

Article

Not peer-reviewed version

An Analysis Framework to Reveal Automobile Users' Tastes from Online User-Generated Content

[Hanyang Luo](#) , [Wugang Song](#) , [Wanhua Zhou](#) ^{*} , [Xudong Lin](#) , [Sumin Yu](#)

Posted Date: 16 August 2023

doi: 10.20944/preprints202308.1102.v1

Keywords: Chinese automobile market; user tastes; online user-generated content; sentiment analysis



Preprints.org is a free multidiscipline platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Article

An Analysis Framework to Reveal Automobile Users' Tastes from Online User-Generated Content

Hanyang Luo ¹, Wugang Song ², Wanhua Zhou ^{2,*}, Xudong Lin ¹ and Sumin Yu ¹

¹ Institute of Big Data Intelligent Management and Decision, College of Management, Shenzhen University, Shenzhen 518060, China; hanyang@szu.edu.cn (H.L.); xudonlin@szu.edu.cn (X.L.); yusumin@szu.edu.cn (S.Y.)

² College of Management, Shenzhen University, Shenzhen 518060, China; 2100132004@email.szu.edu.cn

* Correspondence: 2060132016@email.szu.edu.cn (W.Z.)

Abstract: With the view toward enlivening the Chinese automobile market, this work seeks to develop a framework to reveal the tastes of Chinese car users from online user-generated content (UGC) and identify future product improvement orientation. We construct an importance-satisfaction gap analysis (ISGA) model based on sentiment analysis technique. Specifically, a novel unsupervised word-boundary-identified algorithm for the Chinese language is used to extract domain professional feature words and a set of sentiment scoring rules is constructed. By matching feature-sentiment word pairs, we calculate car users' satisfaction with different attributes based on the rules, and weigh the importance of attributes using TF-IDF method, which lay the statistic foundation for our ISGA model. We also take time into account and further analyze the ISGA model, aiming to provide new insight into the changing tastes of Chinese car users so as to suggest correct design positioning and resource allocation. The final evaluation results show that the method proposed in this paper, which integrates Chinese word-boundary-identified algorithm and sentiment lexicon matching rules, is well supported.

Keywords: Chinese automobile market; user tastes; online user-generated content; sentiment analysis

1. Introduction

Since 1907, when Japan produced the first petrol car, the automobile industry has developed into a decisive force in modern industrial civilization. Characterized by high degree of industrial correlation, automobile plays a strong driving role in upgrading industrial structure and employment of related industries. Therefore, the automobile industry has served as an important pillar for national prosperity in most countries, and automobile sales are closely related to the lives of people. Up to 2019, the number of cars ownership per thousand people is 500-800 in developed countries, while this figure in China is only 173 [1]. In 2021, the number fell by 11. Considering the population size, regional structure and national differences in resources, the automobile sector in China still has a large room for growth as people's income continues to rise and urbanization gradually advances [2]. According to China Association of Automobile Manufacturers, the automobile market in China saw a poor performance recently, with Chinese vehicle industrial supply chains clogged and sales down in the doldrums [3]. It should not simply be attributed to the impact of COVID-19. As early as 2019, it was reported that Chinese people are becoming less interested in buying cars, and the proportion of young car-users under 30 years old has dropped sharply, which poses challenges to auto enterprises in the future [4]. Methods to unearth the tastes of Chinese car users are of great importance for major auto brands, which is also the major concern of this paper.

Traditional methods for automobile companies to judge consumer tastes like studying the annual sales of various types of new cars in major areas that typically only reflects users' basic needs and their pre-purchase behavior, are not adequate to access their experience needs after driving the

cars, let alone adjusting product improvement orientation by identifying the problems found by experienced consumers. Some consulting companies investigate consumers' preferences for cars through voice-of-customers (VOC) interviews, however, it is difficult to make timely responses to the market due to the large labor cost and time cost needed in a VOC study. Most recently, researchers pay attention to User-Generated Content (UGC), which is rich in customer opinion information [5]. UGC-based customer needs are proved to be comparable to VOC-based customer needs [6]. In the study of Artem et al [6], analysts identify more customer needs from UGC than from interviews of a VOC study at a given cost of time and money. Prior studies also provide evidence that it is of great commercial value to mine customers' shopping behavior patterns from user-generated information e.g., [7–9]. These huge, disorganized and fast-updating UGC data contain real and diverse information of user needs. Therefore, it's quite practical to leverage UGC to identify the needs of Chinese car users. Most UGC research claps eyes on the online reviews of B2C platforms where most of the products categories are fast-moving consumer goods (FMCG) with relatively short life cycles [10]. Compared with FMCG, non-fast-moving consumer goods such as automobiles we discuss in this paper lead to different buying patterns. For example, Luo et al [11] consider that non-FMCG is characterized by a low degree of homogeneity and thus consumers have stronger loyalty to the brand. In addition, it's hard to find a B2C platform selling cars online because consumers always need to take their preferential products for a test drive and make deliberate decisions. Fortunately, the online verticals forum autohome.com provides convenient data access to Chinese consumer reviews. Different from comprehensive platforms (e.g., Amazon.com), vertical platforms focus on specific fields or specific needs and provide all the in-depth information and related services. As a new bright spot of the Internet, vertical websites are attracting more and more people's attention. autohome.com is an O2O model (online to offline, i.e. online booking and offline consumption) based e-commerce vertical platform on which Chinese users can share their opinions on the cars they are driving. It not only provides buyers with reliable information of different cars and preferential prices for reference, but also gives researchers who intend to study the Chinese auto market convenience to investigate consumers' cognition, choices and consumer demands in the process of car maintenance and use. There are few fake reviews on such platforms because customers can tell their true experience anonymously online without being worried about retaliation by offline merchants, and such review data is more authentic and reliable. Another benefit is that reviews on these forums do not involve logistics, sellers and other factors unrelated to product performance, filtering out irrelevant information for us. Therefore, mining customer tastes through UGC on such forums can be an important means to find new business opportunities and attract customers in China.

Our main research task is to explore the taste of Chinese automobile consumers and put forward the direction of future product improvement. Therefore, we propose a framework containing concrete methods to solve some problems encountered in Chinese text processing. Our approach is as follows: firstly, feature words are extracted from online reviews on the forum, and attributes are obtained by classification of feature words. Then, the score of attributes of each review is calculated by using the sentiment scoring rules we set, and then the importance-satisfaction gap analysis (ISGA) model is constructed for analysis. Although many related studies concerning customer needs apply similar methods and processes to other consumer goods and achieve good results [12–14], little attention has been paid to the automobile industry, particularly the Chinese market with huge consumption potential. First, user-generated content is not easy to access, only some surface data such as sales are available online. Second, the automobile is a very professional industry, which involves a lot of professional words, and data processing is more difficult. Moreover, since our research subjects are Chinese people, the data are all in the Chinese language, which differs from English in that no character distinguishes the boundary of words (In English, space is the character to identify the boundary of a word). Research using reviews in other languages for data analysis are not adequate to make a conclusion for Chinese consumers. In our paper, we adopt boundary-average entropy (BAE) algorithm to identify feature words of automobiles. BAE algorithm is a newly proposed unsupervised algorithm by Liu [15] in 2016 for Chinese words segmentation and is rarely known for the moment but it works well in his experiment. The extraction of feature words using the

algorithm and the rules for calculating the score of attributes of each review are discussed in Section 3. By applying the method, we verify the feasibility of the BAE algorithm in building a professional feature word dictionary for language in which word boundary is ambiguous. Considering time is a key context variable that can change the nature of a relationship, retrospective information is essential for a richer understanding. To investigate the evolvement of users' tastes in China, we take time into account to analyze the ISGA model.

The structure of the remaining part of this paper is as follows: Section 2 reviews two research streams that inform our work. In Section 3, we discuss the detail of our proposed method of identifying customer needs with Chinese online reviews in professional categories. Section 4 presents the application of our proposed method to the reviews generated by Chinese automobile users. In Section 5, we evaluate our framework and make a conclusion in the last section.

2. Literature Review

The first stream of literature related to our study is customer tastes identification from online UGC. Since we use sentiment scoring as the basis of users' satisfaction calculation, we also review sentiment analysis of products attributes with a focus on feature words extraction methods.

2.1. Customer tastes identification in the automobile industry

After more than 100 years of continuous improvement and innovation, the automobile embodies the wisdom and ingenuity of human beings, and thanks to the support of oil, steel, electronic technology and finance and other industries, it has achieved today's transportation tools with various types and different specifications, widely used in various fields of social and economic life. The development of automobile industry globalization is a systematic science, which has risen to the height of national strategy and strategy between countries. More and more attention is being paid to the industry. According to the retrieval data from Web of Science, as of October 26, 2021, 38,399 articles related to automobiles have been published on the database and the number is increasing year by year. These research largely cover the areas of vehicle designs and mechanical engineering concerning environmental problems [e.g., 16-18], driving safety [e.g., 19] and transportation science [e.g., 20], while less focus on the basic needs of car attributes of car users, which are normally the major concern of enterprises. Most of the papers related to consumer preferences are researches on pricing and decision models [e.g., 34]. However, these studies do not mine deeply into customers' opinions on various automobile attributes, and most of the related cases focus on brand preference, technology preference and vehicle type preference. This paper hopes to have a more detailed understanding of customers' tastes in specific vehicle attributes.

In the customer's own words, customer needs are descriptions of the benefits that a product or service can achieve [21]. Many research strive to propose methods to help companies quickly and accurately find out what do customers care about and complain about [6,22–25], which share a similar purpose with this study. The most traditional way to uncover customer needs is the voice-of-the-customer study, which typically starts with a limited number of qualitative experiential interviews of which the records will be transformed into scripts. Analysts will then review the scripts, manually identify what do people concern about their products, move out redundancy and develop abstract customer needs statements, usually in the form of a hierarchical structure [26,27] (i.e., needs are organized in primary groups, secondary groups and tertiary groups, etc.). Such manual coding methods require extensive labor and thus the size of the sample analyzed is usually small and cannot be representative. While the development of online user-generated content has brought about great interest of researchers in using big data and text mining techniques to mine people's opinions in different categories. Since one of our study aspects is to propose a method to identify Chinese car users' tastes, we have developed Table A1 in the appendix, which summarizes the research results of studies close to our research purpose of customer tastes mining based on Web of Science. Among the fifty-eight research, a big part of them use opinion mining techniques on the hotel industry and FMCGs, only five are related to automobiles. It is suggested that the research methods of customer demand or customer behavior for one product may not be applicable to other products [6,28]. In

terms of the five automobile research, Kühl et al. evaluate eight previously-defined customer needs on electric cars by applying a supervised machine learning method on German tweets [37]. Lee et al. choose Toyota Yaris as the case to study and design a content analyzer based on co-occurrence analysis to find out the most important elements that users in MForum care about [38]. Asghar et al. analyze the sentiment of Twitter users towards Honda, Toyota, BMW, Audi and Mercedes using a Naive Bayes classification method [51]. Sun et al. propose a method for dynamically analyzing changes in customers' sentiment toward the attributes of Trumpchi GS4 and GS8 [80]. In the study of Fang et al., Chinese feature words of cars are obtained based on one of the rules that the word is a noun in the lexicon of a Chinese word segmentation tool (i.e., jieba) [91]. On these research basis, we introduce a framework that identifies the taste of Chinese automobile users.

2.2. Sentiment analysis of products attribute

With the rise of the Internet, people's opinions and ideas began to be analyzed using natural language processing technology. Researchers value people's attitudes and emotions on a topic, such as the degree of public support or opposition to the introduction of policies in order to grasp the follow-up development and take corresponding control measures. In addition to politics, sentiment analysis is also used for commercial purposes. It can detect whether the trend of an aspect or attribute of a product or service is positive, negative or neutral. For example, this paper analyzes the satisfaction degree of car users in various attributes. Sentiment classification and the calculation of sentiment scores based on natural language processing are the two most common outputs in the field, which usually converts people's comments about a subject on the Internet into structured data. Sentiment analysis approaches are divided into three categories: 1) machine learning methods; 2) lexicon-based methods; 3) hybrid methods. Researches based on machine learning (ML) train various well-known classifiers to determine emotional direction. For example, Rai et al. select the best parameterized Naive Bayes Classifier to evaluate the sentiment polarity of live tweets and it performs better compared to Random Forest, Support Vector Machine in terms of accuracy, precision and time consumption [29]. The lexicon-based approach includes a set of sentiment dictionaries used to express positive or negative emotions. For example, words like "good" " nice" and "bad" "sad" are used to express positive and negative emotions, respectively. Dealing with the lexicons allows us to analyze the polarity of the sentiment of sentences or documents. Sometimes synonyms and antonyms are added to improve accuracy [30]. The novel hybrid methods of machine learning and lexicon dictionaries are becoming popular recently. In some of these research [e.g., 31], opinion words are firstly extracted from lexicon resources with the help of sentiment words identifiers like SentiWordNet and AFFIN-111, and then they are tagged by part-of-speech (POS). A machine learning classifier is finally used to identify the sentiments. Some research first calculates the sentiment score of reviews, and then determine the antecedents of dissatisfaction through topic model [e.g., 32,33] However, it has a defect: when an attribute appears frequently in a review, it will be regarded as a topic even if the customer thinks it is positive. For example, a car review with a very low sentiment score wrote: "The only thing I am satisfied with about this car is the seat. The seat design is reasonable, in line with human mechanics, and comfortable for long-distance driving. But the air conditioning noise is too loud, which makes my family restless." The negative attribute of this review is air conditioning, but due to the high occurrence of the seat, the seat will be regarded as the topic. Therefore, not all topics obtained from low score comments are negative topics, and vice versa. Our method first forms feature-sentiment word pairs according to the Chinese turning word and sentence separator to ensure that the sentiment word corresponds to the feature word. Each attribute in each comment is assigned a sentiment score instead of extracting topics followed by calculating the score of a review.

Most of the current work generate lexicons with a set of dictionaries or word identifiers trained by the previous corpus. However, sentiment analysis of a corpus of the emerging industry requires building lexicons by researchers for the lack of dictionary groundwork. High technology industry like automobiles grows rapidly, and their professional terms are constantly updated. If feature words are extracted based on a previously-defined lexicon, some new product feature words will be

ignored, and it is difficult to comprehensively grasp an overview of the advantages and disadvantages of products. The most commonly used Chinese word segmentation system jieba is based on string matching (i.e., the method based on dictionary) and the hidden Markov model (HMM). Although to a certain extent, HMM solves the problem of the unlisted words, but its string matching priority rules will affect the segmentation of specialized domain words. The most common error is to cut long string proper nouns into two or more entries in dictionaries. For example, "大众车(Volkswagen car)" will be divided into two words "大众(general)" and "车(car)", but the original meaning of "大众车" refers to the cars of Germany Volkswagen brand.

Therefore, this paper proposes to use an unsupervised word-boundary-identified algorithm, namely the boundary average entropy (BAE) algorithm, to solve the problem of unlisted word identification To address this issue, this study applies the algorithm to help automatically extract feature words and sentiment words of the professional industry. Based on the lexicon built by these words, we propose a set of matching rules for calculating the performance scores of each product attribute.

3. A Method to Identify Customer Tastes and Product Improvement Orientation

We propose a method that employs four major techniques to identify customer needs from Chinese UGC (specifically, online reviews), namely, BAE algorithm in the step of lexicon words extraction, sentiment scoring rules in calculating the users' satisfaction score of attributes, TF-IDF in calculating attribute importance, and ISGA model with time considered in analyzing customer taste and future product improvement orientation. Figure 1 shows the flow diagram of the procedure.

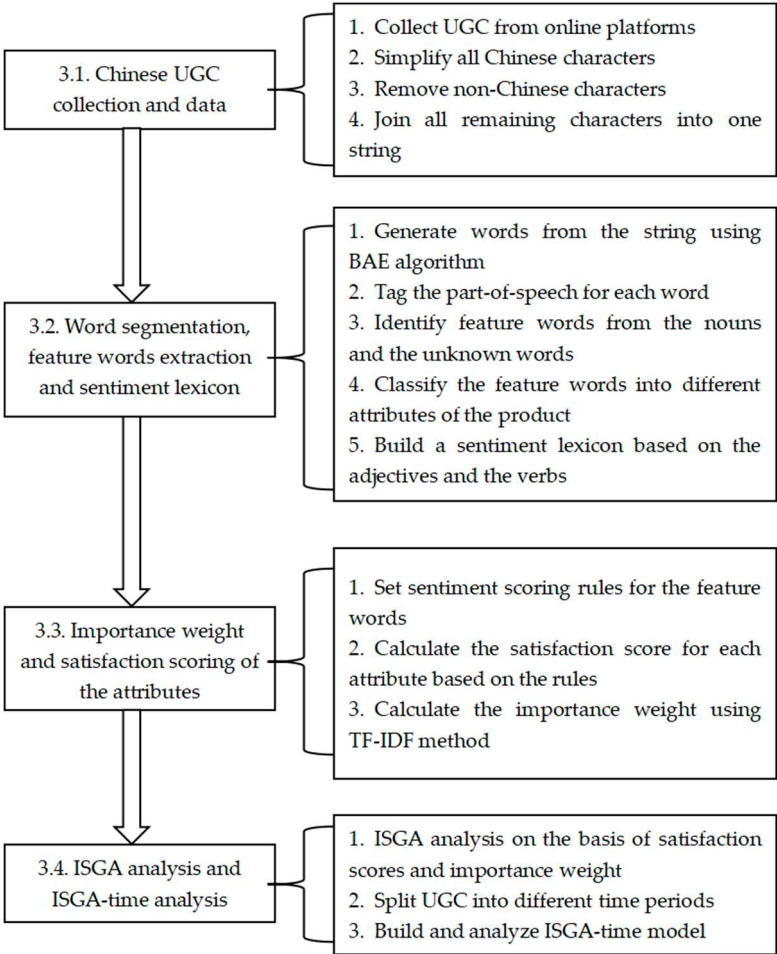


Figure 1. System architecture of the method.

3.1. Chinese UGC data collection and data cleansing

As the research object of this paper is the Chinese consumer group, the data collected are all Chinese user-generated text. The longer the total number of characters, the better the result of subsequent lexicon building. The data cleansing procedure in this paper is relatively simple and slightly different from other studies in that we do not remove stop words which the following training requires. All non-Chinese characters are removed, including punctuation marks and transposes. Next, all sentences are concatenated into a long string containing only Chinese characters, as a result of this step.

3.2. Word segmentation and feature words extraction

We obtain a long string consisting of many Chinese characters in step 3. 1. To identify which two or more consecutive characters can form an independent word in the language environment, we apply the BAE algorithm that is designed based on the principle of comparing the information entropy of concatenating strings before and after to determine word boundary. "Information entropy" stands for the average amount of information contained in the received information without redundancy, which can represent the uncertainty of an event. Assume that for discrete variable X , its value range is $\{x_1, x_2, \dots, x_n\}$, and the corresponding probabilities are $\{P(x_1), P(x_2), \dots, P(x_n)\}$, then the "information entropy" of the variable is:

$$H(X) = -\sum_{i=1}^n P(x_i) \log P(x_i). \quad (1)$$

BAE algorithm first defines the left and right information entropy $LE(s)$ and $RE(s)$ of a string s according to the information entropy:

$$LE(s) = -\sum_{l \in L} P(ls | s) \log P(ls | s), \quad (2)$$

$$RE(s) = -\sum_{r \in R} P(sr | s) \log P(sr | s), \quad (3)$$

where L and R respectively represent the set of word string elements adjacent to the left and right of character s in the corpus. $P(ls|s)$ represents the conditional probability that the left-adjacent string is l when the string s occurs, and $P(sr|s)$ represents the conditional probability that the right-adjacent string is r when the string s occurs.

Assuming that the current string is w_i (the string length of w_i can be several Chinese characters but its initial length is 1), its boundary average entropy is:

$$BAE(w_i) = \frac{1}{2} [LE(w_i) + RE(w_i)]. \quad (4)$$

The boundary average entropy of two consecutive strings w_i and w_j is:

$$BAE(w_i w_j) = \frac{1}{2} [BAE(w_i) + BAE(w_j)]. \quad (5)$$

Similarly, the boundary average entropy of three consecutive strings w_i , w_j and w_k is:

$$BAE(w_i w_j w_k) = \frac{1}{3} [BAE(w_i) + BAE(w_j) + BAE(w_k)]. \quad (6)$$

Next, the function $f(W, n)$ that determines whether the string W made up of $w_1, w_2, \dots, w_n$ can form a word is defined as follows:

$$f(W, 1) = \begin{cases} 1, & \text{freq}(w_i) > \alpha_1, \\ 0, & \text{else} \end{cases} \quad (7)$$

$$f(W,2)=\begin{cases} 1, freq(\overline{w_i w_j}) \geq \alpha_2 \&\& BAE(\overline{w_i w_j}) > BAE(w_i w_j), \\ 0, else \end{cases} \quad (8)$$

$$f(W,3)=\begin{cases} 1, freq(\overline{w_i w_j w_k}) \geq \alpha_3 \&\& BAE(\overline{w_i w_j w_k}) > BAE(w_i w_j w_k) > BAE(w_i w_j w_k), \\ 0, else \end{cases} \quad (9)$$

where $freq(W)$ is the frequency of string W in the corpus, and α_1 , α_2 and α_3 are the parameters we set before training. When the function value is 1, it indicates that the string W consisting of n strings $w_1, w_2, \dots, w_n$ may be a word in the corpus and 0 otherwise. When there is a line over multiple strings (for example, $\overline{w_i w_j}$), they are treated as one string, and the dimension of n decreases. Calculating the $f(\cdot)$ function of each consecutive combination in the long string obtained in step 3.1 yields 1-character, 2-character and 3-character words. Basically two iterations are sufficient for the Chinese language. In this way we can get specific or unlisted words in the field of study, which can not be accomplished by other word segmentation methods based on dictionary matching. To illustrate, with "电源(power source)" and "适配器(adaptor)" represented by w_i and w_j respectively, and the 2-gram string "电源适配器(power adaptor)" as potential word, the determining process is shown in Figure 2:

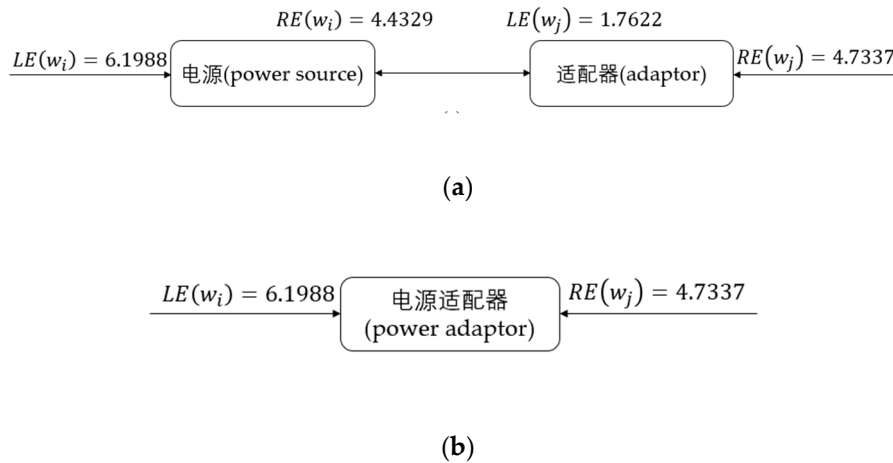


Figure 2. (a) Boundary average entropy before combination of the two strings; (b) Boundary average entropy after combination of the two strings.

- According to Figure 2(a): $BAE(w_i w_j) = \frac{6.1988 + 4.4329 + 1.7622 + 4.7337}{4} = 4.2819$;
- According to Figure 1(b): when w_i and w_j combine to become "电源适配器(power adaptor)", $BAE(\overline{w_i w_j}) = \frac{6.1988 + 4.7337}{2} = 5.4557$.

Since $BAE(\overline{w_i w_j}) > BAE(w_i w_j)$, indicating that the boundary average entropy of combined string is higher, "电源适配器(power adaptor)" can be adopted as a new word to extend the original lexicon..

Next, we use the part-of-speech tagging model of Tsinghua University Lexical Analyzer for Chinese to code the part-of-speech of each word. The words we need in the next step are nouns, adjectives, verbs, adverbs and unknown words. Normally, the words coded as "unknown" are unlisted words that do not exist in the dictionary of common tokenizers and possibly are newly emerging words. After manually abandoning some words irrelevant with the attributes of the product (e.g., one may write in an oral-care review that "my wife likes the make-up mirror", but "wife" is not one of the product attributes and does not contain useful information for product development so it should not be included in the following analysis), we classify all the remaining nouns into several major attributes and these nouns are feature words. In some work, the classification of feature words

is realized by clustering or word co-occurrence. We do it manually for higher accuracy and it does not take much time and effort because a product usually does not have too many feature words people care. Determining whether a word is a feature word costs fewer efforts than determining the attribute it belongs to.

3.3. Calculating the satisfaction score and importance weight for each attribute

To measure users' satisfaction with attributes, we first score their sentiment. There are 3 important points to note when calculating the sentiment score of attributes for each review. First, what attributes are mentioned in the review (Q1)? Second, if an attribute is mentioned (i.e., there are corresponding feature words in the review), is the customer's attitude positive or negative (Q2)? Third, how strong the customer's feelings are (Q3)?

In order to better display the sentiment scoring rules, Table 1 is constructed on the basis of the three questions: **Q1.** In review i , attribute j should have the positive and negative sentiment score of both 0 if it is not mentioned ($att_{ij}^{pos} = 0, att_{ij}^{neg} = 0$). **Q2.** To judge one's attitude towards the attributes mentioned, we build a sentiment dictionary consisting of adjectives (i.e., "优秀(excellent)" "时尚(fashionable)" "物有所值(worthy)") and verbs (i.e., "喜欢(like)" "满意(satisfy)" "讨厌(hate)") we obtained in step 3.2 and classified into positive and negative sentiment words. Then we judge whether the sentiment word adjacent to the feature word f_{ij} is positive or negative and mark the score as $att_{ij}^{pos} = f_{ij}^{pos} = 1$ and $att_{ij}^{neg} = f_{ij}^{neg} = 1$ correspondingly. If there are any privative elements like "几乎不(hardly)" "没有(no)", the score should multiplies $\prod_k w(pri_{ik})$, which is 1 if the number of privative words k is even and -1 otherwise. **Q3.** Adverbs are used to measure how strong is the sentiment. Compared with feature words and sentiment words, degree adverbs, as auxiliary components in sentences, have almost no difference in different fields. We therefore refer to the mature scale of HowNet to divide polarity of emotion into 4 levels (see examples in Table 2). Then the score multiplies $\prod_t Adv_{it}$, where Adv_{it} is the score of level of the adverb t in review i .

Table 1. Illustration of sentiment scoring rules.

	Type of text	Formula	Example
Q1	No feature word	$att_{ij}^{pos} = 0, att_{ij}^{neg} = 0$	"我很满意(I like it very much.)"
	Feature word + positive sentiment word	$att_{ij}^{pos} = f_{ij}^{pos} = 1$	"座椅舒服(The seat is comfortable.)"
	Feature word + negative sentiment word	$att_{ij}^{neg} = f_{ij}^{neg} = 1$	"窗户脏(The windows are dirty.)"
Q2	Feature word + privative words + positive sentiment word	$att_{ij}^{pos} = f_{ij}^{pos} \prod_k w(pri_{ik})$	"不喜欢后备箱(I don't like the trunk.)"
	Feature word + privative words + negative sentiment word	$att_{ij}^{neg} = f_{ij}^{neg} \prod_k w(pri_{ik})$	"价格不贵(The price is not expensive.)"
	Feature word + degree adverbs + positive sentiment word	$att_{ij}^{pos} = f_{ij}^{pos} \prod_t Adv_{it}$	"外观很大气(The appearance is very gorgeous.)"
	Feature word + degree adverbs + negative sentiment word	$att_{ij}^{neg} = f_{ij}^{neg} \prod_t Adv_{it}$	"隔音棉很差(The sound insulation cotton is really poor.)"
Q4	Feature word + privative words + degree adverbs + positive sentiment word	$att_{ij}^{pos} = f_{ij}^{pos} \prod_k w(pri_{ik}) \prod_t Adv_{it}$	"我特别喜欢这个颜色(I love the color very much.)"
	Feature word + privative words + degree adverbs + negative sentiment word	$att_{ij}^{neg} = f_{ij}^{neg} \prod_k w(pri_{ik}) \prod_t Adv_{it}$	"我老婆非常讨厌轮胎(My wife really hates the tires.)"

Table 2. The score of level of adverbs.

	Score	Examples
Score of degree adverbs level	3.0	极度(extremely), 超(super)
	2.0	非常(very), 十分(really)
	1.5	比较(relatively), 颇(relatively)
	0.5	有点(slightly), 稍许(somewhat)

Before applying the sentiment scoring rules for each attribute in a review, we need to process the review sentences. First, most of the reviews are not just talking about one attribute so it is important to separate each feature-sentiment word pair. We divide each review into several sentences by commas, full stops and other typical sentence separators. Chinese is different from Germanic languages in which subject, predicate and object constitute a strictly self-contained sentence. Chinese comma often serves as sentence boundaries. In addition, adversative words and conjunctions (words like "but" and "and") interfere with subsequent feature-sentiment word matching, so we treat them as follows: if there is a complete feature-sentiment word pair before and after the disturbance terms (for example, in review "空调十分凉快但是窗户太小了(The air conditioner is really cool but the windows are so small)", there are two feature-sentiment word pairs: "空调-凉快(air conditioner-cool)" and "窗户-小(window-small)" separated by the word "但是(but)", we treat the disturbance terms as separators and divide one sentence into two; If the feature words only appear in the preceding part of the disturbance term, we add the feature words mentioned above after it and then cut it. For example, in a review "窗户看起来很干净但是有点小(the window looks very clean but a little bit small)", there is no feature word after "但是(but)", we add another "窗户(window)" to it so that it forms two feature-sentiment word pairs, namely, "窗户-干净(window-clean)" and "窗户-小(window-small)". Each Chinese sentence obtained in this way will finally contain only one feature word and one sentiment word, and the score can be calculated by the sentiment scoring rules mentioned above. If several sentences of a review discuss the same attribute and hold with the same attitude, we end up with the highest score. And finally, we get a matrix-like Table 3.

Table 3. Example of sentiment scores of attribute j in review i .

Review	att_{i1}^{pos}	...	att_{ij}^{pos}	...	att_{in}^{pos}	att_{i1}^{neg}	...	att_{ij}^{neg}	...	att_{in}^{neg}
$i = 1$	1.5	...	0	...	2.0	0	...	1.0	...	1.5
$i = 2$	2.0	...	1.0	...	0	1.0	...	1.5	...	0
...	...	...	...	...	...	...	...	...	...	...
$i = m$	0	...	1.5	...	1.0	0	...	2.0	...	1.0

In this study, we measure the performance of an attribute by the following method. As long as a review contains one of the feature words of this attribute, no matter how many there are, we will calculate it as 1. By adding all of the 1s together, we get the number of reviews mentioning this attribute, and we mark this figure as f_{req_j} . The satisfaction scores of attributes are calculated based on the sentiment scores of attributes in each review we have already learned in Section 3.3. We sum up the positive and negative scores for each attribute by adding Table 2 vertically, and get $[att_1^{pos}, att_2^{pos}, \dots, att_n^{pos}, att_1^{neg}, att_2^{neg}, \dots, att_n^{neg}]$. And the equation for calculating the performance of each attribute per_j and adjusted metrics after rescaling to $[0,1]$ $AdjustedPer_j$ are expressed as follows:

$$per_j = \frac{att_j^{pos} - att_j^{neg}}{freq_j}, \quad (10)$$

$$AdjustedPer_j = \frac{per_j - \min_j per_j}{\max_j per_j - \min_j per_j}. \quad (11)$$

To determine the users' importance weight for each attribute, we apply TF-IDF method, which is a metric used to reflect the significance of a word in a document [30]. TF-IDF is generated by two metrics, namely, term frequency (TF) and inverse document frequency (IDF). The former represents the total number of a word that occurs in the document, while the latter is designed to attenuate the effect of meaninglessness resulting from a too high occurrence. Mathematically, given a product p , a product collection P , for an attribute j , the TF , IDF , and imp that determines the importance weight of the attribute are defined as:

$$TF_j = \frac{|f_j|}{\sum_j |f_j|}, \quad (12)$$

$$IDF_j = \log \frac{|f_j|}{1 + |\{p \in P : f_j \in d_p\}|}, \quad (13)$$

$$imp_j = TF_j \times IDF_j, \quad (14)$$

where f_j are the feature words of attribute j , and d_p is the text document of p .

3.4. Product improvement orientation identification

The satisfaction score of attributes we obtained by matching feature words, sentiment words and other auxiliary terms can tell manufacturers the strength and the weakness of their products so that they have a correct direction on what to be maintained and improved to increase customer satisfaction. However, resources are always limited so manufacturers have to wisely allocate them to those with a higher importance weight, which is the core of the importance-performance analysis (IPA) model [92].

In this study, we refer to the adjusted IPA model which integrates gap analysis to better understand the priorities for corrective products improvement actions and develop the importance-satisfaction gap analysis model (ISGA model). Figure 3 shows the basic zone distribution of the model. The median value of importance weight and the median value of satisfaction score are used as coordinate cross-sections. To avoid subjective bias due to the evaluation of the cross-section of coordinates, a diagonal line at an angle of 45° is added, where the values of satisfaction and importance are equal (no gap). As shown in the figure, the four quadrants are re-divided into four zones. Zone 1 and Zone 2 contain elements that should be allocated resources from Zone 3 and Zone 4. Zone 1 is the same as the second quadrant, where elements have high importance but low user satisfaction. Zone 2 consists of the two triangles of the first and the third quadrants above the diagonal line. Elements within Zone 1 have higher resource allocation priorities than elements within Zone 2, i.e. attention should be paid first. Zone 3 consists of the two triangles of the first and the third quadrant below the diagonal line. Zone 4 is the same as the fourth quadrant, where the importance is low but possibly overperform. Elements in Zone 4 have higher resource adjustment priority in relation to those in Zone 3. And for each zone, the distance between the element and the coordinate of the intersection determines the resource allocation/adjustment priorities for actions. Suppose the cross-section of the coordinate is (x_M, y_M) , we measure the distance between M and element (x_*, y_*) using the Euclidean distance formula:

$$d(*) = \sqrt{(x_* - x_M)^2 + (y_* - y_M)^2}. \quad (15)$$

For example, in Figure 4, for elements in Zone 1 and Zone 2, the priority for resource allocation is $A > B > C > D$; for elements in zone 3 and 4, the priority for resource adjustment is $E > F > G > H$.

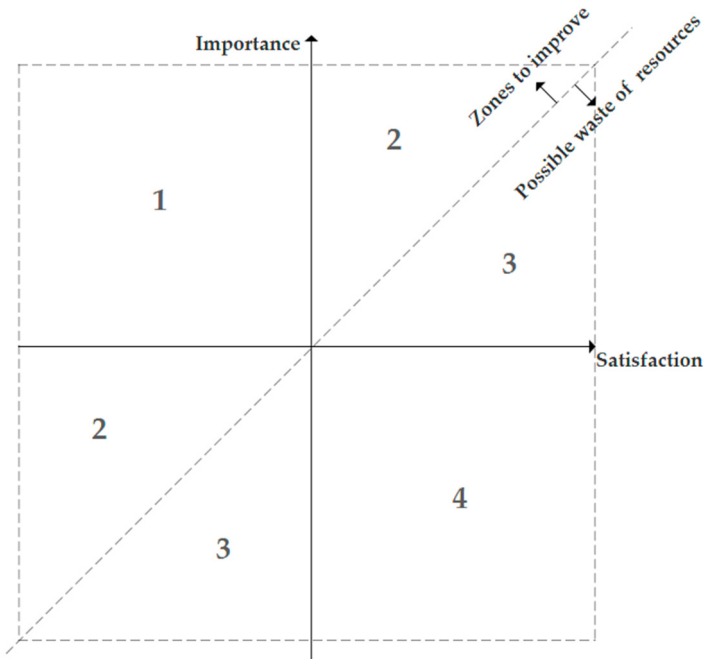


Figure 3. Zone distribution of the ISGA model.

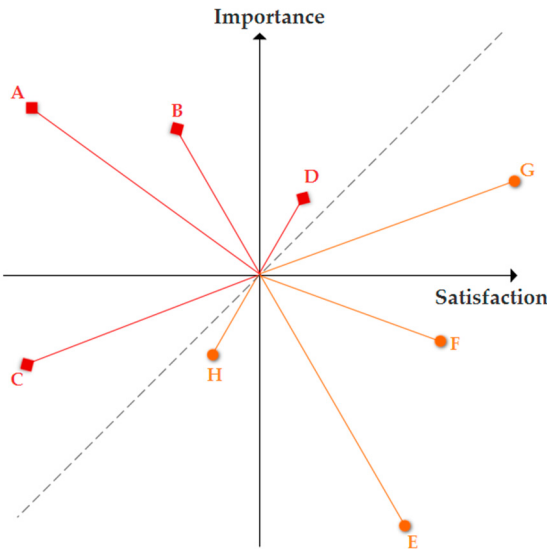


Figure 4. Demonstration of zone priority.

Considering that dynamical changes of importance-satisfaction-gap provide insights into evolving trends of users' tastes, we divide data by different periods. Then a 3D ISGA-time model is constructed with time as the x-axis, customer satisfaction(attribute performance) as the y-axis and attribute importance as the z-axis. See Figure 4, The shaded bevel separates zones to improve from zones that possibly waste resources. The evolvement of consumer tastes of attributes and product improvement orientation thus can be learned through analysis. Specific discussion will be presented through the next section of our application on the automobile market.

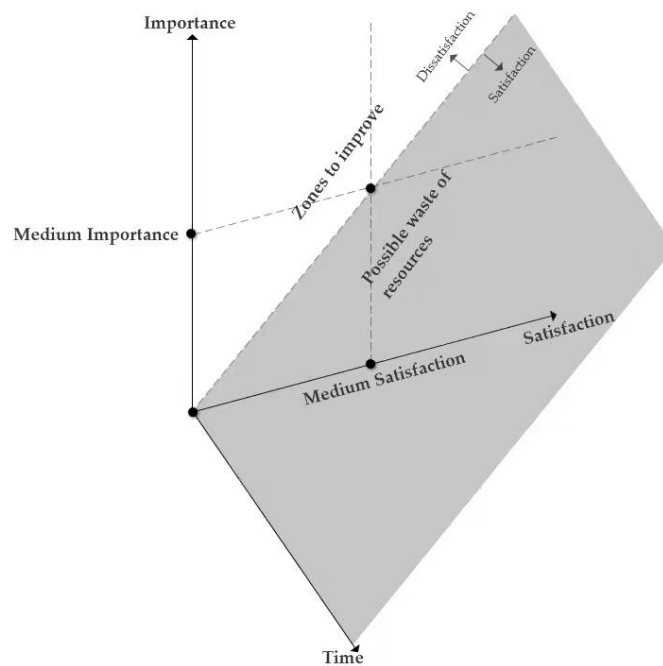


Figure 5. The base ISGA-time model.

4. Empirical Study on The Automobile Industry

4.1. Product improvement orientation identification

In this section, we illustrate the proposed product improvement orientation analysis method in the field of automobile online review and gain insights into the challenges and opportunities of the Chinese automobile market. With the method proposed above, we use content generated by automobile users from autohome.com to address three concerns: Which attributes are most important to automobile users? How satisfied are the automobile-users are with these attributes? How does consumers' satisfaction with these attributes and attribute importance change over time?

The products under study should be as similar as possible in function and price so that manufacturers can compete with similar producers with limited resources. 5,171 data are collected from autohome.com for a small-sized sedan. Its purchase price ranges from RMB 60,000 to RMB 80,000 (around USD 9,400 to USD 12,600). It came into the market in 2017 and was still for sale at the time when the study was conducted in December 2021. All of the review contents are written in Chinese.

4.2. Feature words identification

In this study, the main tool we use for data processing is Python 3.8.5. By removing all the non-Chinese ideographs from the reviews and concatenating the remaining elements, we obtained strings of 1,086,924 Chinese characters. After identifying words by iterating the BAE algorithm twice with $\alpha_1 = 5$, and $\alpha_2, \alpha_3 = 3$ and tagging, we only leave nouns, adjectives, adverbs, verbs and unknown words.

two participants with the experience of driving the car were asked to identify feature words from the nouns. They were told that feature words must be related to the attributes of the car. Any words like "reason" "family" that do not contain information on product development should not be regarded as feature words. For the words that the two persons hold different opinions, we asked a third participant to judge it again. There are not as many professional words as there are everyday words so the time for coding is only approximately 2.5 hours (less than 3 seconds for a word on average) for each participant. And finally, a total of 314 words were identified as feature words. Since some of the feature words belong to the same attribute, we finally sorted out 119 attributes. Therefore,

some attributes are made up of several feature words. A sentiment dictionary is built on adjectives and verbs. The construction of the sentiment dictionary needs to judge whether the word is a sentiment word and its emotional direction. There are only two categories of emotion direction (i.e., positive vs. negative. In the Chinese language context, neutral words always express a tinge of dissatisfaction). For unregulatory feature words that point to an adverse direction of emotion when using the emotional matching rules, we build another sentiment dictionary for them. Following the rules proposed in Section 3.3, for each attribute of the car, satisfaction scores were determined using the sentiment scoring rules and importance weights were calculated using the TF-IDF method.

4.3. Importance-satisfaction gap analysis

Due to space limitation, only twenty top-mentioned attributes are selected for importance-satisfaction gap analysis in this part. Based on all the reviews, we drew an ISGA map for the car. With users' satisfaction as the horizontal axis and importance as the vertical axis and the median of both as the origin, the coordinate matrix is established. We add a 45-degree slash to each of the four zones to indicate different resource allocation/adjustment priorities. The upper-left zone (Zone 1) is the one that requires the most attention, and the lower-right zone (Zone 4) element is the one that attention can be reduced.

Figure 6 shows that price, fuel consumption and sound insulation in Zone 1 are the attributes that users of this small-sized sedan consider very important but are not satisfied with. Among them, price is furthest from the origin, which is the attribute that needs the most attention. Fog light, Model appearance, seat, and Engine dynamics are distributed in Zone 2, and they can be improved after considering the resource allocation for Zone 1 attributes. Shock absorber, underpan, and smell are in Zone 4, indicating that they are over-resourced, and manufacturers can reduce resource investment in these attributes and switch to Zone 1 and Zone 2 attributes. Wheel hub, audio, air conditioner, brand popularity, steering wheel, trunk size, horn, gearbox, trim and interior layout are attributes of Zone 3. Manufacturers can adjust resources in this zone after adjusting resources in Zone 4.

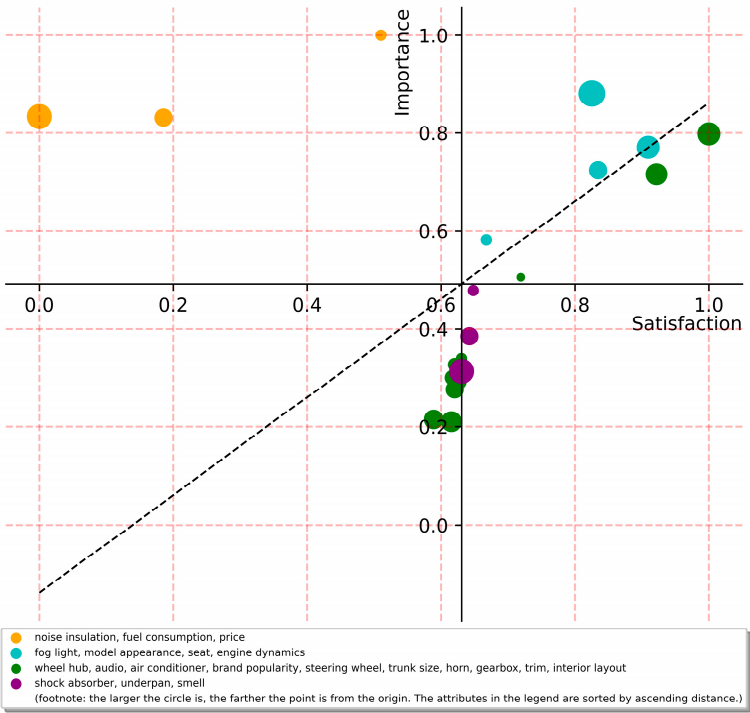


Figure 6. Importance-satisfaction value distribution of the small-sized sedan attributes.

4.4. ISGA-time analysis

In order to analyze the development trend of customer demand for different attributes, ISGA analysis is needed for different periods of time. Therefore, the dynamic ISGA-Time model is constructed in this section to analyze the changing trend of the importance of user demand and satisfaction of the small-sized sedan.

We divide the reviews 5 data sets by time period and construct an ISGA-Time analysis chart (see Figure 7) with time as X-axis, customer satisfaction as Y-axis and importance as Z-axis. The marker on the folding point represents the zone where the attribute is located in that year. At the same time, we also draw the diagram of satisfaction changing over time (Figure 8(a)), and the diagram of importance changing over time (see Figure 8(b)).

For the users of the small-sized sedan, it can be seen that the importance of interior layout, model appearance, price, trim, engine dynamics, smell and underpan have all decreased. However, price is always in Zone 1 or Zone 2, indicating that users' satisfaction with this attribute is still low, that is, they think the price is unreasonable. We tracked down reviews about the price and found that most people thought that even though the price was not high, the specs were not worth it. Moreover, compared with other vehicles entering the market during the same period, the price reduction of this vehicle is lower, and the discount is too small. This attribute became less important to people when the attribute of engine switch from Zone 1 and Zone 2 to the last year's Zone 4 (higher satisfaction was harder to come by because the engine was immutable when it was designed). The importance of fuel consumption, seat, sound insulation, smell and underpan is on the rise. Fuel consumption and sound insulation are almost maintained in Zone 1 in recent years, which belong to properties with high importance but low customer satisfaction, and are in urgent need of resource input and management. The remaining attributes belong to Zone 3 and Zone 4.

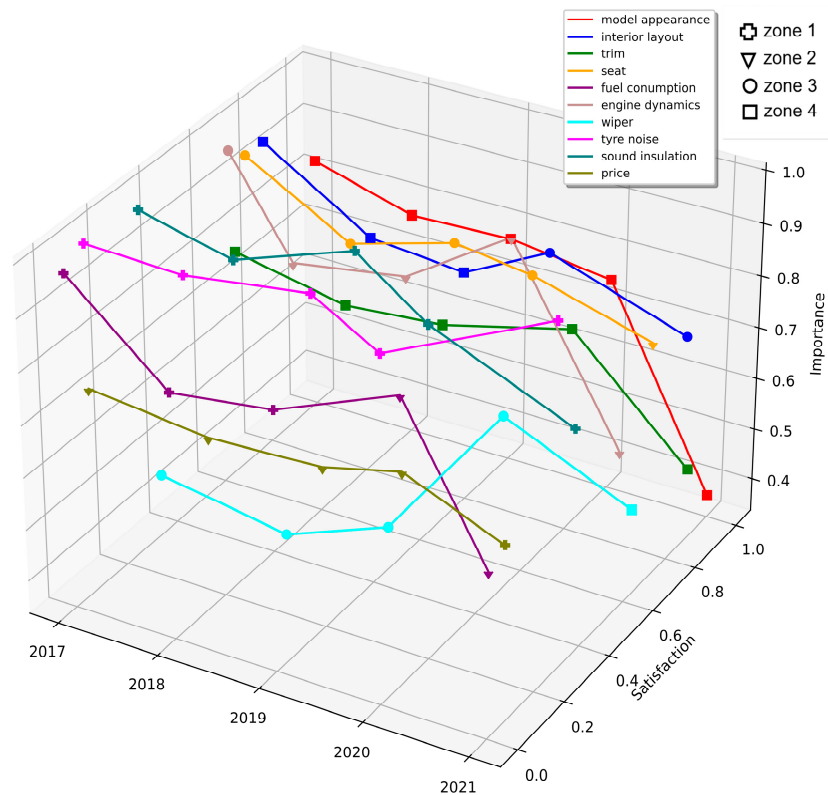


Figure 7. Importance-satisfaction distribution of the small-sized sedan over time.

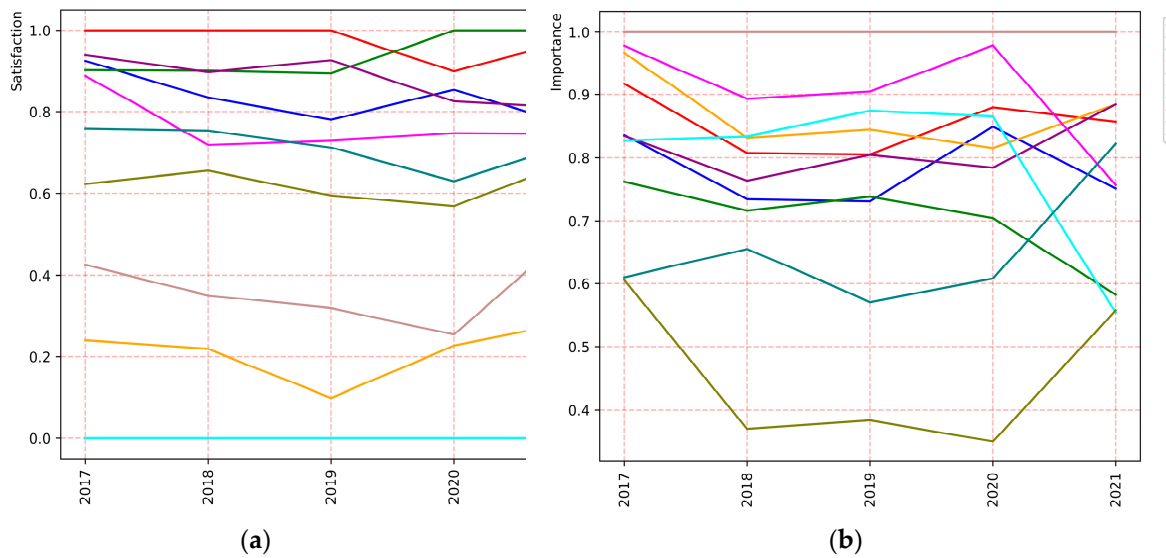


Figure 8. (a) Changes of the car users' satisfaction of the small-sized sedan over time; (b) Changes of attributes importance of the small-sized sedan over time.

We also draw a thermodynamic diagram according to the distance of each attribute from the origin in every year (see Figure 9). Zone 1 and Zone 2 are in red, which are dissatisfactory zones that need to be allocated resources. The blue represents Zone 3 and Zone 4, and is the satisfying zones where resources can be reduced. The farther away from the origin, the darker the square color and the higher the priority of resource allocation/adjustment. As can be seen, most attributes remain in the same color scheme over time.

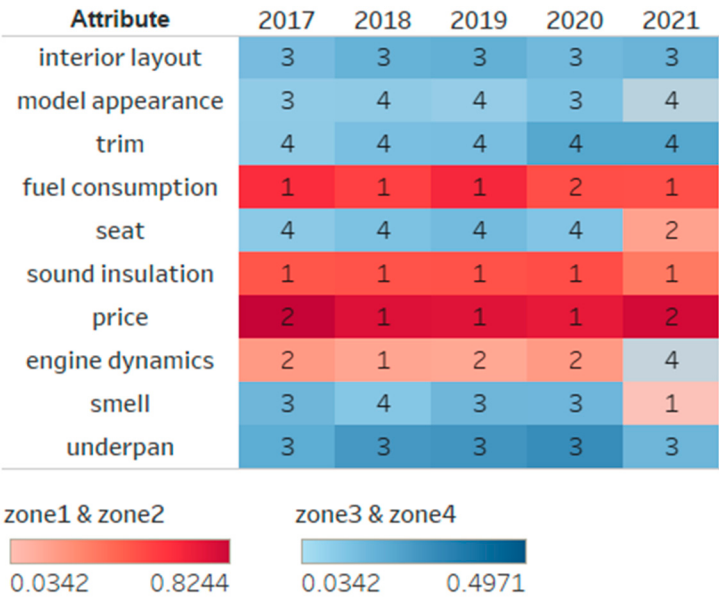


Figure 9. Changes of distance and zones of the small-sized sedan.

In order to better describe the changing behavior of attributes, the product attributes whose importance weight changes outside the range of $[-0.1,0.1]$ are defined as active attributes. The 119 attributes in Figure 10 are sorted by how often they appear in the reviews. It is clear that there is little change in these attributes between 2018 and 2019. While from 2017 to 2018 and from 2019 to 2021 there are more active attributes, especially those mentioned frequently. And the change of their importance weight is positive or negative, indicating that some attributes become more important to people, while others decrease. From 2020 to 2021, the attributes that become significantly more important are seat, smell,

underpan, fog light, back space, dashboard, ,rearview mirror, atenna, armrest box, iron sheet, disc brake, water thermometer and tire noise; and the attributes that become significantly less important are trim, price, engine dynamics, radio, air conditioner, steering wheel, gearbox, wheel hub, exhaust pipe and window. Figure 11 shows the dynamic changes of product attribute importance weights and satisfaction scores over time. Taking the vehicle attribute Engine Dynamics as an example, it can be seen from Figure 11 that the importance of this attribute increased slightly (positive) from 2019 to 2020. However, while people's satisfaction with it remained almost the same, its importance decreased greatly from 2020 to 2021. So manufacturers can focus less on this attribute. The change of importance of some attributes has been maintained at 0, such as sound insulation, indicating that people have always attached the same importance to this attribute.

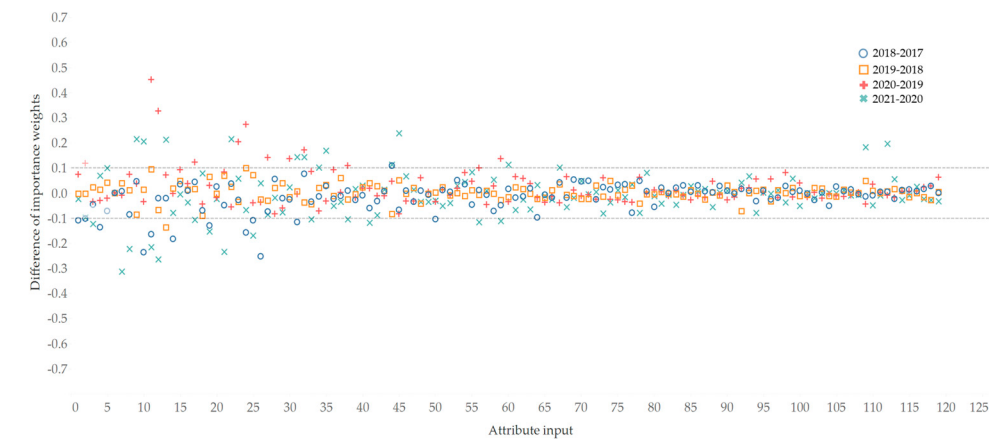


Figure 10. Attribute importance weight difference of the small-sized sedan over the four periods.

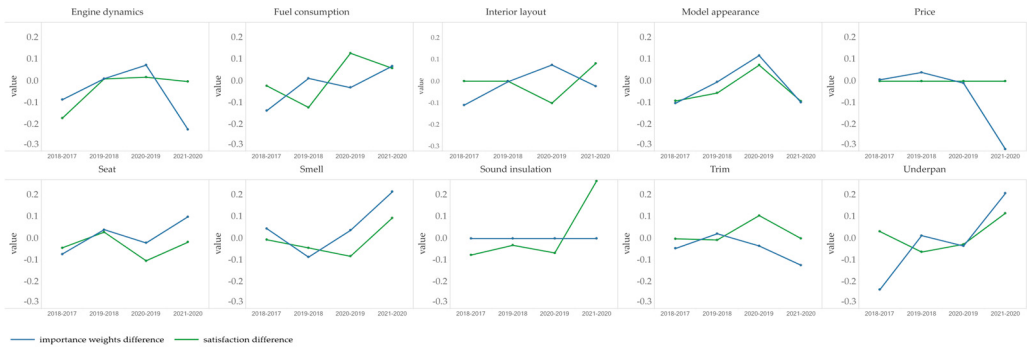


Figure 11. Dynamic performance of importance weights and satisfaction.

5. Evaluation of the Method

In order to evaluate the proposed method, we measure it from three aspects: the word segmentation performance of the BAE algorithm, the accuracy of feature-sentiment word pair matching, and the accuracy of emotional direction of sentiment words. We evaluated the result with 100 of the 5,171 review data.

5.1. Word segmentation performance

We manually took the word segmentation of 100 reviews and counted them as the correct result. We compare the word segmentation results of the 100 reviews obtained by jieba, Thulac and BAE algorithms in Section 4.2 (it should be noted that the BAE algorithm obtains the words based on the

corpus of the whole set of reviews, not only the 100). In this paper, three of the most commonly used indicators, precision, recall and F_1 of the second International Chinese Word Segmentation Assessment are adopted. The calculation methods are as follows:

$$\text{Precision} = \frac{\text{the number of words correctly segmented}}{\text{total number of words segmented}} \times 100\%, \quad (15)$$

$$\text{Recall} = \frac{\text{the number of words correctly segmented}}{\text{the real total number of words}} \times 100\%, \quad (16)$$

$$F_1 = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \times 100\%. \quad (17)$$

The results of different word segmentation methods are shown in Table 4. It can be seen that the BAE word segmentation algorithm performs the best in all indicators

Table 4. Alternative methods to identify Chines word boundaries.

	Precision	Recall	F_1
jieba	70.83%	68.77%	69.79%
THULAC	65.98%	62.40%	64.13%
BAE	72.54%	69.69%	71.08%

5.2. Matching accuracy rate of feature-sentiment words

To calculate the satisfaction score of each attribute, we automatically matched sentiment words for each feature word in each review based on sentence delimiters, Chinese transitional words and conjunctions. There may be two kinds of error of this method: (1)the sentiment words matched by a feature word that belong to others, or (2)there are sentiment words for a feature in the sentence, but no match is found. We first manually identify feature-sentiment word pairs for 100 reviews, and then compare the results with those obtained in Section 4.2. The number of feature words correctly matched is 1,060 and the total frequency of feature words is 1,324. So the accuracy rate is approximately 80.06%. Considering the high cost of manual recognition, the overall performance of this method is relatively good.

5.3. Evaluation of accuracy rate of emotional direction

The plus-minus of satisfaction may be wrong even if feature-sentiment words match correctly,. The collocation of the same sentiment word with different feature words may be positive or negative. Even two feature words with the same attribute, such as "价格(price)" and "性价比(cost performance)", which belong to the same attribute, when matched with the sentiment word "低(low)", the emotional direction will be different. Therefore, we further investigate whether the correct emotional direction is given to the sentiment words matched by the feature words under an attribute when calculating the satisfaction score of an attribute in each review. Among the 1,060 feature words correctly matched with sentiment words, 955 of them were correct in emotional direction, with an accuracy rate of 90.09%.

6. Conclusions

In order to remain competitive in the market, it is necessary for an enterprise to effectively prioritize the improvement of its product or service attributes and allocate resources appropriately to improve customer satisfaction. To deal with the high-speed developing demand of Chinese automobile users, we propose a method to measure their perceptions of attributes quality of different types of cars, which employ bound average entropy algorithm to identify domain professional vocabularies and importance-satisfaction gap analysis to judge Chinese car users' changes in taste

over time. By using the BAE algorithm, we avoid the disadvantage of the frequently used word segmentation system, which heavily depends on a pre-established lexicon and is hardly able to distinguish new terms from a new corpus, especially in an industry like automobile that derives specific terminologies very quickly. Through the importance-satisfaction gap analysis, we evaluate the allocation of resources to car attributes. We also add time as an element to allow us to discuss the dynamic change of importance and satisfaction, so as to provide future product improvement orientation for manufacturers. With the Chinese reviews on automobile.com, the vertical forum where people freely share cars' information that rarely involve logistics and sellers, opinion words were extracted and satisfaction and importance value were calculated based on them. Visual analysis follows based on the ISGA model to predict future trends. In actual production activities, this method provides a potential tool for enterprises to meet the real needs of consumers and allocate resources rationally by analyzing customers' opinions published online. This provides powerful support for product upgrades and improvements.

Overall, this study enriches the field of mining opinions from user-generated content. Three main aspects of novelty are exhibited: (1) the implementation of BAE algorithm on automobile domain professional vocabularies identification; (2) the ISGA analysis of automobile attributes for resources adjustment; and (3) the presentation of ISGA-time model analysis of future designing orientation.

The framework of this study has to be seen in light of some limitations. First, taking customer demand of different attributes brought about by automobile types into consideration, we conducted ISGA analysis on three types of automobile attributes separately. In fact, other factors can lead to different demands on automobile attributes. For instance, customers living in different regions will have differences in income levels, infrastructure and transportation conditions, all of which can generate different customer demand on automobile attributes. In addition, factors from customers themselves namely gender and age can have a similar effect as external factors like regions and automobile types. Due to data limitations, we didn't obtain other detailed dimensions for classification in the process of ISGA analysis. Therefore, future studies can leverage other dimensions to classify the attributes data before analysis in order to give suggestions for enterprises on prioritizing the improvement of its product/service features. Second, the employment of the BAE algorithm can improve word segmentation precision while posing a negative impact on recall, and this algorithm has high time complexity. And in the step of sentiment dictionary matching, the preset score for each degree adverb when building sentiment lexicon is subjective to some extent. Future studies can make further improvements based on product specificity and context grammar structure.

Author Contributions: H.L., supervision, conceptualization, funding acquisition, methodology, validation, writing—review and editing; W.Z., conceptualization, methodology, investigation, resources, data curation, writing—original draft preparation; W.S., software, validation, investigation; X.L., project administration, resources, methodology, visualization; S.Y., software, formal analysis, validation, visualization. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Discipline Co-construction Project for Philosophy and Social Science in Guangdong Province (No. GD20XGL03), the Universities Stability Support Program in Shenzhen (No. 20200813151607001), the Major Planned Project for Education Science in Shenzhen (No. zdfz20017), the National Natural Science Foundation of China (No. 71901151).

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflicts of interest.

Appendix A

Table A1. Research that share similar purpose with this study on Web of Science.

Author(year)	Title	Publication Source	Type of UGC and category
Timoshenko, A and Hauser, JR. (2019) [6]	Identifying customer needs from user-generated Content	MARKETING SCIENCE 38(1), 1-20	Oral care reviews from amazon
Rasool, G.; Pathania, A. (2021) [35]	Reading between the lines: untwining online user-generated content using sentiment analysis	JOURNAL OF RESEARCH IN INTERACTIVE MARKETING 15(3), 401-418	Airline passenger reviews from TripAdvisor
Fels, A.; Briele, K.; Ellerich, M.; Schmitt, R. (2018) [36]	Extracting customer-related information for need Identification	INTERNATIONAL CONFERENCE ON HUMAN SYSTEMS ENGINEERING AND DESIGN: FUTURE TRENDS AND APPLICATIONS IHSED2018 876 , 1108-1112	Vacuum cleaner and smart phone reviews from Amazon
Kühl, N.; Mühlthaler, M.; Goutier, M. (2019) [37]	Supporting customer-oriented marketing with artificial intelligence: automatically quantifying customer needs from social media	ELECTRONIC MARKETS 30(2), 351-367	Electric vehicle tweets
Lee, JYH; Yang, CS and Chen, SY (2017) [38]	Understanding customer opinions from online discussion forums: A design science framework	ENGINEERING MANAGEMENT JOURNAL 29(4), 235-243	Discussions about cars from Mforum
Vollero, A; Sardanelli, D; Siano, A (2021) [39]	Exploring the role of the Amazon effect on customer expectations: An analysis of user-generated content in consumer electronics retailing	JOURNAL OF CONSUMER BEHAVIOUR	Amazon electronic reviews & facebook discussions about electronic
Zhu, D; Lappas, T and Zhang, JH .(2018) [40]	Unsupervised tip-mining from customer reviews	DECISION SUPPORT SYSTEMS 107, 116-124	Tripadvisor hotel reviews
Ekhlassi, A; Zahedi, A. (2018) [41]	A unique method of constructing brand perceptual maps by the text mining of multimedia consumer reviews	INTERNATIONAL JOURNAL OF MOBILE COMPUTING AND MULTIMEDIA COMMUNICATIONS 9(3), 1-22	Amazon digital tablets reviews
Yu, CE; Zhang, XY (2020) [42]	The embedded feelings in local gastronomy: a sentiment analysis of online reviews	JOURNAL OF HOSPITALITY AND TOURISM TECHNOLOGY 11(3), 461-478	Online reviews of restaurants from one of the most popular

			tourism website (not mentioned name)
Hsiao, YH; Chen, MC; Liao, WC. (2017) [43]	Logistics service design for cross-border E-commerce using Kansei engineering with text-mining-based online content analysis	TELEMATICS AND INFORMATICS 34(4), 284-302	Cross-boarder e-commerce online reviews
Zhang, R.; Pang, Z.; Liu, X. (2021) [44]	Mining express service innovation opportunity from online reviews	JOURNAL OF ORGANIZATIONAL AND END USER COMPUTING (JOEUC) 33(6)	Online reviews of express delivery from Baidu Koubei
Valsan, A; Sreepriya, CT; Nitha, L. (2017) [45]	Social media sentiment polarity analysis: A novel approach to promote business performance and consumer decision-making	ARTIFICIAL INTELLIGENCE AND EVOLUTIONARY COMPUTATIONS IN ENGINEERING SYSTEMS, ICAIECES 2016 517, 1-12	Online cameras reviews
Hasan, MR.; Abdunurova, A.; Wang, W.; Zhang, J.; Shams, S.M.R. (2021) [46]	Using deep learning to investigate digital behavior in culinary tourism	JOURNAL OF PLACE MANAGEMENT AND DEVELOPMENT 14(1), 43-65	Online restaurants reviews from TripAdvisor
Kauffmann, E. (2019) [47]	A step further in sentiment analysis application in marketing decision-making	RESEARCH & INNOVATION FORUM 2019: TECHNOLOGY, INNOVATION, EDUCATION, AND THEIR SOCIAL IMPACT, 211-221	Online reviews of cell phones and accessories from Amazon
Vinodhini, G.; Chandrasekaran, RM. (2014) [48]	Measuring the quality of hybrid opinion mining model for e-commerce application	MEASUREMENT 55, 101-109	Publicly available reviews of digital cameras from University of Chicago dataset
Chalupa, S.; Petricek, M.; Chadt, K. (2021) [49]	Improving service quality using text mining and sentiment analysis of online reviews	QUALITY-ACCESS TO SUCCESS 22(182) , 46-49	Online hotel reviews from various booking sites
Dickinger, A and Mazanec, JA (2015) [50]	Significant word items in hotel guest reviews: A feature extraction approach	TOURISM RECREATION RESEARCH 40 (3) , 353-363	Online hotel reviews from tripadvisor
Asghar, Z.; Ali, T.; Ahmad, I.; Tharanidharan, S.; Nazar, S.K.A.; Kamal, S. (2018) [51]	Sentiment analysis on automobile brands using twitter Data	INTELLIGENT TECHNOLOGIES AND APPLICATIONS, INTAP 2018 932 , 76-85	Twitter of automobiles
Aman, JJC.; Smith-Colin, J.; Zhang, WW. (2021) [52]	Listen to e-scooter riders: Mining rider satisfaction factors from app store reviews	TRANSPORTATION RESEARCH PART D-TRANSPORT AND ENVIRONMENT	E-scooter rider app store reviews
Ng, CY.; Law, KMY. (2020) [53]	Investigating consumer preferences on product designs by analyzing	COMPUTERS & INDUSTRIAL ENGINEERING 139	Comments on smart phones from Facebook

	opinions from social networks using evidential reasoning		
Becken, S.; Alaei, AR.; Wang, Y. (2019) [54]	Benefits and pitfalls of using tweets to assess destination sentiment	JOURNAL OF HOSPITALITY AND TOURISM TECHNOLOGY 11(1), 19-34	Tourism tweets
Wang, W.; Feng, Y.; Dai, W. (2018) [55]	Topic analysis of online reviews for two competitive products using latent Dirichlet allocation	ELECTRONIC COMMERCE RESEARCH AND APPLICATIONS 29, 142-156	Online reviews of wireless mice from Amazon.com
Zhu, D.; Lappas, T.; Zhang, J. (2018) [56]	Unsupervised tip-mining from customer reviews	DECISION SUPPORT SYSTEMS 107, 116-124	Travel guides reviews from TripAdvisor
Al-Obeidat, F.; Spencer, B.; Kafeza, E. (2018) [57]	The opinion management framework: Identifying and addressing customer concerns extracted from online product reviews	ELECTRONIC COMMERCE RESEARCH AND APPLICATIONS 27, 52-64	Online hotel reviews
Vo, A.; Nguyen, Q.; Ock, C. (2018) [58]	Opinion-aspect relations in cognizing customer feelings via reviews	IEEE ACCESS 6, 5415-5426	Cameras reviews from dataset and SemEval-2016 laptop reviews
Oh, Y.K.; Yi, J. (2021) [59]	Asymmetric effect of feature level sentiment on product rating: an application of bigram natural language processing (NLP) analysis A machine learning approach	INTERNET RESEARCH	Reviews of wireless earbud products on Amazon.com
Singh, A.; Tucker, C.S. (2017) [60]	to product review disambiguation based on function, form and behavior classification	DECISION SUPPORT SYSTEMS 97, 81-91	Laptop reviews from amazon.com
Eldin, S.S.; Mohammed, A.; Hefny, H.; Ahmed, A.S.E. (2021) [61]	An enhanced opinion retrieval approach on Arabic text for customer requirements expansion	JOURNAL OF KING SAUD UNIVERSITY-COMPUTER AND INFORMATION SCIENCES 33(3), 351-363	Several social product resources including souq.com
Riaz, S.; Fatima, M.; Kamran, M.; Nisar, N.W. (2019) [62]	Opinion mining on large scale data using sentiment analysis and k-means clustering	CLUSTER COMPUTING-THE JOURNAL OF NETWORKS SOFTWARE TOOLS AND APPLICATIONS 22, S7149-S7164	Reiveiws of camera, mobile phone, laptop, tablet, tv, and video surveillance devices from Amazon, Ebay, Alibaba
Jin, J.; Ji, P.; Liu, Y. (2015) [63]	Translating online customer opinions into engineering characteristics in QFD: A	ENGINEERING APPLICATIONS OF ARTIFICIAL INTELLIGENCE 41, 115-127	Printer reviews from Amazon.com

	probabilistic language analysis approach Mining online reviews with a		
Jin, J; Jia, D.P.; Chen, K.J. (2021) [64]	Kansei-integrated Kano model for innovative product design	INTERNATIONAL JOURNAL OF PRODUCTION RESEARCH, 1-20	Smartphones reviews
Zhang, L.; Chu, X.N.; Xue, D.Y. (2019) [65]	Identification of the to-be-improved product features based on online reviews for product redesign	INTERNATIONAL JOURNAL OF PRODUCTION RESEARCH 57(8), 2464-2479	Smartphones reviews
Zhou, F.; Jiao, R.J.; Linsey, J.S. (2015) [66]	Latent customer needs elicitation by use case analogical reasoning from sentiment analysis of online product reviews	JOURNAL OF MECHANICAL DESIGN 137(7), 071401	Kindle fire hd reviews from Tablet
Zhou, F.; Ayoub, J.; Xu, Q. Jessie Yang, X. (2020) [67]	A machine learning approach to customer needs analysis for product ecosystems	JOURNAL OF MECHANICAL DESIGN 142(1), 011101	Kindle fire tablets reviews from Amazon
Liu, Y.; Jin, J.; Ji, P.; Harding, J.A.; Fung, R.Y. (2013) [68]	Identifying helpful online reviews: A product designer's perspective	COMPUTER-AIDED DESIGN 45(2), 180-194	Phone reviews collected from Amazon
Ireland, R.; Liu, A. (2018) [69]	Application of data analytics for product design: Sentiment analysis of online product reviews	CIRP JOURNAL OF MANUFACTURING SCIENCE AND TECHNOLOGY 23, 128-144	Reviews on coleman chair from Amazon
Joung, J.; Jung, K.; Ko, S.; Kim, K. (2019) [70]	Customer complaints analysis using text mining and outcome-driven innovation method for market-oriented product development	SUSTAINABILITY 1(1), 40	Reviews of stand-type air conditioners
Jain, P.K.; Pamula, R.; Srivastava, G. (2021) [71]	A systematic literature review on machine learning applications for consumer sentiment analysis using online reviews	COMPUTER SCIENCE REVIEW 41, 100413	Reviews cover many industries (Hotel, Airline, Restaurant, Airport, Tourist, Art and Musuem)
Foris, D.; Crihalmean, N.; Foris, T. (2020) [72]	Exploring the environmental practices in hospitality through booking websites and online tourist reviews	SUSTAINABILITY 12(24), 10282	Online hotel reviews from booking.com
Han, Y.; Moghaddam, M. (2021) [73]	Eliciting attribute-level user needs from online reviews	JOURNAL OF MECHANICAL DESIGN 143(6), 061403	Online sneaker reviews

	with deep language models		
	and information extraction		
	An approach to discovering		
	product/service Improvement		Reviews of Huawei P series
Wu, J.; Wang, Y.; Zhnag, R.; Cai, J. (2018) [74]	priorities: Using dynamic importance-performance analysis	SUSTAINABILITY 10(10), 3564	smartphones from Jingdong.com
	Extracting product features		Customer reviews of 5
Htay, S.S.; Lynn, K.T. (2013) [75]	and opinion words using pattern knowledge in customer reviews	SCIENTIFIC WORLD JOURNAL, 1-5	electronic products: Two digital cameras, one dvd player, one mp3 player, and one cellular phone from Amazon
	Supporting the construction		
	of affective product		
Wang, W.M.; Tian, Z.G.; Li, Z.; Wang, J.W. (2019) [76]	taxonomies from online customer reviews: an	JOURNAL OF ENGINEERING DESIGN 30(10-12), 445-476	Reviews on toys and games on Amazon
	affective-semantic approach		
	Augmenting feature model		
Zhou, F.; Jiao, J.R.; Yang, X.J.; Lei, B. (2017) [77]	through customer preference mining by hybrid sentiment analysis	EXPERT SYSTEMS WITH APPLICATIONS 89, 306-317	Reviews of the first generation kindle fire hd tablets from Amazon
	What makes consumers		
Jin, J.; Ji, P.; Kwong, C.K. (2016) [78]	unsatisfied with your products: Review analysis at a fine-grained level	ENGINEERING APPLICATIONS OF ARTIFICIAL INTELLIGENCE 47, 38- 48	Customer reviews of six mobile phones from Amazon
	OpinionSeer: interactive		
Wu, Y.; Wei, F.; Liu, S.; Au, N.; Cui, W.; Zhou, H.; Qu, H. (2010) [79]	visualization of hotel customer feedback	IEEE TRANSACTIONS ON VISUALIZATION AND COMPUTER GRAPHICS 16(6), 1109-1118	Reviews of Hong Kong hotels from TripAdvisor
	Dynamical mining of ever-		
Sun, H.; Guo, W.; Shao, H.; Rong, B. (2020) [80]	changing user requirements: A product design and improvement perspective	ADVANCED ENGINEERING INFORMATICS 46, 101174	Online reviews of Trumpchi gs4 and gs8
	Extracting user requirements		
Anh, K.Q.; Nagai, Y.; Le Minh, N. (2019) [81]	from online reviews for product design: A supportive framework for designers	JOURNAL OF INTELLIGENT & FUZZY SYSTEMS 37(6), 7441-7451	Nokia phones reviews
	A text analytics-based		
Nam, S.; Lee, H.C. (2019) [82]	importance performance analysis and its application to airline service	SUSTAINABILITY 11(21), 6153	Reviews of airline services from TripAdvisor
	Identifying service		
Nam, S.; Yoon, S.; Raghavan, N.; Park, H. (2021) [83]	opportunities based on	SUSTAINABILITY 13(1), 391	Online hotel reviews from TripAdvisor

	outcome-driven innovation framework and deep learning: A case study of hotel service		
Hong, W.; Zheng, C.; Wu, L.; Pu, X. (2019) [84]	Analyzing the relationship between consumer satisfaction and fresh E- commerce logistics service using text mining techniques	SUSTAINABILITY 11(13), 3570	Reviews of logistics from JD fresh supermarket
Malik, H.; Afthanorhan, A.; Amirah, N.A.; Fatema, N. (2021) [85]	Machine learning approach for targeting and recommending a product for project management	MATHEMATICS 9(16), 1958	Reviews of cellphones on Amazon
Trappey, A.J.C.; Trappey, C.V.; Fan, C.Y.; Lee, I.J. (2018) [86]	Consumer driven product technology function deployment using social media and patent mining	ADVANCED ENGINEERING INFORMATICS 36, 120-129	Reviews of three smartphones from Amazon
Fang, Z.; Zhang, Q.; Tang, X.; Wang, A.; Baron, C. (2020) [87]	An implicit opinion analysis model based on feature-based implicit opinion patterns	ARTIFICIAL INTELLIGENCE REVIEW 53(6), 4547-4574	Online car reviews from PCauto.com.cn
Sankar, H.; Subramaniaswamy, V.; Vijayakumar, V.; Arun Kumar, S.; Logesh, R.; Umamakeswari, A.J.S.P. (2020) [88]	Intelligent sentiment analysis approach using edge computing-based deep learning technique	SOFTWARE-PRACTICE & EXPERIENCE 50(5), 645-657	Reviews from Internet movie database, rotten tomatoes data set and polarity data set
Shah, A.M.; Yan, X.; Tariq, S.; Ali, M. (2021) [89]	What patients like or dislike in physicians: Analyzing drivers of patient satisfaction and dissatisfaction using a digital topic modeling approach	INFORMATION PROCESSING & MANAGEMENT 58(3), 102516	Reviews from iwantgreatcare.org
Gregoriades, A.; Pampaka, M.; Herodotou, H.; Christodoulou, E. (2021) [90]	Supporting digital content marketing and messaging through topic modelling and decision trees	EXPERT SYSTEMS WITH APPLICATIONS 184, 115546	Online cyprus reviews from TripAdvisor
Wang, W.M.; Li, Z.; Tian, Z.G.; Tsui, E. (2018) [91]	Mining of affective responses and affective intentions of products from unstructured text	JOURNAL OF ENGINEERING DESIGN 29(7), 404-429	Reviews on 24 different product categories from Amazon

References

- McKinsey China auto consumer insights 2019. Available online: <https://www.mckinsey.com/~media/mckinsey/industries/automotive%20and%20assembly/our%20insights/china%20auto%20consumer%20insights%202019/mckinsey-china-auto-consumer-insights-2019.pdf>
- Press conference of the State Council's joint prevention and control mechanism on April 9. Available online: <http://www.gov.cn/xinwen/gwylflkjz86/index.htm>
- Economic performance of automobile industry in July 2021. Available online: http://www.caam.org.cn/chn/4/cate_154/con_5234340.html
- Insight report of Chinese car users' age. Available online: <https://max.book118.com/html/2019/0805/8021023136002040.shtm>
- Tirunillai, S.; Tellis, G.J. Does chatter really matter? Dynamics of user-generated content and stock performance. *Marketing Science* **2012**, *31*, 198-215. [CrossRef]
- Timoshenko, A.; Hauser, J.R. Identifying customer needs from user-generated content. *Marketing Science* **2019**, *38*, 1-20. [CrossRef]
- Liu, B.Q.; Karahanna, E. The dark side of reviews: the swaying effects of online product reviews on attribute preference construction. *MIS Quarterly* **2017**, *41*, 427-448. [CrossRef]
- Duan, W.; Gu, B.; Whinston, A.B. Do online reviews matter? — An empirical investigation of panel data. *Decision Support Systems* **2008**, *45*, 1007-1016. [CrossRef]
- Archak, N.; Ghose, A.; Ipeirotis, P.G. Deriving the pricing power of product features by mining consumer reviews. *Management Science* **2011**, *57*, 1485-1509. [CrossRef]
- McDonald, M. H. B.; de Chernatony, L.; Harris, F. Corporate marketing and service brands—Moving beyond the fast moving consumer goods model. *European Journal of Marketing* **2001**, *35*, 335-352. [CrossRef]
- Luo, H.; Cheng, S.; Zhou, W.; Song, W.; Yu, S.; Lin, X. Research on the Impact of Online Promotions on Consumers' Impulsive Online Shopping Intentions. *J. Theor. Appl. Electron. Commer. Res.* **2021**, *16*, 2386-2404. [CrossRef]
- Bracewell, D.B.; Minato, J.; Ren, F.; Kuroiwa, S. Determining the emotion of news articles. In Proceedings of the International Conference on Intelligent Computing, Kunming, China, 16-19 August 2006; pp. 918-923. [CrossRef]
- Rao, Y.; Lei, J.; Liu, W.; Li, Q.; Chen, M. Building emotional dictionary for sentiment analysis of online news. *World Wide Web* **2014**, *17*, 723-742. [CrossRef]
- Ji, Q.; Raney, A.A. Developing and validating the self-transcendent emotion dictionary for text analysis. *PloS one* **2020**, *15*, e0239050. [CrossRef]
- Liu, T.; Zhang, C. Wu, M. Product feature extraction algorithm based on boundary average information entropy in online reviews. *Systems Engineering — Theory&Practice* **2016**, *36*, 2416-2423. [CrossRef]
- Jing, R.; Yuan, C.; Rezaei, H.; Qian, J.; Zhang, Z. Assessments on emergy and greenhouse gas emissions of internal combustion engine automobiles and electric automobiles in the USA. *Journal of Environmental Sciences* **2020**, *90*, 297-309. [CrossRef]
- Du, Z.; Lin, B. Changes in automobile energy consumption during urbanization: Evidence from 279 cities in China. *Energy Policy* **2019**, *132*, 309-317. [CrossRef]
- Tong, R.; Cheng, M.; Ma, X.; Yang, Y.; Liu, Y.; Li, J. Quantitative health risk assessment of inhalation exposure to automobile foundry dust. *Environmental Geochemistry and Health* **2019**, *41*, 2179-2193. [CrossRef]
- Bao, G.Z.; Liu, W.; Wei, L.; Zhao, J.G. Automobile brake protection based on laser pulse real-time ranging. *Lasers in Engineering (Old City Publishing)* **2020**, *45*, 353-365.
- James, A.T.; Kumar, G.; Arora, A.; Padhi, S. Development of a design based remanufacturability index for automobile systems. *Journal of Automobile Engineering* **2021**, *235*, 3138-3156. [CrossRef]
- Zhang, X.; Lou, Z.; Sun, Z.; Dai, X. Pricing and investment decision issues of an automobile manufacturer for different types of vehicles. *IEEE Access* **2021**, *9*, 73083-73089. [CrossRef]
- Griffin, A., & Hauser, J. R. The Voice of the Customer. *Marketing Science* **1993**, *12*, 1-27. [CrossRef]
- Hu, N.; Zhang, T.; Gao, B.; Bose, I. What do hotel customers complain about? Text analysis using structural topic model. *Tourism Management* **2019**, *72*, 417-426. [CrossRef]
- Kitsios, F.; Kamariotou, M.; Karanikolas, P.; Grigoroudis, E. Digital marketing platforms and customer satisfaction: Identifying eWOM using big data and text mining. *Applied Sciences* **2021**, *11*, 8032. [CrossRef]
- Xiang, Z.; Schwartz, Z.; Gerdes Jr, J.H.; Uysal, M. What can big data and text analytics tell us about hotel guest experience and satisfaction? *International Journal of Hospitality Management* **2015**, *44*, 120-130. [CrossRef]
- Xu, X.; Li, Y. The antecedents of customer satisfaction and dissatisfaction toward various types of hotels: A text mining approach. *International journal of hospitality management* **2016**, *55*, 57-69. [CrossRef]
- Jiao, J.; Chen, C.H. Customer requirement management in product development: A review of research issues. *Concurrent Engineering* **2006**, *14*, 173-185. [CrossRef]
- Min Kim, J.; Han, J.; Jun, M. Differences in mobile and nonmobile reviews: The role of perceived costs in review-posting. *International Journal of Electronic Commerce* **2020**, *24*, 450-473. [CrossRef]

29. Shamantha, R.B.; Shetty, S.M.; Rai, P. Sentiment analysis using machine learning classifiers: Evaluation of performance. In Proceedings of the 2019 IEEE 4th International Conference on Computer and Communication Systems (ICCCS), Singapore, 23-25 February 2019; pp. 21-25. [CrossRef]
30. Wijayanti, R.; Arisal, A. Automatic Indonesian sentiment lexicon curation with sentiment valence tuning for social media sentiment analysis. *ACM Transactions on Asian and Low-Resource Language Information Processing* **2021**, *20*, 1-16. [CrossRef]
31. Sahu, T.P.; Khandekar, S. A machine learning-based lexicon approach for sentiment analysis. *Journal of Technology and Human Interaction (IJTHI)* **2020**, *16*, 8-22. [CrossRef]
32. Xu, X.; Li, Y. The antecedents of customer satisfaction and dissatisfaction toward various types of hotels: A text mining approach. *International Journal of Hospitality Management* **2016**, *55*, 57-69. [CrossRef]
33. Mankad, S.; Han, H.S.; Goh, J.; Gavirneni, S. Understanding online hotel reviews through automated text analysis. *Service Science* **2016**, *8*, 124-138. [CrossRef]
34. Ma, J.; Hou, Y.; Wang, Z.; Yang, W. Pricing strategy and coordination of automobile manufacturers based on government intervention and carbon emission reduction. *Energy Policy* **2021**, *148*, 111919. [CrossRef]
35. Rasool, G.; Pathania, A. Reading between the lines: untwining online user-generated content using sentiment analysis. *Journal of Research in Interactive Marketing* **2021**, *15*, 401-418. [CrossRef]
36. Fels, A.; Briele, K.; Ellerich, M.; Schmitt, R. Extracting customer-related information for need identification. In Proceedings of the International Conference on Human Systems Engineering and Design: Future Trends and Applications, Reims, France, 25-27 October 2018; pp. 1108-1112. [CrossRef]
37. Kühl, N.; Mühlthaler, M.; Goutier, M. Supporting customer-oriented marketing with artificial intelligence: automatically quantifying customer needs from social media. *Electronic markets* **2020**, *30*, 351-367. [CrossRef]
38. Lee, J.Y.H.; Yang, C.S.; Chen, S.Y. Understanding customer opinions from online discussion forums: A design science framework. *Engineering Management Journal* **2017**, *29*, 235-243. [CrossRef]
39. Vollero, A.; Sardanelli, D.; Siano, A. Exploring the role of the Amazon effect on customer expectations: An analysis of user-generated content in consumer electronics retailing. *Journal of Consumer Behaviour* **2021**, 1-12. [CrossRef]
40. Zhu, D.; Lappas, T.; Zhang, J. Unsupervised tip-mining from customer reviews. *Decision Support Systems* **2018**, *107*, 116-124. [CrossRef]
41. Ekhlassi, A.; Zahedi, A. A unique method of constructing brand perceptual maps by the text mining of multimedia consumer reviews. *International Journal of Mobile Computing and Multimedia Communications (IJMCMC)* **2018**, *9*, 1-22. [CrossRef]
42. Yu, C.E.; Zhang, X. The embedded feelings in local gastronomy: a sentiment analysis of online reviews. *Journal of Hospitality and Tourism Technology* **2020**, *11*, 461-478. [CrossRef]
43. Hsiao, Y.H.; Chen, M.C.; Liao, W.C. Logistics service design for cross-border E-commerce using Kansei engineering with text-mining-based online content analysis. *Telematics and Informatics* **2017**, *34*, 284-302. [CrossRef]
44. Zhang, N.; Zhang, R.; Pang, Z.; Liu, X.; Zhao, W. Mining express service innovation opportunity from online reviews. *Journal of Organizational and End User Computing (JOEUC)* **2021**, *33*, 1-15. [CrossRef]
45. Valsan, A.; Sreepriya, C.T.; Nitha, L. Social media sentiment polarity analysis: A novel approach to promote business performance and consumer decision-making. In *Artificial Intelligence and Evolutionary Computations in Engineering Systems*; Dash, S. S.; Vijayakumar, K.; Panigrahi, B. K.; Das, S.; Springer: Singapore, 2017; pp. 1-12, ISBN 978-981-10-3173-1. [CrossRef]
46. Hasan, M.R.; Abdunurova, A.; Wang, W.; Zheng, J.; Shams, S.R. Using deep learning to investigate digital behavior in culinary tourism. *Journal of Place Management and Development* **2021**, *14*, 43-65. [CrossRef]
47. Kauffmann, E.; Gil, D.; Peral, J.; Ferrández, A.; Sellers, R. A step further in sentiment analysis application in marketing decision-making. In Proceedings of the International Research & Innovation Forum, Rome, Italy, 24-26 April 2019; pp. 211-221. [CrossRef]
48. Vinodhini, G.; Chandrasekaran R.M. Measuring the quality of hybrid opinion mining model for e-commerce application. *Measurement* **2014**, *55*, 101-109. [CrossRef]
49. Chalupa, S.; Petricek, M.; Chadt, K. Improving service quality using text mining and sentiment analysis of online reviews. *Quality-Access to Success* **2021**, *22*, 46-49.
50. Dickinger, A.; Mazanec, J.A. Significant word items in hotel guest reviews: A feature extraction approach. *Tourism Recreation Research* **2015**, *40*, 353-363. [CrossRef]
51. Asghar, Z.; Ali, T.; Ahmad, I.; Tharanidharan, S.; Nazar, S. K. A.; Kamal, S. Sentiment analysis on automobile brands using Twitter data. In Proceedings of the International Conference on Intelligent Technologies and Applications, Bahawalpur, Pakistan, 23-25 October 2018; pp. 76-85. [CrossRef]
52. Aman, J.J.C.; Smith-Colin, J.; Zhang, W. Listen to E-scooter riders: Mining rider satisfaction factors from app store reviews. *Transportation Research Part D: Transport and Environment* **2021**, *95*, 102856. [CrossRef]
53. Ng, C.Y.; Law, K.M.Y. Investigating consumer preferences on product designs by analyzing opinions from social networks using evidential reasoning. *Computers & Industrial Engineering* **2020**, *139*, 106180. [CrossRef]

54. Becken, S.; Alaei, A.R.; Wang, Y. Benefits and pitfalls of using tweets to assess destination sentiment. *Journal of Hospitality and Tourism Technology* **2019**, *11*, 19-34. [CrossRef]
55. Wang, W.; Feng, Y.; Dai, W. Topic analysis of online reviews for two competitive products using latent Dirichlet allocation. *Electronic Commerce Research and Applications* **2018**, *29*, 142-156. [CrossRef]
56. Zhu, D.; Lappas, T.; Zhang, J. Unsupervised tip-mining from customer reviews. *Decision Support Systems* **2018**, *107*, 116-124. [CrossRef]
57. Al-Obeidat, F.; Spencer, B.; Kafeza, E. The opinion management framework: Identifying and addressing customer concerns extracted from online product reviews. *Electronic Commerce Research and Applications* **2018**, *27*, 52-64. [CrossRef]
58. Vo, A.D.; Nguyen, Q.P.; Ock, C.Y. Opinion–aspect relations in cognizing customer feelings via reviews. *IEEE Access* **2018**, *6*, 5415-5426. [CrossRef]
59. Oh, Y.K.; Yi, J. Asymmetric effect of feature level sentiment on product rating: an application of bigram natural language processing (NLP) analysis. Internet Research, 2021. [CrossRef]
60. Singh, A.; Tucker, C.S. A machine learning approach to product review disambiguation based on function, form and behavior classification. *Decision Support Systems* **2017**, *97*, 81-91. [CrossRef]
61. Eldin, S.S.; Mohammed, A.; Hefny, H.; Ahmed, A.S.E. An enhanced opinion retrieval approach on Arabic text for customer requirements expansion. *Journal of King Saud University-Computer and Information Sciences* **2019**, *33*, 351-363. [CrossRef]
62. Riaz, S.; Fatima, M.; Kamran, M.; Nisar, N.W. Opinion mining on large scale data using sentiment analysis and k-means clustering. *Cluster Computing* **2019**, *22*, 7149-7164. [CrossRef]
63. Jin, J.; Ji, P.; Liu, Y.; Lim, S.J. Translating online customer opinions into engineering characteristics in QFD: A probabilistic language analysis approach. *Engineering Applications of Artificial Intelligence* **2015**, *41*, 115-127. [CrossRef]
64. Jin, J.; Jia, D.; Chen, K. Mining online reviews with a Kansei-integrated Kano model for innovative product design. *International Journal of Production Research* **2021**: 1-20. [CrossRef]
65. Zhang, L.; Chu, X.; Xue, D. Identification of the to-be-improved product features based on online reviews for product redesign. *International Journal of Production Research* **2019**, *57*, 2464-2479. [CrossRef]
66. Zhou, F.; Jiao, R.J.; Linsey, J.S. Latent customer needs elicitation by use case analogical reasoning from sentiment analysis of online product reviews. *Journal of Mechanical Design* **2015**, *137*, 071401. [CrossRef]
67. Zhou, F.; Ayoub, J.; Xu, Q.; Jessie Yang, X. A machine learning approach to customer needs analysis for product ecosystems. *Journal of Mechanical Design* **2020**, *142*, 011101. [CrossRef]
68. Liu, Y.; Jin, J.; Ji, P.; Harding, J.A.; Fung, R.Y. Identifying helpful online reviews: a product designer's perspective. *Computer-Aided Design* **2013**, *45*, 180-194. [CrossRef]
69. Ireland, R.; Liu, A. Application of data analytics for product design: Sentiment analysis of online product reviews. *CIRP Journal of Manufacturing Science and Technology* **2018**, *23*, 128-144. [CrossRef]
70. Joung, J.; Jung, K.; Ko, S.; Kim, K. Customer complaints analysis using text mining and outcome-driven innovation method for market-oriented product development. *Sustainability* **2019**, *11*, 40. [CrossRef]
71. Jain, P.K.; Pamula, R.; Srivastava, G. A systematic literature review on machine learning applications for consumer sentiment analysis using online reviews. *Computer Science Review* **2021**, *41*, 100413. [CrossRef]
72. Foris, D.; Crihalmean, N.; Foris, T. Exploring the environmental practices in hospitality through booking websites and online tourist reviews. *Sustainability* **2020**, *12*, 10282. [CrossRef]
73. Han, Y.; Moghaddam, M. Eliciting attribute-level user needs from online reviews with deep language models and information extraction. *Journal of Mechanical Design* **2021**, *143*, 061403. [CrossRef]
74. Wu, J.; Wang, Y.; Zhang, R.; Cai, J. An approach to discovering product/service improvement priorities: Using dynamic importance-performance analysis. *Sustainability* **2018**, *10*, 3564. [CrossRef]
75. Htay, S.S.; Lynn, K.T. Extracting product features and opinion words using pattern knowledge in customer reviews. *The Scientific World Journal* **2013**, *2013*, 1-5. [CrossRef]
76. Wang, W.M.; Tian, Z.G.; Li, Z.; Wang, J.W. Supporting the construction of affective product taxonomies from online customer reviews: an affective-semantic approach. *Journal of Engineering Design* **2019**, *30*, 445-476. [CrossRef]
77. Zhou, F.; Jiao, J.R.; Yang, X.J.; Lei, B. Augmenting feature model through customer preference mining by hybrid sentiment analysis. *Expert Systems with Applications* **2017**, *89*, 306-317. [CrossRef]
78. Jin, J.; Ji, P.; Kwong, C.K. What makes consumers unsatisfied with your products: Review analysis at a fine-grained level. *Engineering Applications of Artificial Intelligence* **2016**, *47*, 38-48. [CrossRef]
79. Wu, Y.; Wei, F.; Liu, S.; Au, N.; Cui, W.; Zhou, H.; Qu, H. OpinionSeer: Interactive visualization of hotel customer feedback. *IEEE transactions on visualization and computer graphics* **2010**, *16*, 1109-1118. [CrossRef]
80. Sun, H.; Guo, W.; Shao, H.; Rong, B. Dynamical mining of ever-changing user requirements: A product design and improvement perspective. *Advanced Engineering Informatics* **2020**, *46*, 101174. [CrossRef]
81. Anh, K.Q.; Nagai, Y.; Le Minh, N. Extracting user requirements from online reviews for product design: a supportive framework for designers. *Journal of Intelligent & Fuzzy Systems* **2019**, *37*, 7441-7451. [CrossRef]

82. Nam, S.; Lee, H.C. A text analytics-based importance performance analysis and its application to airline service. *Sustainability* **2019**, *11*, 6153. [CrossRef]
83. Nam, S.; Yoon, S.; Raghavan, N.; Park, H. Identifying service opportunities based on outcome-driven innovation framework and deep learning: A case study of hotel service. *Sustainability* **2021**, *13*, 391. [CrossRef]
84. Hong, W.; Zheng, C.; Wu, L.; Pu, X. Analyzing the relationship between consumer satisfaction and fresh e-commerce logistics service using text mining techniques. *Sustainability* **2019**, *11*, 3570. [CrossRef]
85. Malik, H.; Afthanorhan, A.; Amirah, N.A.; Fatema, N. Machine learning approach for targeting and recommending a product for project management. *Mathematics* **2021**, *9*, 1958. [CrossRef]
86. Trappey, A.J.C.; Trappey, C.V.; Fan, C.Y.; Lee, I.J. Consumer driven product technology function deployment using social media and patent mining. *Advanced Engineering Informatics* **2018**, *36*, 120-129. [CrossRef]
87. Fang, Z.; Zhang, Q.; Tang, X.; Wang, A.; Baron, C. An implicit opinion analysis model based on feature-based implicit opinion patterns. *Artificial Intelligence Review* **2020**, *53*, 4547-4574. [CrossRef]
88. Sankar, H.; Subramaniaswamy, V.; Vijayakumar, V.; Arun Kumar, S.; Logesh, R.; Umamakeswari, A.J.S.P. Intelligent sentiment analysis approach using edge computing-based deep learning technique. *Software: Practice and Experience* **2020**, *50*, 645-657. [CrossRef]
89. Shah, A.M.; Yan, X.; Tariq, S.; Ali, M. What patients like or dislike in physicians: Analyzing drivers of patient satisfaction and dissatisfaction using a digital topic modeling approach. *Information Processing & Management* **2021**, *58*, 102516. [CrossRef]
90. Gregoriades, A.; Pampaka, M.; Herodotou, H.; Christodoulou, E. Supporting digital content marketing and messaging through topic modelling and decision trees. *Expert Systems with Applications* **2021**, *184*, 115546. [CrossRef]
91. Wang, W.M.; Li, Z.; Liu, L.; Tian, Z.G.; Tsui, E. Mining of affective responses and affective intentions of products from unstructured text. *Journal of Engineering Design* **2018**, *29*, 404-429. [CrossRef]
92. Martilla, J. A.; James, J. C. Importance-Performance Analysis. *Journal of Marketing* **1977**, *41*, 77-79. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.