https://doi.org/10.1007/s43684-022-00032-4

Autonomous Intelligent Systems

ORIGINAL ARTICLE Open Access

A dynamic customer requirement mining method for continuous product improvement

Qian Zhao[1,2,3], Wu Zhao[1,2,3*], Xin Guo[1,2,3], Kai Zhang[1,2,3] and Miao Yu[1,2,3]

Abstract

The key to successful product development is better understanding of customer requirements and efficiently identifying the product attributes. In recent years, a growing number of researchers have studied the mining of customer requirements and preferences from online reviews. However, since customer requirements often change dynamically on multi-generation products, most existing studies failed to discover the correlations between customer satisfaction and continuous product improvement. In this work, we propose a novel dynamic customer requirement mining method to analyze the dynamic changes of customer satisfaction of product attributes based on sentiment and attention expressed in online reviews, aiming to better meet customer requirements and provide the direction and content of future product improvement. Specifically, this method is divided into three parts. Firstly, text mining is adopted to collect online review data of multi-generation products and identify product attributes. Secondly, the attention and sentiment scores of product attributes are calculated with a natural language processing tool, and further integrated into the corresponding satisfaction scores. Finally, the improvement direction for next-generation products is determined based on the changing satisfaction scores of multi-generation product attributes. In addition, a case study on multi-generation phone products based on online reviews was conducted to illustrate the effectiveness and practicality of the proposed methodology. Our research completes the field of requirements analysis and provides a new dynamic approach to requirements analysis for continuous improvement of multi-generation products, which can help enterprises to accurately understand customer requirements and improve the effectiveness and efficiency of continuous product improvement.

Keywords: Requirement mining, Sentiment analysis, Continuous product improvement, Multi-generation products, Online reviews

1 Introduction

In the beginning stages of product design, customer requirements are critical in determining the design direction of a new product [1, 2]. The key to successful product development is better understanding of customer requirements and efficiently identifying the product attributes [3, 4]. In the product design process, accurate customer requirements not only guide designers in determining the functionality and structure of the product [5–7], but also improve customer satisfaction by improving product attributes, thereby increasing the competitiveness of the product in the marketplace.

It is worth observing that with the widespread availability of information technology, the Internet has permeated various industries, and its openness and freedom have resulted in the growing number of customers using online reviews to express personal consumer experiences [8]. Ac-

*Correspondence: zhaowu@scu.edu.cn
[1] School of Mechanical Engineering, Sichuan University, Chengdu, 610065, China
[2] The Innovation Method and Creative Design Key Laboratory of Sichuan Province, Sichuan University, Chengdu, 610065, China
Full list of author information is available at the end of the article

cording to a recent report, 77% of customers read online reviews and 75% of them trust online reviews more than personal recommendations [9]. Numerous studies have shown that most customer behaviors are directly related to online reviews [10, 11], especially in purchase decisions [12, 13] and product sales [14, 15], and Güneş [16] suggested that online reviews can provide designers with heuristic ideas, which suggests that online reviews contain a large amount of valuable content and also provide the source of data for customer requirements analysis. Cao et al. [17] suggested that to unravel the complex challenges addressed by product innovation, it is oftentimes essential for customers to participate in the design process. Online reviews are also a form of customer involvement in the design process. Compared with traditional customer requirements data, online reviews are more efficient and lower cost to acquire. In recent years, a growing number of researchers have studied the mining of customer requirements from online reviews. The main focus is to analyze online review data in terms of extracting customer preferences [18, 19], identifying product attributes [4, 20], and developing strategies for companies and sales [21–23]. However, in the process of continuous improvement of multi-generation products, new products are usually developed by improving existing product. Since customer requirements often change dynamically on multi-generation products, most existing studies failed to discover the correlations between customer satisfaction and continuous product improvement.

To address the issue, in this paper a dynamic customer requirement mining method for continuous product improvement is proposed. Two unique characteristics are described, 1) dynamic customer requirement mining, which examines the change of the satisfaction score of product attributes over multi-generation products; 2) applying dynamic change trends to the method, dynamic changes trend is mostly applied in the fields of climate [24, 25] and disease [26–28], mainly by predicting the future based on existing data so that appropriate countermeasures and suggestions can be prepared in advance. This study combines these two characteristics in the process of multi-generation product improvement to help companies accurately understand customer requirements, quickly develop improvement strategies, and win market leadership.

The study is organized as follows: A review of related studies present in Sect. 2. Section 3 proposes a framework that the dynamic customer requirement mining method for continuous product improvement and introduces the methods of data process and dynamic change trend of product attributes, which includes the extraction of product attributes, the calculation of satisfaction score for product attributes. Section 4 mainly describes the application of the proposed framework and methodology in the case study of product. The conclusion is presented in Sect. 5.

2 Related work

2.1 Analysis of customer requirements

Customer requirements are an important part of product innovation [29, 30]. Traditional methods of customer requirement analysis include customer surveys, customer interviews, focus groups, questionnaires, etc., but they are usually time-consuming and subjective. Compared to traditional customer requirement analysis data, online review data is more accurate, real, reliable and easily accessible [31].

Indeed, there are many researchers who conduct customer requirement analysis based on online product reviews. Timoshenko and Hauser [32] proposed a customer requirement analysis method based on user-generated content using machine learning, which proved its value and efficiency compared to traditional content extraction methods. However, the method has some limitations in terms of the scope of application. For some specialized equipment for oil exploration, it is not applicable because the number of customers for such products is small, and such customers may not blog, tweet, or post reviews. Anh et al. [33] presented how to extract useful comments from customers and the implications for the early stages of product design and proposed a framework for analyzing comments on online shopping sites and it can automatically extract useful information from review documents and it can also collaboratively work between designers and opinion customers. Hananto et al. [34] developed a fashion style model to investigate fashion styles by using online customer reviews to analyze customer preferences. The obtained fashion style models can potentially help marketing and product design specialists better understand customer preferences in the e-commerce fashion industry. Chan et al. [35] have formulated a multi-objective optimization problem which attempts to equally learn imbalances online review data collected from popular and unpopular products in marketplace and addressed the problem of how to predict customer satisfaction with product design attributes through online data imbalance in order to improve the success of the product in the marketplace. Joung and Kim [20] proposed to identify product attributes from online reviews using an approach based on latent Dirichlet assignment and the case study of Android smartphones demonstrates that the proposed method yields better LDA results to identify product attributes than previous methods. Ying et al. [36] performed fine-grained sentiment analysis based on e-commerce product review data, which can provide many valuable information to merchants. Similarly, Zhang et al. [37] proposed a method based on fine-grained sentiment analysis and a Kano model to extract consumers' requirements for product attributes from online reviews. The proposed method is illustrated by online reviews of air purifiers and they also use a dataset of refrigerators to further validate the effectiveness of the method. Li et al. [38]

proposed a model for identifying key customer requirements based on online product reviews and applied the key customer identification problem for smartphones to demonstrate the feasibility and effectiveness of the model.

In order to mine customer requirements from online reviews, most current research has focused on methods to mine customer requirements, however, lack research on the dynamic changes of customer requirements over multi-generation products. In this paper, a dynamic customer requirement mining method for continuous product improvement is proposed to address these issues.

2.2 Feature extraction from online reviews

Product features are attributes of products (which also become entities or topics), product components, or product parts that are viewed online by customers [39]. Similarly, Han and Moghaddam [40] stated that product features can refer to any component or feature of a product or service. In previous studies, scholars have proposed many methods to extract product features. The first one is extraction based on linguistic rules, mainly by sentence and word syntax, and this method is designed to analyze the syntactic relationships or dependencies between feature words and viewpoint words [41]. The second one is extraction by supervised learning model, and Budhi et al. [42] proposed a comparison framework that combines text pre-processing and feature extraction methods. The feature extraction stage is to use TF to generate features for each pre-processed word token first, and then features for each review text are extracted by checking for the existence of each feature word. The third one is extraction using topic model, Wu et al. [43] proposed an LDA short text clustering algorithm (SKP-LDA) based on sentiment word co-occurrence and knowledge pair feature extraction, and comparing with JST, LSM, LTM and ELDA, SKP-LDA performs better in terms of Accuracy, Precision, Recall and F-measure. The experiments demonstrate that SKP-LDA has better semantic analysis ability and emotional topic clustering effect. The fourth one is based on noun or noun phrase extraction, Hu and Liu [44] extracted product features based on the frequency of nouns and noun phrases being volume in reviews, and kept product features with frequencies greater than the threshold. Product features are further divided into explicit features, which are features directly mentioned by customers when reviewing products, and implicit features, which are features that are not directly mentioned but can be inferred from the reviews [45]. For example, for phones, the battery is one of the product attributes, and if customers directly comment that the battery is durable, it is an explicit feature, but some customers comment that the phone has a long standby time, which is called an implicit feature, because the battery can also be inferred from this product attribute.

Based on the overview of feature extraction methods, this paper uses extraction methods based on nouns or noun phrases, also known as frequent pattern extraction. The advantage of this approach is that when there are a large number of customer reviews for products, the targets of the reviews are usually focused on the product features that customers are more concerned about. As the same time, this paper also takes into account that online product reviews contain implicit features, we can add new words in the jieba library by adding words that can represent the product features in the dictionary, which can then achieve as many product features as possible.

2.3 Sentiment analysis from online reviews

Sentiment analysis refers to the use of relevant textual methods to transform sentiment words in reviews into computable sentiment values, or to analyze the sentiment expressed by consumers in texts by mining their emotions based on sentiment words [46]. Online product review sentiment analysis can be divided into coarse-grained and fine-grained sentiment analysis. Coarse-grained sentiment analysis refers to the tendency analysis of the text, which is mainly to determine whether the customer sentiment toward the product is positive, negative or neutral; fine-grained sentiment analysis is mainly to extract the sentiment object and sentiment words of the text, and then determine the customer sentiment tendency [47]. Research on sentiment analysis is divided into the following two main types. The first one is lexicon and rule-based sentiment analysis, which mainly uses sentiment lexicon for sentiment calculation, and Asghar et al. [48] also integrated emoticons, negation words, sentiment words and domain terms for sentiment analysis based on lexicon and rules. The advantage of this approach is that it has a high accuracy rate, but the disadvantage is also obvious that it needs to build different lexicon and rules for different domains, which is poor in generality. The second type of sentiment analysis is based on machine learning and deep learning, and the general idea is to train the model through sentiment labeled data and then test the model using a test set. Rathor et al. [49] focused on examining the efficiency of three machine learning techniques(Support Vector Machines(SVM), Naïve Bayes(NB) and Maximum Entropy(ME)) for classification of online reviews using a web model using supervised learning methods, and the results have shown that SVM has resulted maximum accuracy. Similarly, Wawre and Deshmukh [50] compared two supervised machine learning approaches SVM, NB for sentiment classification of reviews. Results states that Naïve Bayes approach outperformed the SVM. Zeng et al. [51] put forward the PosATT-LSTM model and extensive experiments have been conducted on the SemEval 2014 datasets, and the results indicate the efficacy of the proposed model. Feng et al. [52] proposed a method of combining the convolutional neural network and attention model for sentiment analysis and the experimental results show that the method is more effective than the machine learning method and the simplex convolution neural network method. The results of this sentiment analysis method are strongly influenced by the training set. For sentiment analysis most researchers currently still use a lexicon and rule-based approach by extracting feature words and sentiment words in reviews, and then assigning sentiment words based on sentiment lexicon to get customer sentiment scores on product features.

Due to the increasing number of studies on natural language processing tasks, especially in the context of sentiment analysis of texts, more and more researchers are focusing on the second approach. Therefore there are also many researchers who have developed sentiment analysis tools. In this paper, we use the SnowNLP, which has a high accuracy rate for sentiment analysis of shopping-type reviews because its corpus is mainly on shopping.

3 Methodology

Previous research on customer requirements analysis of online reviews has not taken into account the dynamics changes of customer satisfaction with product attributes over multi-generation products and how to analyze this dynamic change for continuous product improvement. In order to address these two issues, this study develops a dynamic customer requirement mining method for continuous product improvement containing three stages: identifying product attributes, calculating the satisfaction score of product attributes and determining the direction of continuous product improvement. First, online customer reviews from multi-generation products as resource and the product attributes are extracted after data pre-processing and data processing. Second, the attention and sentiment scores of product attributes are calculated with a natural language processing tool, and further integrated into the corresponding satisfaction scores. Finally, in order to guide the direction for improving next generation product, the evolution of product attributes in terms of satisfaction scores over multi-generation products is determined. Our analytical framework is illustrated in Fig. 1.

3.1 Data acquisition and processing

Python is utilized to write regular expressions so as to run a web crawler and effectively obtain online review data. The general steps of web text crawling are shown in Fig. 2. Figure 3 illustrates a typical online review posted on a shopping platform. Part A is the registered named of the reviewer. Part B is the evaluation score of the reviewer for the product. Part C is the review content of reviewer. Part D is the color and memory of the product, and part E is the review time. We mainly obtain the review data in part C. Reviews can be represented as a set $R = \{R_1, R_2, \ldots, R_i, \ldots, R_m\}$, where R_i denotes the i-th review.

After obtaining the comment data, data pre-processing and data processing are performed and divided into the following five steps:

1) Remove the reviews that are irrelated to the product, such as "The customer feels good", "The customer did not fill in the evaluation content in time" and other similar reviews.

2) Splitting comments into sentence-based units, which provides a convenient way to match sentiment with product attributes at post.

3) Text deduplication and tokenizer.

4) Removing stop words. Stop words include "I", "this" and other words that are widely present in the data and have no research value, as well as some auxiliary words and prepositions such as "end", "in" and other words that have no real meaning. Usually, there is a list of stop words in the lexicon to help improve the accuracy of text analysis.

5) Part-of-speech (POS) tagging and feature extraction.

There are many existing word separation tools, such as THULAC introduced by Tsinghua University, NLPIR developed by Beijing University of Technology, Harbin Institute of Technology's LTP, jieba, and so on. Each of them has its own advantages and disadvantages. Since the review text was collected from Chinese online websites, the jieba tool was selected to process the review text through segmenting and extracting product attributes. Jieba is also currently used by a relatively large number of people. The principle of its word separation is to calculate the association probability between each Chinese word based on the existing Chinese lexicon. There are three modes of jieba word separation, namely, the most precise cut sentence and the most accurate cut sentence. The exact mode, which is the most accurate way to slice the sentences, the full mode, which is to get all the words that can become words, and the search engine mode. search engine mode. The exact mode is the most suitable for text analysis.

3.2 Extraction of product attributes

This section performs product attributes extraction using the results of POS tagging and feature extraction obtained in the previous section. Since product attributes can refer to any component or feature of a product or service [40], we only perform extraction on nouns, verbs and verbal nouns. The identification of product attributes based on TextRank method. Then perform filtering and semantic similarity analysis to determine product attributes.

TextRank [53] proposed by Rada Mihalcea and Paul Tarau in 2004, which is a text ranking algorithm, improved from Google PageRank algorithm for ranking the importance of web pages, extracts keywords, key phrases of a given text and extracts key lines of that text using an extractive automatic digest method.

The TextRank algorithm model is generally expressed in the form of $G = (V, E)$, for natural language processing V

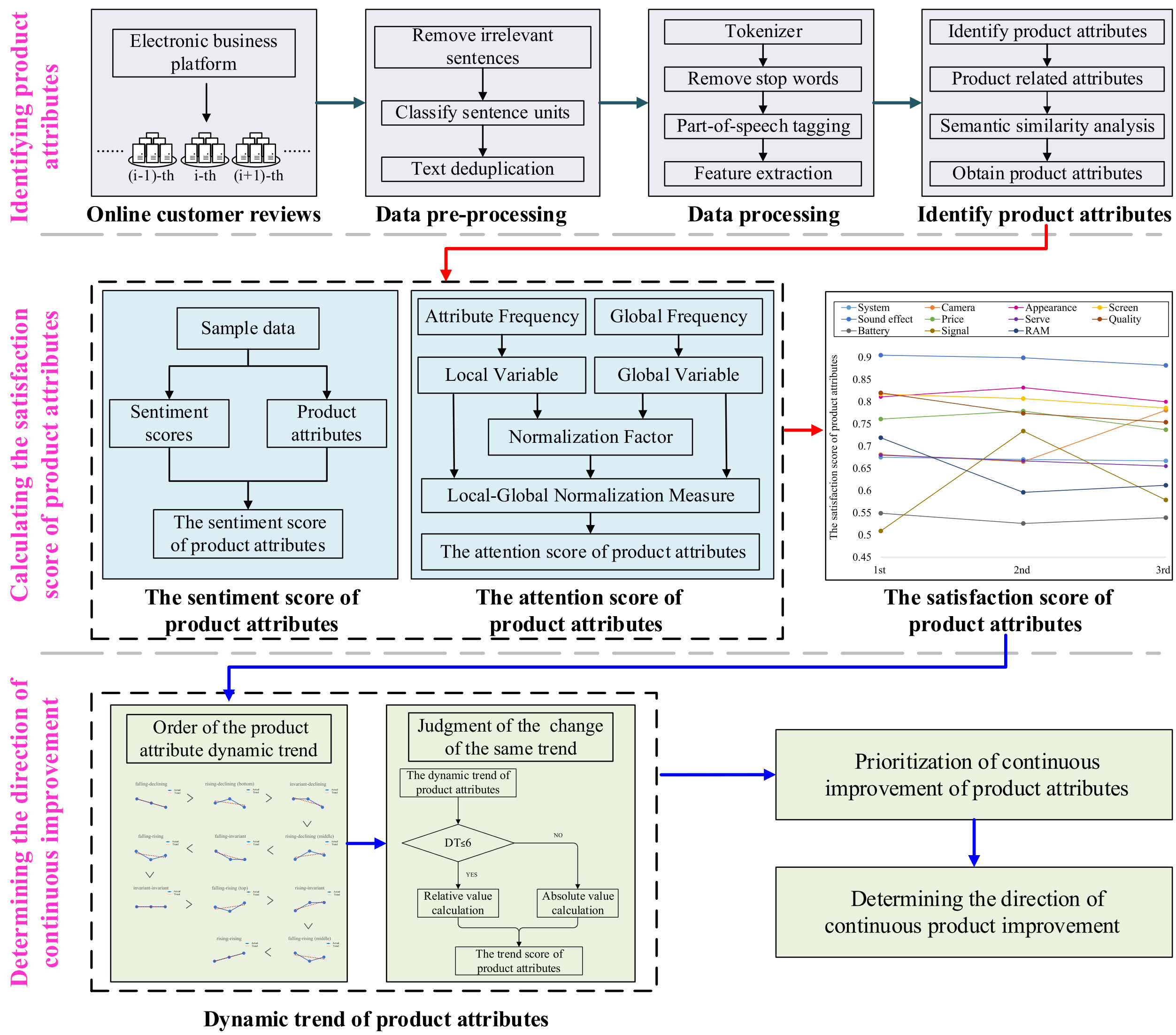


Figure 1 Proposed analytical framework

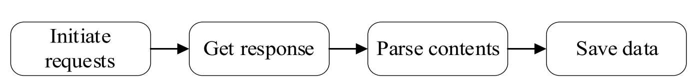


Figure 2 Crawler steps

represents the set of all word-based nodes in the graph, and E represents the set of connected edges between points in the graph. The scoring of nodes V_i is shown below:

$$WS(V_i) = (1 - d) + d * \sum_{V_j \in \ln(V_i)} \frac{w_{ji}}{\sum_{V_k \in Out(V_j)} w_{jk}} WS(V_j), \quad (1)$$

where w_{ji} represents the weight of the edge between node V_j and node V_i, usually the similarity between the two nodes represents the weight; $In(V_i)$ represents the set of all nodes pointing to node V_i, and $Out(V_i)$ represents is the set of all nodes pointed to by node V_i; d represents the damping coefficient ($0 \leq d \leq 1$), which is the probability of pointing from a specific point to any other point in the graph, and d usually takes a value of 0.85. In addition, it

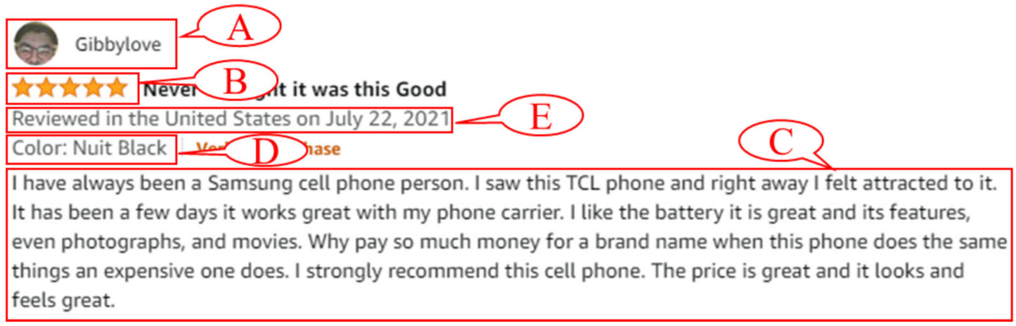


Figure 3 Online consumer reviews

should be noted that in the process of setting the initial value of the TextRank algorithm, the score of all nodes is generally initialized to 1, and the threshold for determining the convergence is 0.0001. When the error rate is less than the threshold, it will converge and stop the iterative operation.

After the product attributes extraction, we put all the product attributes into a collection, and conduct word frequency statistics, and output the top 100 product attributes after the statistics, and then filter the features that have nothing to do with the product, such as "problem", "really", etc. Then, the product attributes to be determined are removed by screening. Since there are different words in the product attributes to be determined, such as "camera" and "photo", "shape" and "appearance", although the two words are different, they represent the same product attributes, so we need to cluster such words and finally get the similarity matrix to determine the product attributes by semantic similarity calculation.

In this paper, we focus on calculating the similarity of words using word vectors [54]. To train word vectors, we need a large corpus, currently there are many English corpora, but less Chinese corpora, here we use Chinese pages from Wikipedia as training corpus, download at https://dumps.wikimedia.org/zhwiki/. After processing the corpus, each word is mapped to a continuous vector space using a word embedding algorithm. In this paper, we use Word2Vec CBOW as the word embedding algorithm [55, 56]. The goal of CBOW model training is to predict intermediate words based on the context. The architecture of CBOW model is represented in Fig. 4 and consists of an input layer, a hidden layer and an output layer. The training objective is defined as:

$$J_\theta = \frac{1}{T}\sum_{t=1}^{T} \log p(c_t|c_{t-n},\ldots,c_{t-1},c_{t+1},\ldots,c_{t+n}), \tag{2}$$

where n is the window of words around the current word c_t at each time step t. The term $p(c_t|c_{t-n},\ldots,c_{t-1},c_{t+1},\ldots,c_{t+n})$

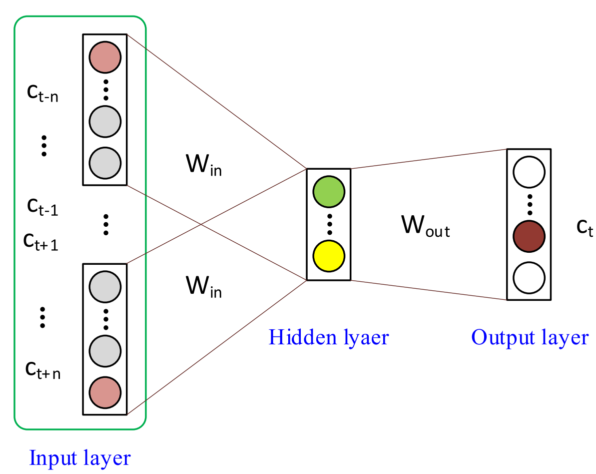


Figure 4 The architecture of CBOW model

is calculated by a softmax function. After training the word vector model, we calculate the semantic similarity between two product attributes by cosine similarity and the equation is expresses as follows:

$$sim(A,B) = \cos(\theta) = \frac{A \cdot B}{|A|^2 + |B|^2 - A \cdot B}. \tag{3}$$

The semantic similarity between each two product attributes is obtained by Eq. (3) for all the extracted product attributes, and then the semantic similarity matrix of the product attributes is obtained. Then the product attributes are clustered by Hierarchical Cluster Analysis (HCA) method to reduce the number of product attributes [57]. The result of HCA is usually presented in a dendrogram and there are two main approaches to resolve the grouping problem in HCA, agglomerative or divisive [58]. In this study we use the agglomerative approach for cluster analysis and also choose to use the Pretty Heatmap module in ImageGP [59] for data visualization. Expressing the semantic similarity matrix and HCA in the form of heatmap allows a more visual observation of the semantic similarity product attributes and the results of HCA.

3.3 Calculating the satisfaction score of product attributes

The satisfaction score of the product attributes is to evaluate the degree of customer satisfaction with product attributes, and obtained by integrating the attention and sentiment scores of product attributes. In this paper, we take a quantitative way to calculate the sentiment scores of customers on product attributes. The obtained sentiment scores for product attributes represent the sum of the sentiment score of customers for the current product attribute in the dataset. Since customers pay different attention to each product attribute, the product attributes with high attention are mentioned more times in the reviews, so the sentiment score at this point does not represent the satisfaction of customers with the product attribute. When the sentiment score of a product attribute is obtained, the attention score of the same product attribute is calculated. Then, the ratio of sentiment score to attention score for the same product attribute is used to express the satisfaction of customers with the product attribute. The product attribute satisfaction S_i is given by:

$$\mathrm{S}_i = \frac{Q_i}{PA_i}, \tag{4}$$

where Q_i is the sentiment score corresponding to the product attribute i, and PA_i is the attention score of the product attribute i.

Sentiment analysis is mainly divided into sentiment polarity analysis and sentiment intensity analysis. Customers often express their emotions after the product experience in product reviews, so it is necessary to perform sentiment analysis on customer reviews. The sentiment scores of product attributes are calculated with a natural language processing tool of the text sentiment analysis tool SnowNLP. Since product attributes and sentiment scores are extracted separately, a correlation between them must be established to analyze the sentiment scores of consumers on product attributes. Since the previous pre-processing was done in terms of customer review sentences, each sentence contains only one or two product attributes, so we corresponded each customer review score to the extracted product attribute of that sentence, thus corresponding the product attribute to the sentiment score, and then the product attribute sentiment score was calculated. Therefore, the equation is proposed and defined as:

$$Q_i = \sum_{j=1}^{M} R_j[A_i] \times W_j, \tag{5}$$

where R_j indicates that it is the jth comment, A_i is the product attribute, $R_j[A_i]$ indicates that the product attribute in the jth comment is A_i, and W_j is the sentiment score of the jth customer comment. and M means there are M comments in total. Mathematically $R_j[A_i]$ is defined as:

$$R_j[A_i] = \begin{cases} 1, & A_i \in R_j, \\ 0, & A_i \in R_j \end{cases} \tag{6}$$

consequently, we can obtain the sentiment scores of the product attributes.

Product attribute attention indicates how much customers pay attention to a particular attribute of the product, and the three different measures defined by Rai [60] are *Frequency Measure* (FM), *Review Appearance Rate Measure* (RARM), and *Local Global Normalization Measure* (LGNM). The *Frequency Measure* knowledge counts the number of occurrences of each noun phrase representing a given product attribute in all customer reviews, and requires two additional important assumptions. The *Review Appearance Rate Measure* is simply a count of the number of different customer reviews in which a noun phrase appears painstakingly. In order to reduce the distortion caused by these two assumptions, the *Local-Global Normalization Measure* (LGNM) is proposed. the *LGNM* consists of three different types of term weights: local weights, global weights and normalization weights. In our study, the *LGNM* is used to calculate the product attribute attention PA_i:

$$PA_i = \sum_{j=1}^{M} L_{ij} * G_i * N_{ij}, \tag{7}$$

where L_{ij} is the local weight for attribute i in the review j, G_i is the global weight that measures the importance of attribute i in all the reviews, and N_{ij} is the normalisation factor to compensate the discrepancies due to the lengths of the reviews.

The local factor L_{ij} is defined as:

$$L_{ij} = \log_2(1 + f_{ij}), \tag{8}$$

where the local opinion frequency f_{ij} is defined as the number of occurrences of the attribute i in each customer review j.

The global factor G_i is defined as:

$$G_i = 1 + \sum_{j=1}^{M} \frac{\frac{f_{ij}}{F_i} \log \frac{f_{ij}}{F_i}}{\log M}, \tag{9}$$

where the global opinion frequency F_i is the frequency of attribute i throughout all the reviews:

$$F_i = \frac{\sum_{j=1}^{M} R_j[A_i]}{M}. \tag{10}$$

The normalisation factor N_{ij} is defined as:

$$N_{ij} = \frac{1}{G_i * L_{ij}}. \tag{11}$$

So far, we can acquire the attention of each product attribute, and then according to the formula (4), we can calculate the satisfaction score of the product attributes.

3.4 Dynamic change trend of product attributes

Since each generation of products is released at a different time, the social environment and similar competing products at that time are also different, resulting in customer satisfaction with the product attributes is constantly changing. In order to more accurately understand customer requirements and improve the effectiveness and efficiency of continuous product improvement, we should analyze the changing trends of customer satisfaction with product attributes over multiple generations.

Determining priorities for future product attribute improvements by dynamic trends in customer satisfaction of product attributes. There are three main types of dynamic trend changes, namely, rise-trend, fall-trend, and no change. Due to the dynamic trend changes we consider for three generations of products, there will be nine trend changes for each product attribute, namely rising-rising, rising-invariant, rising-declining, invariant-rising, invariant-invariant, invariant-declining, falling-rising, falling-invariant, and falling-declining. and are shown in Fig. 5. Within the rising-declining and falling-rising trends, each trend variation is subdivided into two types, rising-declining (middle), rising-declining (bottom), falling-rising (middle), and falling-rising (top). The details are shown in Fig. 6.

Zhang et al. [39] proposed that product attributes with reduced customer satisfaction should be improved substantially, and attributes with increased customer satisfaction should be maintained or slightly improved, so we prioritized the improvement of the above 11 product attributes dynamic trend changes as follows: 1. falling-declining > 2. rising-declining (bottom) > 3. invariant-declining > 4. rising-declining (middle) > 5. falling-invariant > 6. falling-rising (middle) > 7. invariant-invariant > 8. falling-rising (top) > 9. rising-invariant > 10. invariant-rising > 11. rising-rising.

If there are several different product attributes in the same product attribute dynamic trend changes, in this study our solution is as follows. If this product attribute dynamic trend change is among the first six, we use the relative value calculation method, specifically the difference between the highest and lowest satisfaction value of the product attribute is used as an indicator for further evaluation, and the product attribute with a large difference is given priority for improvement. If the product attribute trend changes in the last five positions, we use absolute values to calculate. This is where we compare only the highest customer satisfaction, with low satisfaction product attributes prioritized for improvement. The process is illus-

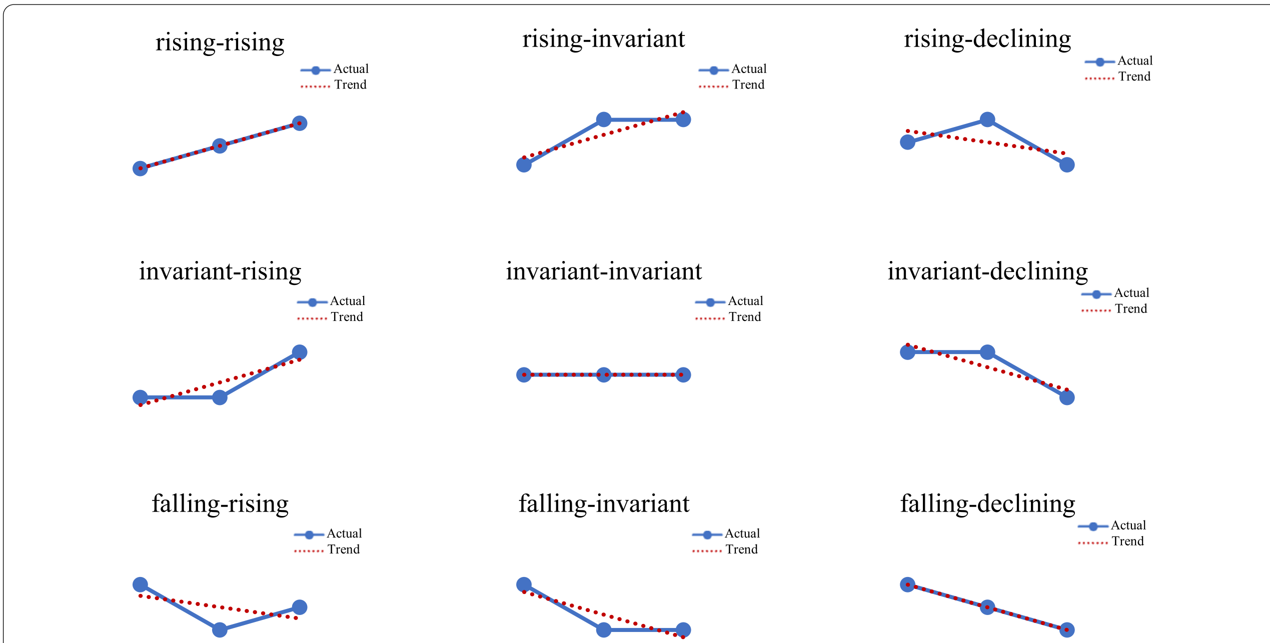


Figure 5 Data trend analysis type

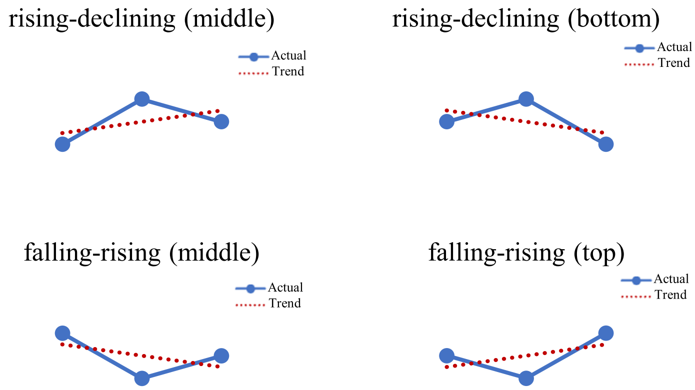


Figure 6 Rising-declining and falling-rising

trated via the following equation:

$$PI_i = \begin{cases} Max(P_{j-1}[S_i], P_j[S_i], P_{j+1}[S_i]) \\ \quad - Min(P_{j-1}[S_i], P_j[S_i], P_{j+1}[S_i]), & DT \in [1,6], \\ Max(P_{j-1}[S_i], P_j[S_i], P_{j+1}[S_i]), & DT \in [7,11], \end{cases} \tag{12}$$

where PI_i denotes the dynamic trend score of product attribute i, $P_j[S_i]$ denotes customer satisfaction with product attribute i of the jth generation of products, and DT represents the product attribute dynamic trend change described earlier.

Similarly, we also take into account the existence of special cases, such as several product attributes belonging to the same product attribute dynamic trend changes, and more coincidentally, their maximum and minimum values are also the same, which may make the PI_i values we obtained the same, and then we will further compare these several product attributes by analysis to give the improved ranking of these product attributes. Of course, the probability of this situation existing is tiny.

4 Case study

In this section, to illustrate the proposed method, a phone case study is presented which captures the dynamic trend change of customer satisfaction of product attributes from customer online review. Analyzing this dynamic trend change to determine the direction of next-generation product improvement helps companies accurately understand customer requirements and improve the efficiency of continuous product improvement.

4.1 Data preparation

Customer online reviews were collected from a shopping platform. First of all, the data is obtained by crawling program to obtain customer reviews of three generations of a brand of phones. The latest product is called product 3rd Gen, and then the past two generations are called product 1st Gen and product 2nd Gen. Since the brand is updated annually, the review data is generated within the last three years. After getting the comments, we got a total of 63,467 pieces of raw data, including 21,405 pieces of product 1st Gen, 22,985 pieces of product 2nd Gen and 19,077 pieces of product 3rd Gen. Then the data is cleaned and pre-processed. Finally, we got a total of 50,797 product reviews, including 17,822 for product 1st Gen, 18,810 for product 2nd Gen and 14,165 for product 3rd Gen. As shown in Table 1.

Table 1 Number of customer reviews

	Product 1st Gen	Product 2nd Gen	Product 3rd Gen	Total
Raw reviews	21,405	22,985	19,077	63,467
Pre-processed reviews	17,822	18,810	14,165	50,797

4.2 Identifying product attributes

Due to our analysis of customer satisfaction of product attributes on multi-generation products, all the pre-processed review data were applied to product attribute extraction in this work. The following steps were taken to process the data. First, we used the Natural Language Processing tool of jieba to tokenizer and store the obtained data into a new text. Second, a lexicon of 1893 words was created by jointing three dictionaries of Harvard deactivation word list, Chinese deactivation word list and Baidu deactivation word list, which aimed to remove noise such as punctuation and stop words. Third, the importance weight of product attributes was determined using the TextRank method. In this case, the top 100 product attributes with high importance weights were selected for further research, as shown in Table 2.

We find a number of irrelevant product attributes in Table 2, such as mobile phone, apple, effect, etc., so we should

Table 2 Top 100 product attributes extracted from online reviews

The top 100 product attributes extracted from online reviews									
mobile phone	nice	*Genuine*	cost-effective	performance	serve	*praise*	*function*	quality	resolution
speed	sound effect	*worth*	*shopping*	*hope*	*rest assured*	*cost performance*	black	*overall*	*evaluate*
Apple	price	quality	*battery*	purple	green	video	*overall*	*frame*	*need not*
photograph	color	package	*experience*	white	signal	*time*	RAM	*promote*	*commodity*
appearance	*feel*	*activity*	face value	camera	*recommend*	*place an order*	*friend*	screen display	*easy to use*
effect	logistics	*buy*	*discount*	affordable	*atmosphere*	show	*choose*	*habit*	charger
Run	system	delivery	customer service	*very beautiful*	*product*	deliver goods	*Work*	service attitude	
Like	*receive*	standby time	*feature*	*start*	*support*	*play games*	*classic*	*get*	*look at*
screen	feel	express delivery	red	trust	texture	pixel	*open*	*the arrival*	*enough*
shape	*satisfy*	charge	sound quality	*self-employed*	*design*	*suit*	*discover*	attitude	photo

Table 3 Filtered product attributes

Filtered product attributes							
speed	nice	feel	charge	sound quality	trust	show	screen display
photograph	sound effect	quality	cost-effective	performance	serve	deliver goods	service attitude
appearance	price	packaging	battery	purple	green	pixel	attitude
run	color	delivery	face value	white	signal	black	resolution
screen	logistic	standby time	customer service	camera	texture	RAM	charger
shape	system	express delivery	red	affordable	video	quality	photo

remove these attributes. Finally, Table 3 shows the 48 product attributes were obtained. As shown in Table 3, it is not difficult to find cases where the product attributes are expressed differently but have similar meanings, so we need to reduce the dimensionality of these product attributes.

To address this issue, in this paper an approach of semantic similarity analysis is performed. Firstly, the semantic similarity is calculated, and obtain the semantic similarity of each product attribute with other product attributes. Figure 7 shows the heatmap and cluster map of the similarity matrix of product attributes. The closer the result is to 1, the more the meanings of the two product attributes are similar and the deeper the color shown in Fig. 7. Then the HCA is performed, and the leftmost and uppermost line segments indicate the results of the HCA. We show several of the results, as shown in the dashed boxes in Fig. 7, which provide the basis for determining the product attributes. Finally, we obtained 11 product attributes. The specific product attributes and the corresponding specific interpretations are shown in Table 4 [61].

4.3 Calculating the satisfaction of multi-generation product attributes

After obtaining the product attributes, customer satisfaction is calculated for each product attribute of product 1st generation, product 2nd generation and product 3rd generation. To make the experimental results more accurate, in this work we control the number of online reviews for each product generation to be equal, so we take the product with the least number of customer reviews as the benchmark. In order to obtain customer satisfaction of product attributes, we need to first calculate the attention and sentiment scores of product attributes.

The sentiment scores of product attributes are mainly determined by Eqs. (3), (4), and the sentiment scores for each product attribute of multi-generation products were obtained by calculation, as shown in Fig. 8.

On the one hand, Fig. 8 shows the dynamic changes in the sentiment scores of product attributes over multi-generation products. Looking at Fig. 8, we clearly see that the sentiment scores of the product attributes don't increase with the upgrades of product and are decreasing for most product attributes, such as system, screen, sound effect, price, quality and RAM. The possible reason for this phenomenon is that the upgrade of the product doesn't meet the customer requirements. On the other hand, the sentiment scores of product attributes also reflect the customer preferences for the product attributes. As shown in Fig. 8, we obviously see that customers have high sentiment scores of product attributes such as appearance, system and serve. However, customers have lower sentiment scores of product attributes such as RAM, signal and battery. This indicates that customers may prefer to product attributes such as appearance and system.

Then we calculation the attention scores of multi-generation product attributes. Firstly, the values of the local factor L_{ij}, global factor G_i and normalisation factor N_{ij} for each product attribute of multi-generation products were obtain from Eqs. (6), (7), (8), (9). Then, the attention scores of product attribute of multi-generation products is calculated from Eq. (5), as shown in Fig. 8.

Figure 9 shows the differences of the attention scores of multi-generation product attributes, which illustrate that

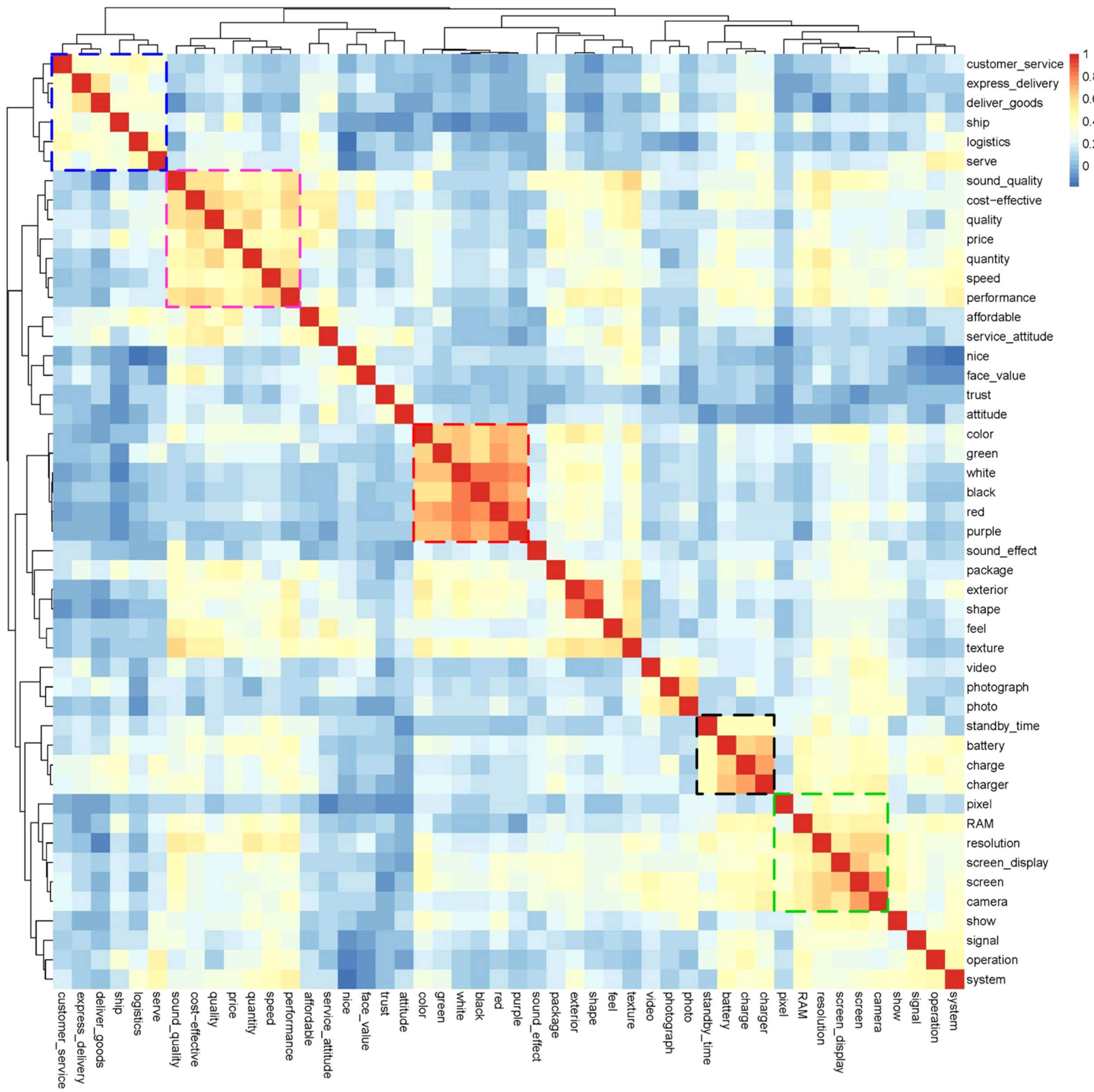


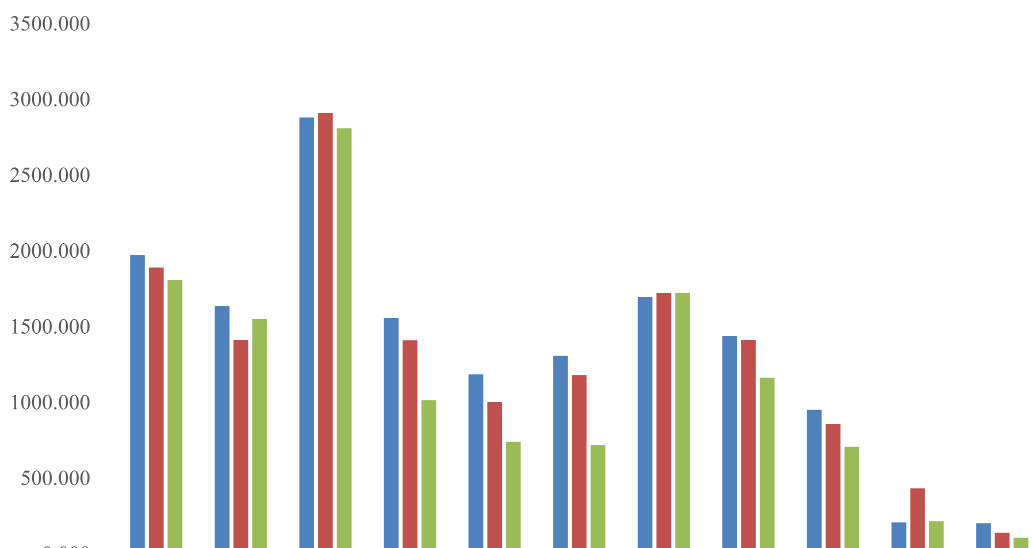
Figure 7 Heatmap and cluster map of the similarity matrix of product attributes

customers pay different attention to different product attributes. As shown in Fig. 9, customers pay more attention to appearance, system, service and camera, which on the one hand indicates that customer attach more importance to these product attributes, and on the other hand shows that these product attributes may be more likely to attract customer attention because they are currently changing more. The other lower product attributes could be due to the product being more mature and without larger changes.

After the attention and sentiment scores of product attributes were determined, we calculated the satisfaction scores of product attributes to analyze the dynamic changes of product attributes. Figure 10 shows the satisfaction scores S_i, which suggests that customer satisfaction varies across product attributes, while customer satisfaction for the same product attribute is dynamic across multi-generation products. For example, customer satisfaction with the sound effect attribute of the product is higher and with the battery attribute is lower. Similarly

Table 4 Product attributes

No.	Product Attribute	Related product attributes	Explanation
1	System	System, Speed, Run, System, Performance	IOS or Android, and some applications and functions in system.
2	Camera	Camera, Photograph, video, pixel, photo	Camera number, image accuracy, aperture size, image stabilization performance, etc.
3	Appearance	Appearance, Shape, good looking, color, face value, red, purple, white, green, black	The shape, color, and material of the smartphone.
4	Screen	Screen, Show, screen display, resolution	Screen size, resolution, color gamut, and so on.
5	Sound effect	Sound effect, sound quality	Audio and video playback quality.
6	Price	Price, Cost performance, affordable	The price of the smartphone.
7	Serve	Serve, Logistic, packaging, delivery, express delivery, customer service, trust, deliver goods, Service attitude, Attitude	Pre-sale consultation, after-sales service, etc.
8	Quality	Quality, Cost-effective, hand feel, texture	Product durability and product quality.
9	Battery	Battery, Standby time, Charge, charger	Capacity and endurance of the battery.
10	Signal	Signal	The network performance of the smartphone.
11	RAM	RAM	Memory capacity, read/write speed, etc.



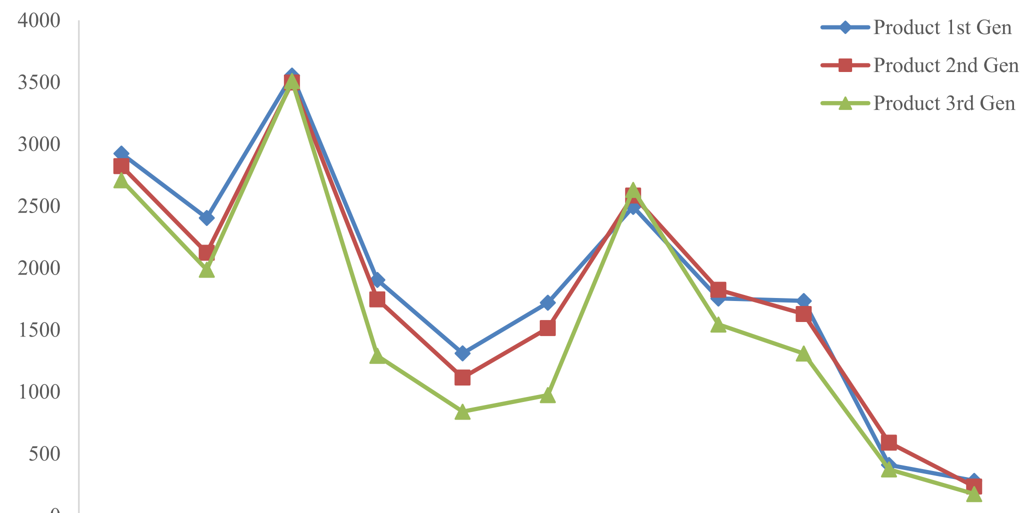
	System	Camera	Appearance	Screen	Sound effect	Price	Serve	Quality	Battery	Signal	RAM
Product 1st Gen	1972.412	1636.580	2880.926	1556.410	1186.155	1308.592	1696.293	1437.420	951.285	208.697	202.796
Product 2nd Gen	1890.795	1411.135	2911.577	1410.282	1002.823	1180.144	1723.927	1411.807	857.549	433.127	140.057
Product 3rd Gen	1806.780	1550.374	2809.443	1015.053	740.561	718.266	1724.894	1164.024	706.622	216.433	107.177

Figure 8 The sentiment scores for each product attribute of multi-generation products

in multi-generation products, customer satisfaction of the product attributes such as system, screen, sound, and service are always decreasing, while satisfaction of the product attributes such as toward signal and RAM is changing, and the change means that customer satisfaction may rise first and then decline or fall first and then rise.

4.4 Data analysis and improving direction of product attributes

The calculation process of changes in product attributes was discussed above. According to Fig. 10, customer satisfaction with the same product attributes over multi-generation products is dynamic, so we can perform trend analysis based on dynamic data. Figure 11 shows the dynamic trend changes of the same product attributes of multi-generation products by customers. As can be seen from Fig. 11, we can see that customer satisfaction with the product attributes of quality, screen, service, sound, and system is always decreasing, which shows that there is a distance between the customer expectation of these product attributes and the actual product attributes obtained, resulting in the customer satisfaction has been decreasing; the user satisfaction of price, appearance and signal is ris-



	System	Camera	Appear-ance	Screen	Sound effect	Price	Serve	Quality	Battery	Signal	RAM
Product 1st Gen	2924	2404	3554	1904	1311	1720	2494	1754	1734	410	282
Product 2nd Gen	2823	2123	3498	1747	1115	1515	2585	1824	1629	590	235
Product 3rd Gen	2709	1986	3511	1291	840	974	2632	1544	1310	374	175

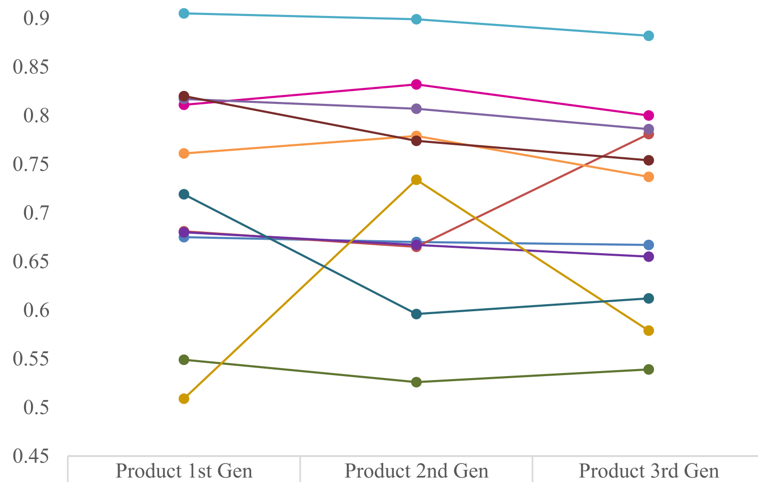
Figure 9 The attention scores for each product attribute of multi-generation products

ing first and then decreasing, the difference is that the first two product attributes are decreasing to the lowest point, and signal is down to the middle value. The customer satisfaction of battery, memory and camera decreases first and then increases, the difference is that the first two product attributes increase to the middle value, and the customer satisfaction of camera increases to the highest value. Therefore, according to the product attribute data trend analysis and Eq. (10), we can get the priority of continuous product improvement attributes as quality, screen, service, sound, system, price, appearance, signal, battery, memory, and camera.

The analysis of consumer requirements based on online product reviews can provide designers with innovative ideas [62], and the future product improvements can also be influenced by customer sentiment and attention of product attributes. Designers can evaluate previous product improvement strategies by analyzing changes in customer satisfaction with product attributes and also bring new product improvement strategies to designers.

According to the changes in the data trends of customer satisfaction with product attributes, it can be seen that customer satisfaction with the product attributes of quality, screen, service, sound, and system has been decreasing. There are two main reasons for customer dissatisfaction with the product: the first reason is that the product attributes have not been upgraded or improved, which leads to a decrease in customer satisfaction; the second reason is that the product attributes have been improved or upgraded, but the customer expectations have also increased over time. However, customer expectations have also increased with the change of time or the emergence of other competing products, so the product attributes still do not meet customer expectations, leading to a decrease in customer satisfaction. Manufacturers should pay more attention to the attributes of such products in the next generation of product design. For example, in terms of screen, to focus on the quality of the screen and more attention to the materials used by competitive products, to improve. In terms of service, service here mainly includes customer service, logistics and after-sales, etc., to focus on customer service is the tone of voice and attitude when communicating with users, logistics should be timely, after-sales also need to actively communicate with users to find a way to deal with the problem to the satisfaction of both sides.

For the product attributes of price, appearance and signal customer satisfaction rise first and then fall, where for the price, customer satisfaction rises first because the price of the product is always fall, because the first time the customers did not expect the price of the product will be decline so customer satisfaction will increase, because it does not reach the level of reduction in customer expectations, which is also the reason for the second change in customer satisfaction, and similarly appearance is also the same. For signal, the satisfaction score of product attribute changed from "0.509" to "0.734" to "0.579", which is indicative the product introduced new signal technology makes



	Product 1st Gen	Product 2nd Gen	Product 3rd Gen
System	0.675	0.67	0.667
Camera	0.681	0.665	0.781
Appearance	0.811	0.832	0.8
Screen	0.817	0.807	0.786
Sound effect	0.905	0.899	0.882
Price	0.761	0.779	0.737
Serve	0.68	0.667	0.655
Quality	0.82	0.774	0.754
Battery	0.549	0.526	0.539
Signal	0.509	0.734	0.579
RAM	0.719	0.596	0.612

Figure 10 Customer satisfaction for each product attribute of multi-generation products

customer satisfaction increase in the first time, while later no new technology was introduced to make customer satisfaction decrease. It also provides product designers with new directions for improvement strategies, when there is a new technology introduced into the product design, it will improve customer satisfaction with the product. For this type of product attributes manufacturers and product designers should first analyze the reasons for the recent decline in customer satisfaction and then develop appropriate strategies to address this reason when designing the next generation of products to meet customer satisfaction. For the product attributes of battery, memory and camera customer satisfaction is first fall and then rise, it can be inferred that the designer got improvement in time after learning that customers are not satisfied with the product attributes, thus improving the customer satisfaction at the end, which shows that the product improvement strategy is correct. Therefore, the product designer only needs to continue to maintain the improvement strategy for such product attributes when designing the next generation of products.

5 Conclusions

In this paper, we focus on dynamically mining user requirements from online reviews, and discover the correlations between product attribute satisfaction and product improvement, which guide the direction and contents of continuous product improvement. We first collected a large number of multi-generation product customer review data from shopping platform, which were used to capture the dynamic changes of consumer satisfaction of product attributes by measuring consumer attention and sentiment. Then the product improvement strategies were analyzed based on the changes of consumer satisfaction with product attributes for next-generation product, which could meet customer requirements with high efficiency, and improve the effectiveness and efficiency of continuous product improvement.

The contribution of this paper to the field of consumer requirements analysis is mainly in two aspects. First, customer satisfaction with product attributes is quantified by combining customer attention and sentiment analysis. Second, correlating changes in customer satisfaction with continuous product improvement provides a dynamic approach to customer requirements analysis for continuous

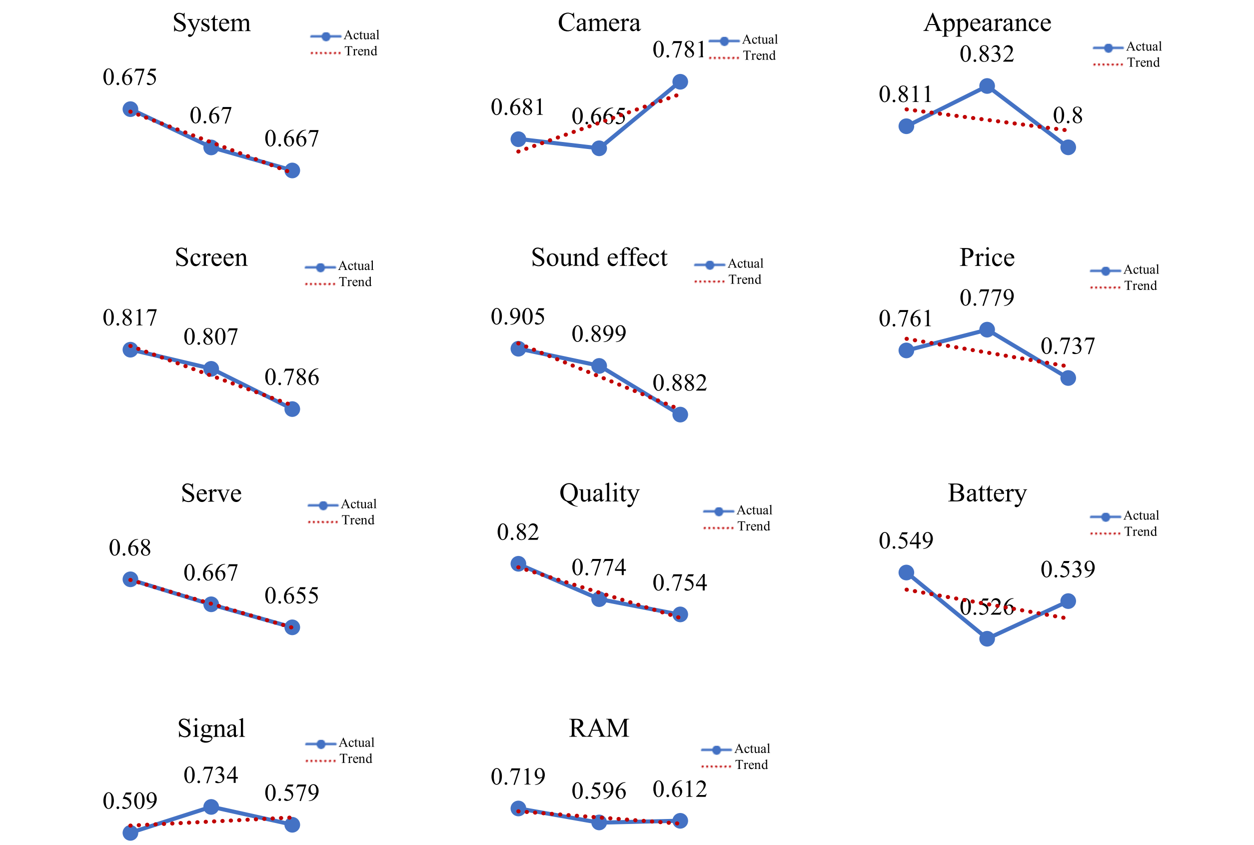


Figure 11 Dynamic trend of product attributes of multi-generation products

improvement of multi-generational products. Moreover, the feasibility of the method is demonstrated by the example of multi-generation phone products. The results showed that can help enterprises and designers to make better and faster decision in future product improvement. However, our study has some limitations. First, there are manual steps in the process of identifying product attributes that may affect the results. Second, only text data was used in this study, and data such as pictures and videos exist in the shopping platform that were not used in the study. Future works can focus on the following directions:

1) Improve the automation of the dynamic customer requirements mining method and reduce the human involvement in the process.
2) Develop a method for analyzing customer requirements based on multi-modal online data including text reviews, pictures and videos.

Funding
This work was supported by two Sichuan Province Science Technology Support Programs (2021YFG0056 and 2021YFG0051) and two City-University Research Funds (2020CDYB-12 and 2021CDLZ-2).

Availability of data and materials
Not applicable.

Code availability
Not applicable.

Declarations

Competing interests
The authors declare that they have no competing interests.

Authors' contributions
All authors contributed to the study conception and design. Material preparation and analysis were performed by QZ, WZ, XG, KZ and MY. The first draft of the manuscript was written by QZ and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Author details
[1]School of Mechanical Engineering, Sichuan University, Chengdu, 610065, China. [2]The Innovation Method and Creative Design Key Laboratory of Sichuan Province, Sichuan University, Chengdu, 610065, China. [3]Yibin Institute of Industrial Technology, Sichuan University Yibin Park, Yibin, 643000, China.

Publisher's Note

Received: 25 April 2022 Revised: 20 May 2022 Accepted: 26 May 2022
Published online: 01 July 2022

References

1. H. Sun, W. Guo, H. Shao, B. Rong, Dynamical mining of ever-changing user requirements: a product design and improvement perspective. Adv. Eng. Inform. **46**, 101174 (2020). https://doi.org/10.1016/j.aei.2020.101174
2. Q. Yang, Z. Ye, X. Li, Z. Zhang, W. Chang, S. Ruan, Product module attribute parameter configuration model considering customer requirements preferences. Math. Probl. Eng. **2021**, 6632057 (2021)
3. L. Wang, Y. Xu, H. Lee, A. Li, Preferred product attributes for sustainable outdoor apparel: a conjoint analysis approach. Sustain. Prod. Consump. **29**, 657–671 (2022)
4. J. Joung, H.M. Kim, Automated keyword filtering in latent Dirichlet allocation for identifying product attributes from online reviews. J. Mech. Des. **143**(8), 084501 (2021). https://doi.org/10.1115/1.4048960
5. Y. Shi, Q. Peng, Definition of customer requirements in big data using word vectors and affinity propagation clustering. J. Process Mech. Eng. **235**(5), 1279–1291 (2021). https://doi.org/10.1177/09544089211001776
6. X. Guo, Y. Liu, W. Zhao, J. Wang, L. Chen, Supporting resilient conceptual design using functional decomposition and conflict resolution. Adv. Eng. Inform. **48**, 101262 (2021). https://doi.org/10.1016/j.aei.2021.101262
7. X. Guo, W. Zhao, H. Hu, L. Li, Y. Liu, J. Wang, K. Zhang, A smart knowledge deployment method for the conceptual design of low-carbon products. J. Clean. Prod. **321**, 128994 (2021). https://doi.org/10.1016/j.jclepro.2021.128994
8. S. Stieglitz, M. Mirbabaie, B. Ross, C. Neuberger, Social media analytics – challenges in topic discovery, data collection, and data preparation. Int. J. Inf. Manag. **39**, 156–168 (2018). https://doi.org/10.1016/j.ijinfomgt.2017.12.002
9. S. Xiao, C. Wei, M. Dong, Crowd intelligence: analyzing online product reviews for preference measurement. Inf. Manag. **53**(2), 169–182 (2016). https://doi.org/10.1016/j.im.2015.09.010
10. L. Fink, L. Rosenfeld, G. Ravid, Longer online reviews are not necessarily better. Int. J. Inf. Manag. **39**, 30–37 (2018). https://doi.org/10.1016/j.ijinfomgt.2017.11.002
11. W. He, X. Tian, A. Hung, V. Akula, W. Zhang, Measuring and comparing service quality metrics through social media analytics: a case study. Inf. Syst. E-Bus. Manag. **16**(3), 579–600 (2017). https://doi.org/10.1007/s10257-017-0360-0
12. T. Escobar-Rodríguez, R. Bonsón-Fernández, Analysing online purchase intention in Spain: fashion e-commerce. Inf. Syst. E-Bus. Manag. **15**(3), 599–622 (2016). https://doi.org/10.1007/s10257-016-0319-6
13. K.Z.K. Zhang, S.J. Zhao, C.M.K. Cheung, M.K.O. Lee, Examining the influence of online reviews on consumers' decision-making: a heuristic–systematic model. Decis. Support Syst. **67**, 78–89 (2014). https://doi.org/10.1016/j.dss.2014.08.005
14. J. Liu, C. Liu, L. Zhang, Y. Xu, Research on sales information prediction system of e-commerce enterprises based on time series model. Inf. Syst. E-Bus. Manag. **18**(4), 823–836 (2019). https://doi.org/10.1007/s10257-019-00399-7
15. J. Ren, W. Yeoh, M. Shan Ee, A. Popovič, Online consumer reviews and sales: examining the chicken-egg relationships. J. Assoc. Inf. Sci. Technol. **69**(3), 449–460 (2018). https://doi.org/10.1002/asi.23967
16. S. Güneş, Extracting online product review patterns and causes: a new aspect/cause based heuristic for designers. Des. J. **23**(3), 375–393 (2020). https://doi.org/10.1080/14606925.2020.1746611
17. J. Cao, W. Zhao, H. Hu, Y. Liu, X. Guo, Using linkography and situated FBS co-design model to explore user participatory conceptual design process. Processes **10**(4), 713 (2022). https://doi.org/10.3390/pr10040713
18. Y. Du, D. Liu, H. Duan, A textual data-driven method to identify and prioritise user preferences based on regret/rejoicing perception for smart and connected products. Int. J. Prod. Res. 1–21 (2022). https://doi.org/10.1080/00207543.2021.2023776
19. B. Gao, M. Zhu, S. Liu, M. Jiang, Different voices between Airbnb and hotel customers: an integrated analysis of online reviews using structural topic model. J. Hosp. Tour. Manag. **51**, 119–131 (2022). https://doi.org/10.1016/j.jhtm.2022.03.004
20. J. Joung, H.M. Kim, Automated keyword filtering in latent Dirichlet allocation for identifying product attributes from online reviews. J. Mech. Des. **143**(8), 084501 (2021). https://doi.org/10.1115/1.4048960
21. W. Fu, E. Choi, H. Kim, Text mining with network analysis of online reviews and consumers' satisfaction: a case study in busan wine bars. Information **13**(3), 127 (2022). https://doi.org/10.3390/info13030127
22. D. Zhang, P. Wu, C. Wu, The role of key online reviews in affecting online hotel booking: an empirical investigation. Ind. Manag. Data Syst. **122**(2), 499–520 (2022)
23. C. Forman, A. Ghose, B. Wiesenfeld, Examining the relationship between reviews and sales: the role of reviewer identity disclosure in electronic markets. Inf. Syst. Res. **19**(3), 291–313 (2008). https://doi.org/10.1287/isre.1080.0193
24. J. Mallick, S. Talukdar, M. Alsubih, R. Salam, M. Ahmed, N.B. Kahla, M. Shamimuzzaman, Analysing the trend of rainfall in Asir region of Saudi Arabia using the family of Mann-Kendall tests, innovative trend analysis, and detrended fluctuation analysis. Theor. Appl. Climatol. **143**(1–2), 823–841 (2020). https://doi.org/10.1007/s00704-020-03448-1
25. Q. Qi, B. Jiang, W. Ma, G. Marley, Trend analysis and spatial distribution of meteorological disaster losses in China, 2004–2015. Atmosphere **13**(208), 208 (2022). https://doi.org/10.3390/atmos13020208
26. S.H.L. Yim, T. Huang, J.M.W. Ho, A.S.M. Lam, S.T.Y. Yau, T.W.H. Yuen, G.H. Dong, K.K.F. Tsoi, J.J.Y. Sung, Rise and fall of lung cancers in relation to tobacco smoking and air pollution. A global trend analysis from 1990 to 2012. Atmos. Environ. **269**, 118835 (2022). https://doi.org/10.1016/j.atmosenv.2021.118835
27. S. Romano, H. Galante, D. Figueira, Z. Mendes, A.T. Rodrigues, Time-trend analysis of medicine sales and shortages during Covid-19 outbreak: data from community pharmacies. Res. Social Adm. Pharm. **17**(1), 1876–1881 (2021). https://doi.org/10.1016/j.sapharm.2020.05.024
28. M. Amini, F. Zayeri, M. Salehi, Trend analysis of cardiovascular disease mortality, incidence, and mortality-to-incidence ratio: results from global burden of disease study 2017. BMC Public Health **21**(1), 401 (2021). https://doi.org/10.1186/s12889-021-10429-0
29. X. Guo, J. Wang, W. Zhao, K. Zhang, C. Wang, Study of medical device innovation design strategy based on demand analysis and process case base. Multimed. Tools Appl. **75**(22), 14351–14365 (2016)
30. X. Guo, Y. Jiang, Y. Liu, Conceptual design for innovation: process and a knowledge-based approach, in (Springer, 2022), pp. 179-198
31. H. Liu, Z. Ren, Y. Li, Research on the evaluation of customer requirement importance of smart products based on online comments and improved quantitative Kano model. Converter **2021**, 382–392 (2021). https://doi.org/10.17762/converter.301
32. A. Timoshenko, J.R. Hauser, Identifying customer needs from user-generated content. Mark. Sci. **38**(1), 1–20 (2019). https://doi.org/10.1287/mksc.2018.1123
33. K.Q. Anh, Y. Nagai, L.M. Nguyen, Extracting customer reviews from online shopping and its perspective on product design. Vietnam J. Comput. Sci. **6**(01), 43–56 (2019). https://doi.org/10.1142/S2196888819500088
34. V.R. Hananto, S. Kim, M. Kovacs, U. Serdult, V. Kryssanov, A machine learning approach to analyze fashion styles from large collections of online customer reviews, in *2021 6th International Conference on Business and Industrial Research (ICBIR)* (IEEE, 2021), pp. 153–158. https://doi.org/10.1109/ICBIR52339.2021.9465830
35. K.Y. Chan, C.K. Kwong, H. Jiang, Analyzing imbalanced online consumer review data in product design using geometric semantic genetic programming. Eng. Appl. Artif. Intell. **105**, 104442 (2021). https://doi.org/10.1016/j.engappai.2021.104442
36. Y. Sun, J.h. Dai, A fine-grained emotional analysis of E-commerce product review data based on dictionaries, in *2021 2nd International Conference on Education, Knowledge and Information Management (ICEKIM)* (IEEE, 2021), pp. 257–262. https://doi.org/10.1109/ICEKIM52309.2021.00063
37. J. Zhang, A. Zhang, D. Liu, Y. Bian, Customer preferences extraction for air purifiers based on fine-grained sentiment analysis of online reviews. Knowl.-Based Syst. **228**, 107259 (2021). https://doi.org/10.1016/j.knosys.2021.107259
38. N. Li, X. Jin, Y. Li, Identification of key customer requirements based on online reviews. J. Intell. Fuzzy Syst. **39**(3), 3957–3970 (2020). https://doi.org/10.3233/JIFS-200057
39. H. Zhang, H. Rao, J. Feng, N.Y. Yen, V. Sugumaran, Z. Xu, Z. Xu, V. Sugumaran, N.Y. Yen, Product innovation based on online review data mining: a case study of Huawei phones. Electron. Commer. Res. **18**(1), 3–22 (2017). https://doi.org/10.1007/s10660-017-9279-2
40. Y. Han, M. Moghaddam, Eliciting attribute-level user needs from online reviews with deep language models and information extraction. J. Mech. Des. **143**(6), 061403 (2021). https://doi.org/10.1115/1.4048819
41. C. Mosha, Y. Tianfang, Combining dependency parsing with shallow semantic analysis for Chinese opinion-element relation identification, in

2010 4th International Universal Communication Symposium (IEEE, 2010), pp. 299–305. https://doi.org/10.1109/IUCS.2010.5666009
42. G.S. Budhi, R. Chiong, I. Pranata, Z. Hu, Using machine learning to predict the sentiment of online reviews: a new framework for comparative analysis. Arch. Comput. Methods Eng. **28**(4), 2543–2566 (2021)
43. D. Wu, R. Yang, C. Shen, Sentiment word co-occurrence and knowledge pair feature extraction based LDA short text clustering algorithm. J. Intell. Inf. Syst. **56**(1), 1–23 (2021)
44. M. Hu, B. Liu, Mining and summarizing customer reviews, in *Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (ACM, New York, 2004), pp. 168–177. https://doi.org/10.1145/1014052.1014073
45. M. Hu, B. Liu, Mining opinion features in customer reviews, in (2004), pp. 755–760
46. M. Taboada, Sentiment analysis: an overview from linguistics. Annu. Rev. Linguist. **2**, 325–347 (2016). https://doi.org/10.1146/annurev-linguistics-011415-040518
47. C.R. Fink, D.S. Chou, J.J. Kopecky, A.J. Llorens, Coarse- and fine-grained sentiment analysis of social media text. Johns Hopkins APL Tech. Dig. **30**(1), 22–30 (2011)
48. M.Z. Asghar, A. Khan, S. Ahmad, M. Qasim, I.A. Khan, Lexicon-enhanced sentiment analysis framework using rule-based classification scheme. PLoS ONE **12**(2), e171649 (2017). https://doi.org/10.1371/journal.pone.0171649
49. A.S. Rathor, A. Agarwal, P. Dimri, Comparative study of machine learning approaches for Amazon reviews. Proc. Comput. Sci. **132**, 1552–1561 (2018). https://doi.org/10.1016/j.procs.2018.05.119
50. S.V. Wawre, S.N. Deshmukh, Sentiment classification using machine learning techniques. Int. J. Sci. Res. **5**(4), 819–821 (2016)
51. J. Zeng, X. Ma, K. Zhou, Enhancing attention-based LSTM with position context for aspect-level sentiment classification. IEEE Access **7**, 20462–20471 (2019)
52. X. Feng, Z. Zhang, S. J, Text sentiment analysis based on convolutional neural networks and attention model. Appl. Res. Comput. (2018)
53. R. Mihalcea, P. Tarau, Textrank: bringing order into text (2004), pp. 404–411
54. M. Tu, X. Liu, S. Liu, *Natural Language Processing Core Technology and Algorithm with Python* (2018)
55. T. Mikolov, I. Sutskever, K. Chen, G.S. Corrado, J. Dean, Distributed representations of words and phrases and their compositionality, in *Advances in Neural Information Processing Systems*, vol. 26 (2013)
56. T. Mikolov, K. Chen, G. Corrado, J. Dean, Efficient estimation of word representations in vector space (2013) https://doi.org/10.48550/arXiv.1301.3781. arXiv preprint. arXiv:1301.3781
57. L. Zhang, X. Chu, D. Xue, Identification of the to-be-improved product features based on online reviews for product redesign. Int. J. Prod. Res. **57**(8), 2464–2479 (2019). https://doi.org/10.1080/00207543.2018.1521019
58. D. Granato, J.S. Santos, G.B. Escher, B.L. Ferreira, R.M. Maggio, Use of principal component analysis (PCA) and hierarchical cluster analysis (HCA) for multivariate association between bioactive compounds and functional properties in foods: a critical perspective. Trends Food Sci. Technol. **72**, 83–90 (2018). https://doi.org/10.1016/j.tifs.2017.12.006
59. T. Chen, Y.X. Liu, L. Huang, ImageGP: an easy-to-use data visualization web server for scientific researchers. iMeta **1**(1), e5 (2022). https://doi.org/10.1002/imt2.5
60. R. Rai, Identifying key product attributes and their importance levels from online customer reviews, in *International Design Engineering Technical Conferences and Computers and Information in Engineering Conference* (American Society of Mechanical Engineers, 2012), pp. 533–540. https://doi.org/10.1115/DETC2012-70493
61. C. Yang, L. Wu, K. Tan, C. Yu, Y. Zhou, Y. Tao, Y. Song, Online user review analysis for product evaluation and improvement. J. Theor. Appl. Electron. Commer. Res. **16**(5), 1598–1611 (2021). https://doi.org/10.3390/jtaer16050090
62. M. Zhang, B. Fan, N. Zhang, W. Wang, W. Fan, Mining product innovation ideas from online reviews. Inf. Process. Manag. **58**(1), 102389 (2021). https://doi.org/10.1016/j.ipm.2020.102389