes corresponding to the requirement with low initial satisfaction and high satisfaction enhancement efficiency are ranked in the top of the priority list. According to the designed formula for calculating the initial average satisfaction and weight of attributes, the initial average satisfaction of attribute $\overline{M_{ik}}$ and the final demand weight I_i can be calculated, and the calculation results are shown in Table 10 in reverse order of demand weight. The top six demand weights of consumers for automobiles are, in order, pedals (3.332), cigarette lighter (3.015), antenna (2.539), windows (2.121), low beams (1.866), and keys (1.542). It can be noticed that the top prioritized needs are more related to the need to add antenna, cigarette lighter, pedals, and enhance accessory features. Then there is the need to reduce rattles, increase seat comfort, improve power, enhance shift flexibility and other aspects of the car's handling experience, as well as needs related to the quality of the car's engine, such as dimming the lights, key malfunctions, and tuning capabilities. Therefore, automakers should prioritize accessory features, handling experience, and vehicle quality as their main focus when iterating their products. In the selection of specific iterative solutions or demand response prioritization, consider the demand corresponding to the attributes with the highest priority calculation results from this study. For example, solving the problem of pedal vibration, adding cigarette lighter and antenna, and solving the problem of window rattling.

Table 10. Demand weight calculation results

Feature	Attribute initial average satisfaction $\overline{M_{ik}}$	I_i
Near light	0.325	1.866
Key	0.471	1.542
Pedal	0.321	3.332
Cigarette lighter	0.258	3.015
Aerial	0.143	2.539
Window	0.239	2.121
Sofa	0.566	0.758
Frustration	0.374	0.596
One key start	0.244	-1.628
Stern door	0.261	-1.253
Oil box cover	0.239	-1.525
Noise	0.207	-1.429
Cosmetic mirror	0.235	-1.303
Electric bottle	0.178	-1.427

5. Conclusion

In this paper, we screen consumer demand to obtain data and use LDA topic model to capture product attributes and consumer demand information from online reviews. The KANO model is used to classify the consumer demand and calculate the sorting demand weights to obtain the product optimization strategy.

Combine the consumer demand multi-source data and take advantage of online review data to preprocess online reviews.The LDA model shows strong differentiation on all three types of datasets: hotel, shoes, and food, and yields the topic distribution of different documents, which is able to effectively capture the hidden semantic information inside the documents.

The user demand classification scheme based on KANO model is established to convert consumer demand in the form of satisfaction, and the consumer demand weights are calculated. Take the air conditioner of the car as an example, the satisfaction value of the air conditioner is -1.352. Consumer demand prioritization calculation and evaluation get the consumer demand desire for the car in the order of footrest > cigarette lighter > antenna > window > low beam > key >. Thus, automobile manufacturers should prioritize automobile pedals, cigarette lighter, antenna, and windows when upgrading the design of automobiles to enhance consumers' desired attributes and optimize the implementation of the necessary attributes of automobiles.

References

[1] Wang, Y., & Li, L. (2024). Digital economy, industrial structure upgrading, and residents' consumption: Empirical evidence from prefecture-level cities in China. International Review of Economics & Finance, 92, 1045-1058.

[2] Doborjeh, Z., Hemmington, N., Doborjeh, M., & Kasabov, N. (2022). Artificial intelligence: a systematic review of methods and applications in hospitality and tourism. International Journal of Contemporary Hospitality Management, 34(3), 1154-1176.

[3] Chen, H., Liu, Y., & Wang, Z. (2024). Can Industrial Digitalization Boost a Consumption-Driven Economy? An Empirical Study Based on Provincial Data in China. Journal of Theoretical and Applied Electronic Commerce Research, 19(3), 2377-2399.

[4] Fazayeli, S., Eydi, A., & Kamalabadi, I. N. (2018). Location-routing problem in multimodal transportation

network with time windows and fuzzy demands: Presenting a two-part genetic algorithm. Computers & Industrial Engineering, 119, 233-246.

[5] Yong, L., & Huipeng, Y. (2022). Digital Consumption and Manufacturing Industry's Digital Transition. China Economist, 17(6), 54-76.

[6] Stefanov, K., Huang, B., Li, Z., & Soleymani, M. (2020, October). Opensense: A platform for multimodal data acquisition and behavior perception. In Proceedings of the 2020 International Conference on Multimodal Interaction (pp. 660-664).

[7] Onete, C. B., Teodorescu, I., & Vasile, V. (2016). Analysis Componentsof the Digital Consumer Behavior in Romania. Amfiteatru Economic Journal, 18(43), 654-662.

[8] Liu, X. Y. (2021). Agricultural products intelligent marketing technology innovation in big data era. Procedia Computer Science, 183, 648-654.

[9] Shi, Y., Gao, Y., Luo, Y., & Hu, J. (2022). Fusions of industrialisation and digitalisation (FID) in the digital economy: Industrial system digitalisation, digital technology industrialisation, and beyond. Journal of Digital Economy, 1(1), 73-88.

[10] Jiang, W., Zhang, Y., Han, H., Huang, Z., Li, Q., & Mu, J. (2024). Mobile traffic prediction in consumer applications: a multimodal deep learning approach. IEEE Transactions on Consumer Electronics.

[11] Hilbert, M. R. (2001). From industrial economics to digital economics: an introduction to the transition. ECLAC.

[12] Fang, X., Cao, C., Chen, Z., Chen, W., Ni, L., Ji, Z., & Gan, J. (2020). Using mixed methods to design service quality evaluation indicator system of railway container multimodal transport. Science Progress, 103(1), 0036850419890491.

[13] Daymond, J., Meisiek, S., & Knight, E. (2024). Into the Customers' Shoes: Multimodal practices for customer-centric strategizing. Organization Studies, 45(11), 1579-1609.

[14] Barefoot, K., Curtis, D., Jolliff, W., Nicholson, J. R., & Omohundro, R. (2018). Defining and measuring the digital economy. US Department of Commerce Bureau of Economic Analysis, Washington, DC, 15, 210.

[15] Brida, J. G., Martín, J. C., Román, C., & Scuderi, R. (2017). Air and HST multimodal products. A segmentation analysis for policy makers. Networks and Spatial Economics, 17, 911-934.

[16] Su, J., Su, K., & Wang, S. (2021). Does the digital economy promote industrial structural upgrading?—A test of mediating effects based on heterogeneous technological innovation. Sustainability, 13(18), 10105.

[17] Cai, W., Song, Y., & Wei, Z. (2021). Multimodal Data Guided Spatial Feature Fusion and Grouping Strategy for E-Commerce Commodity Demand Forecasting. Mobile Information Systems, 2021(1), 5568208.

[18] Wang, Y. (2021). Survey on deep multi-modal data analytics: Collaboration, rivalry, and fusion. ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM), 17(1s), 1-25.

[19] Zhao, Q., Wang, Q., Wu, P., Fan, J., Shi[1], L., Li, H., ... & Gao, Y. (2024). A Method for Analyzing Consumer Behavior and Evaluating Marketing Effectiveness of Online Websites Based on Multi-Modal Data Synchronization. Usability and User Experience, 107.

[20] Grewal, R., Gupta, S., & Hamilton, R. (2021). Marketing insights from multimedia data: text, image, audio, and video. Journal of Marketing Research, 58(6), 1025-1033.

[21] Napitupulu, L. H., Bako, E. N., Ars, N. R., & Zein, T. (2018). A multimodal analysis of advertisement of online marketplace Shopee. KnE Social Sciences, 452-460.

[22] Li, L., Chi, T., Hao, T., & Yu, T. (2018). Customer demand analysis of the electronic commerce supply chain using Big Data. Annals of Operations Research, 268, 113-128.

[23] Giannakos, M. N., Sharma, K., Pappas, I. O., Kostakos, V., & Velloso, E. (2019). Multimodal data as a means to understand the learning experience. International Journal of Information Management, 48, 108-119.

[24] Arrese, A., Medina, M., & Sánchez-Tabernero, A. (2019). Consumer demand and audience behavior. In A research agenda for media economics (pp. 59-76). Edward Elgar Publishing.

[25] Zhou, R. (2022). Sustainable economic development, digital payment, and consumer demand: Evidence from China. International Journal of Environmental Research and Public Health, 19(14), 8819.

[26] Shoumy, N. J., Ang, L. M., Seng, K. P., Rahaman, D. M., & Zia, T. (2020). Multimodal big data affective analytics: A comprehensive survey using text, audio, visual and physiological signals. Journal of Network and Computer Applications, 149, 102447.

[27] Byrne, D., & Corrado, C. (2017, March). Accounting for innovation in consumer digital services. In 5th IMF Statistical Forum.

[28] Yuldasheva, O., Rodionov, D., Konnikova, O., Konnikov, E., Kryzhko, D., & Pirogov, D. (2023). Quantifying Consumer Preferences For Goods And Services Across Categories In The Digital Information Era. International Journal of eBusiness and eGovernment Studies, 15(1), 217-233.

[29] Alsokkar, A., Law, E. L., Almajali, D., & Alshinwan, M. (2023). The effect of multimodality on customers' decision-making and experiencing: A comparative study. International Journal of Data and Network Science, 7(1), 1-14.

[30] Cochoy, F., Licoppe, C., McIntyre, M. P., & Sörum, N. (2020). Digitalizing consumer society: equipment and devices of digital consumption. Journal of Cultural Economy, 13(1), 1-11.

[31] Basu, S., & Bundick, B. (2017). Uncertainty shocks in a model of effective demand. Econometrica, 85(3), 937-958.

[32] Jamil, R. (2020). Hydroelectricity consumption forecast for Pakistan using ARIMA modeling and supply-demand analysis for the year 2030. Renewable Energy, 154, 1-10.

[33] Ho, M., Britz, W., Delzeit, R., Leblanc, F., Roson, R., Schuenemann, F., & Weitzel, M. (2020). Modelling consumption and constructing long-term baselines in final demand. Journal of Global Economic Analysis, 5(1), 63-108.

[34] Pinjari, A. R., & Bhat, C. (2021). Computationally efficient forecasting procedures for Kuhn-Tucker consumer demand model systems: Application to residential energy consumption analysis. Journal of choice modelling, 39, 100283.

[35] Zhang, X., & Guo, C. (2024). Research on Multimodal Prediction of E-Commerce Customer Satisfaction Driven by Big Data. Applied Sciences, 14(18), 8181.

[36] Liu, X., Liu, Y., Qian, Y., Jiang, Y., & Ling, H. (2023). Learning consumer preferences through textual and visual data: a multi-modal approach. Electronic Commerce Research, 1-30.

[37] Chen, Y., Iyengar, R., & Iyengar, G. (2017). Modeling multimodal continuous heterogeneity in conjoint analysis—a sparse learning approach. Marketing Science, 36(1), 140-156.

[38] Din, I. U., Almogren, A., Rodrigues, J. J., & Altameem, A. (2024). Advancing Secure and Privacy-Preserved Decision-Making in IoT-Enabled Consumer Electronics via Multimodal Data Fusion. IEEE Transactions on Consumer Electronics.

[39] Chen, G., Huang, L., Xiao, S., Zhang, C., & Zhao, H. (2024). Attending to customer attention: A novel deep learning method for leveraging multimodal online reviews to enhance sales prediction. Information Systems Research, 35(2), 829-849.

[40] Li Hongming,Xiu Taiyu,Sun Yin,Zhang Xuan,Gao Yunting,Wu Jieting & Zhang Yurong Abby. (2025). The

Analysis of Emotion-Aware Personalized Recommendations via Multimodal Data Fusion in the Field of Art. Journal of Organizational and End User Computing (JOEUC)(1),1-29.

[41] Li Meiwen,Xia Liye,Wu Qingtao,Wang Lin,Zhu Junlong & Zhang Mingchuan. (2025). MD-LDA: a supervised LDA topic model for identifying mechanism of disease in TCM. Data Technologies and Applications(1),1-18.

[42] Liang Gao,Fang Cui & Chengbo Zhang. (2024). Library Similar Literature Screening System Research Based on LDA Topic Model. Journal of Information & Knowledge Management(05).