A Computer-Aided Approach for Acquisition and Importance Ranking of Customer Requirements from the Online Comment Mining

Li Xiangdong[1,2], He Youlong[1], Qingjin Peng[3] and Liu Wei[1,2]

[1]School of Economics and Management, Hebei University of Technology, lwofhebut@126.com
[2]National Technology Innovation Method and Implementation Tool Engineering Technology Research Center,
[3]University of Manitoba

Corresponding author: Liu Wei, lwofhebut@126.com

Abstract. Analysis of customer requirements (CRs) is the premise of product innovation and the key to improve the marketing competitiveness of products. This paper proposes an integrative approach in which word association model, importance ranking method, Kano model, sentiment analysis and quality function deployment (QFD) are collaboratively used to acquire and analyze CRs. To do so, the word association model is firstly applied to mine CRs from the online customer comments. Then, a new importance ranking method is designed to calculate the importance weights of CRs, in which the enterprise's expectation for product improvement and key points of sales are taken into account. Subsequently, Kano model and sentiment analysis are employed to determine the final importance ranking of CRs. Finally, a mobile phone online comment is used as a case study to address the practical value of the proposed approach for enterprise to promote product iterative innovation.

Keywords: online comment mining; customer requirements; requirement acquisition; importance ranking; quality function deployment
DOI: https://doi.org/10.14733/cadaps.2022.132-151

1 INTRODUCTION

Nowadays, product design has developed from product-oriented design to customer-oriented design [1], therefore, the analysis of CRs is more important for enterprise to improve the satisfaction of customers [2]. CRs are the basis of product development [3], poor understanding of CRs leads to negative impacts on the performances of product design and development [4].

With the continuous development of e-commerce technology, online comments have become an important information source for customers to find product information before they make purchase decisions [5, 6]. More and more customers' comments on products or services are published on websites (e.g., virtual communities, forums, BBS, online shopping platforms) [7]. The online comments have spawned new opportunities to acquire and analyze CRs [8].

Existing research on online comments has mostly focused on customers' purchase decision [9, 10], personalized recommendation [11, 12], product design [13, 14], and enterprise sales revenue forecasting [15]. Another important use in online comments is acquiring and analyzing CRs. Online comments are appealing sources of CRs, especially for manufacturers who must continually renovate their products in the competitive market [16]. Conjoint analysis and the Kano model are two widely known methods based on online comments used to acquire and analyze CRs. Li et al. [17] combined sentiment analysis with the technical characteristics of product to calculate utility values for the technical characteristics, and integrated Kano model to identify the key customer requirements. Qi et al. [18] applied the Kano method based on the conjoint analysis model to analyze online comments to develop appropriate product improvement strategies. However, these methods are mostly based on customers and products themselves, and ignoring the influence of factors such as corporate sales, product improvement ratio and customer preference. In addition, previous efforts fail to combine CRs with importance analysis, which reduces the accuracy and efficiency of CRs' analysis.

This research proposes a computer-aided approach, which is an integrative method to design a complete system from acquisition of CRs to product design and development. It is achieved by a hybrid analytical approach consisting of text mining, word association model, importance ranking and QFD [19]. First, we use text mining technology to process online customer comments, including word segmentation, data cleansing, part of speech (POS) tagging, word frequency and weight statistics. All the text analysis results are used to build the word association model to acquire CRs. This model can solve the problem of the expression difference of customers by combining a variety of words and their POS. Second, design the importance calculation method based on the weight of words, and take the key points of sales and the proportion of product improvement into account. Third, employ the Kano model and sentiment analysis to improve the importance ranking of CRs. This importance calculation method comprehensively considers the customers' sentiments and preferences, improves the accuracy and rationality of importance ranking, and improves the efficiency of importance calculation with the aid of computer. Finally, integrate CRs and their importance analysis results into the QFD to build an integrated system for the transformation of CRs. This integrated system can reduce the time and cost of product design and eliminate the information distortion by simplifying the transformation process from CRs to technical characteristics of product. And a case study implemented the acquisition and analysis of CRs for a smartphone to demonstrate the feasibility and effectiveness of the proposed methodology.

The rest of this research is organized as follows. In Section 2, a brief review of related researches is provided, it includes acquisition and importance ranking of CRs. In Section 3, methods to acquire CRs and calculate their importance are introduced. In Section 4, a case study is implemented to demonstrate the effectiveness of the proposed approach. In Section 5, discussions and conclusions are presented.

2 LITERATURE REVIEW

2.1 Acquisition of Customer Requirements

Previous methods for acquiring CRs include conversational methods, observation methods, analytical methods and collaboration methods [20]. Technologies to acquire CRs can be divided into the traditional requirement acquisition technology, the cognitive analysis technology, the modern group inspiration technology, and the social research [21].

With the development of computer technologies such as the big data, the artificial intelligence and the machine learning, ways to acquire CRs from a large number of online comments has been facilitated by text mining. Online comments are the evaluation information of the products or services by customers on e-commerce platforms [22]. And most of them are unstructured texts in large volume and complex structure. The steps to acquire CRs from online comments are generally: (1) comment data extraction; (2) data preprocessing such as clean up invalid comments; (3) word segmentation; (4) stop words filtering; (5) feature clustering; (6) sentiment analysis; (7) CRs generating. Liu et al. [23] and Jiang et al. [24] proposed methods to automatically evaluate online comments for the inability to effectively and accurately identify a large number of online comments, and dig out the key CRs from the perspective of designers. Jiang et al. [25] proposed a demand-centered approach for requirements evolution by mining and analyzing online reviews. Peng et al. [26] proposed a method for acquiring CRs based on feature models and collaborative filtering, and in order to improve the accuracy and efficiency of the software requirement acquisition process, designed requirement acquisition algorithms. Li et al. [27] proposed a model for identifying critical customer requirements.

Compared with traditional acquisition methods of CRs, the methods based on online sources have advantages of low cost, wide information, strong timeliness, high reliability, and high customer initiative [28]. At present, methods to acquire CRs based on online comment mining mainly includes the sentiment analysis and opinion mining [29, 30], the requirements feature extraction [31], the data mining [32] and the product attribute extraction [33]. Methods for acquiring CRs can be divided into two categories:

- Extracting product feature words separately, and then to select the sentiment words corresponding to the feature words [34].
- The other is to extract product feature words and sentiment words at the same time [35].

The methods for extracting product feature words include manual labeling methods, extraction from nouns [36], extraction based on dependencies [37] and topic model extraction [38]. Among them, the methods of manual labeling and extracting from nouns are relatively simple, but they are large workload and low efficiency. Although the machine learning improves the efficiency of feature extraction, but they still have a poor attribute classification of technical characteristics of product, and high repetition rate.

2.2 Importance Ranking of Customer Requirements

The acquisition of CRs is the first step of product development, which makes the obtainment of customer perception even more important. In order to calculate importance weights of CRs, there are many methods for the importance ranking of CRs, such as score ranking method [39], analytic hierarchy process (AHP) [40, 41], rough set [42], fuzzy set [43], and entropy method [42]. The grading method and the AHP are importance ranking methods based on subjective scoring, their accuracies are affected by the subjective wishes of the scorers, and they cannot effectively deal with the problem of strong ambiguity in CRs. The AHP method has problems such as large computation and complicated process, when CRs more, the calculation time is longer. Therefore, AHP is combined with other methods (e.g., fuzzy numbers, extent analysis) to improve the efficiency and accuracy of importance ranking [44].

But customers are always ambiguous or contain multiple semantics when expressing their requirements, using the traditional method for calculating the importance weights of CRs, it inevitably appears that, CRs cannot be quantified accurately, and wrong decisions can be made in product design. In order to solve this problem, fuzzy sets are employed to enhance the expression of CRs [45, 46]. The fuzzy sets convert the linguistic assessment of CRs to triangular fuzzy numbers, which are used to build the pairwise comparison matrix of AHP [44, 47]. This approach can improve the imprecise ranking of CRs inherited from studies based on the conventional AHP. But the approach does not consider the influence of customer preference, and most of the membership functions of the fuzzy set are given by experience with obvious subjectivity. Unlike the fuzzy set, the rough set are relatively subjective method as the membership function is determined based on experience. However, both the fuzzy set and the rough set have the disadvantages of difficulty in data collection and heavy workload.

As mentioned above, the analysis of CRs has received much attention from researchers. However, the previous efforts have mostly focused on improving the accuracy of analysis of CRs, ignoring the importance of efficiency of analysis. Most of the traditional methods for acquiring CRs need to survey or interview, which increases the time to collect information of CRs. Although the online comment mining are used to acquire CRs, they ignore the differences in expression of customers. Existing methods for importance ranking of CRs has the disadvantages of difficulty in data collection and heavy workload. Furthermore, CRs is divorced from product design, which reduces the speed of product development.

3 METHODOLOGY

In this section, we introduce the proposed approach, it includes four parts: data collection and pretreatment, acquisition method of CRs, importance ranking method and an integrated system for transforming CRs to technical characteristics of product. The structure of the integrated approach is shown in Figure 1.

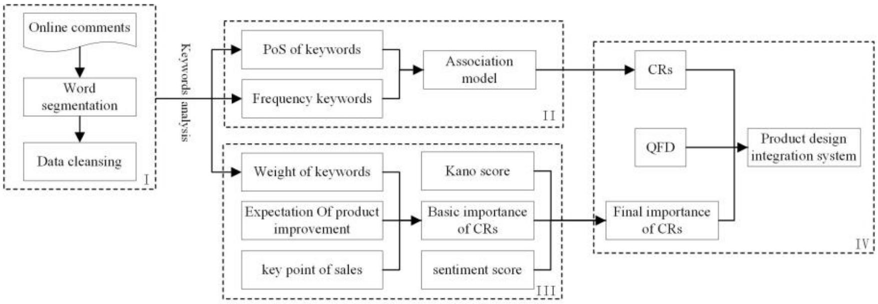

Figure 1: Flow chart of the method.

3.1 Integrated Model for Acquiring Customer Requirements

By analyzing the expression habits of CRs in Chinese online comments, it is found that different customers have different ways to express, such as short sentences, phrases composed of nouns, verbs and adjectives. Therefore, this research proposes a word association model based on word frequency and POS to acquire CRs. In this model, nouns, verbs and adjectives with similar word frequency are formed into a feature item, and then all feature items are clustered to acquire CRs. The acquisition process of CRs is shown in Figure 2.

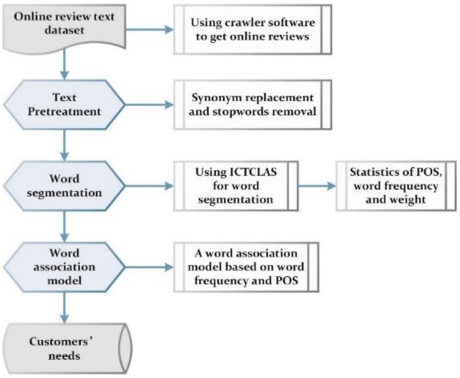

Figure 2: Acquisition process of CRs.

3.1.1 Data collection and pretreatment

(1) Data collection

Online comments from a wide range of sources, including e-commerce platforms, forums and online virtual communities, which store a large number of customer comments data. In this research, the Octoparse [48], a crawler software for data collection from China, was used to crawl customers' comments. Data labels include comment content, comment time, number of likes, number of comments, and additional comment. The Figure 3 shows the data crawling process.

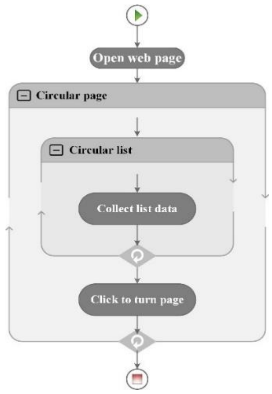

Figure 3: Data crawling process.

(2) Data cleansing

Data cleansing is necessary to maximize the quality and value of the online comments. Data cleansing process includes two specific aspects. The first one is synonym replacement. Due to the diversity of Chinese expressions, synonyms often appear in a comment text. In order to improve the accuracy of CRs' acquisition, synonyms can be filtered through replacement. The second one is stopword removal, which aims to eliminate noise words unrelated to CRs. Some words cannot form the expression of CRs for product attributes, such as "mobile phone", "Huawei", "Apple", and so on. Therefore, stopword removal is carried out to eliminate their interferences on accuracy of CRs' acquisition.

Algorithm 1 shows the main steps of synonym replacement and stopword removal:

Input: comments, synonyms, stopwords

```
1.seg_list = jieba.cut(comments)
2.f = "/".join(seg_list)
3.combine_dict = {}
4.for line in open(synonyms):
5.    seperate_word = line.strip().split(" ")
6.    num = len(seperate_word)
7.    for i in range(1, num):
8.      combine_dict[seperate_word[i]] = seperate_word[1]
9.final_sentence = " "
10.for word in f.split("/"):
11.    if word not in stopwords:
12.       if word in combine_dict:
13.           word = combine_dict[word]
14.           final_sentence += word
15.       else:
16.           final_sentence += word
17.   else:
```

```
    f.split("/").remove(word)
print(final_sentence)
```

(3) Semantic analysis

Semantic analysis is a common method to extract important information from comment content. In this research, ICTCLAS, a Chinese lexical analysis system, is used for semantic analysis. Its principle was Chinese lexical analysis based on the Cascaded Hidden Markov Model (CHMM), which includes five levels: atomic segmentation, simple unknown word recognition, recognition of nested unknown words, class-based hidden horse segmentation and hidden horse tagging of POS [49]. Figure 4 shows the Chinese lexical analysis framework based on CHMM.

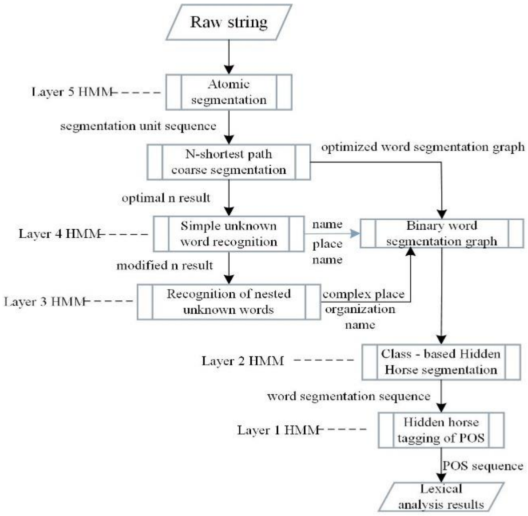

Figure 4: Chinese lexical analysis framework based on CHMM.

ICTCLAS system was used for semantic analysis, and the analysis content included word segmentation, POS tagging, new word recognition, keyword extraction, emotion analysis. Due to keywords with low word frequency are often not representative, this research filters keywords mentioned lowly based on the statistical results of word frequency, and then selects keywords with word frequency greater than 10 according to the POS.

3.1.2 Association model for acquiring customer requirements

The acquisition method of CRs based on online comment mining consists of two parts, the extraction of product attributes and the expression of CRs for product attributes. Existing research on online comments pay more attention to the extraction of product attributes and the analysis of overall sentiment, while there are few researches on the expression of requirements for specific product attributes. Product attributes are usually expressed in two ways in the comment text. The first one is expressed by nouns, such as “system”, “appearance”, “battery”, etc. The second is implicit in the comments, such as “(appearance) nice”, “long (battery) standby time”. In the first category, the extraction of technical characteristics of product is relatively easy, it only is the clustering of all nouns obtained by word segmentation. In the second category, technical characteristics of product are extracted through variety of POS combinations such as “noun + adjective” and “verb + noun”. However, technical characteristics of product extracted by this way are noisy and incomplete.

This research proposed a model to express CRs by combinations of feature words which include noun, verb and adjective. The first one is "noun + verb + adjective", such as "system/n run/v fluent/a"; the second is "noun + adjective", such as "screen/n beautiful/a"; the third is "verb + adjective", such as "charge/v fast /a". These three combinations can not only summarize the majority of product attributes, but also fully reflect customer sentiment expression of product attributes.

Therefore, we define that CRs can be expressed in feature terms of nouns, verbs, and adjectives.

$$T = \{n, v, a\} \tag{1}$$

where T is the feature term; n is a noun; v is a verb; n is an adjective.

Based on the above definition, a word association model can be built to acquire CRs, as shown in Figure 5.

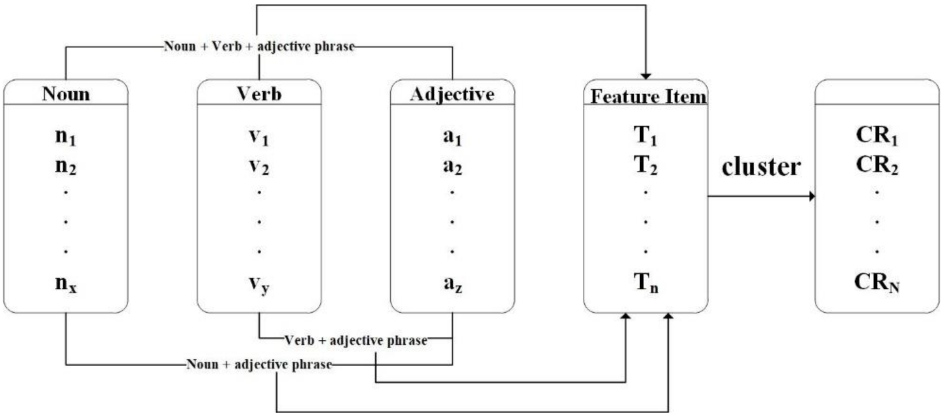

Figure 5: Word association model.

In this model, n_x, v_y and a_z mean that the filtered keywords contain x nouns, y verbs and z adjectives respectively. $\varnothing$ means that the keywords that make up the CRs can default.

Based on the word association model, the specific acquisition process of CRs can be determined:

- Putting the filtered keywords into the corresponding positions according to the part of speech;
- Using the combination of nouns, verbs and adjectives to form feature items, which is based on the fact that keywords with similar word frequency can be combined into one term, and keywords without similar word frequency or with effective collocation can be default;
- Finally, all feature items are clustered to acquire CRs.

3.2 Method for Ranking Importance of Customer Requirements

According to the CRs acquired in section 3.1, this research proposes an importance ranking method combining basic importance and customer preference based on the word weights generated by the ICTCLAS system.

3.2.1 Basic importance of customer requirements

Assume each customer requirement can be expressed by n feature terms. For example, the appearance can be expressed by feature terms such as exquisite workmanship, suitable size, etc.

$$CR_j = \{T_1, T_2, \cdots, T_i, \cdots, T_n\} \tag{2}$$

where CR_j is the j^{th} customer requirement; T_i is the i^{th} feature item which constitutes the customer requirement.

The weight of the i^{th} feature term is determined by the weight of the key words. For example, the weight of the feature "beautiful color" is determined by the weight of "beautiful" and "color".

$$W'_{T_i} = W_n \bullet r_n + W_v \bullet r_v + W_a \bullet r_a \tag{3}$$

where W'_{T_i} is the weight of the feature term T_i; W_n, W_v and W_a represent the weights of nouns, verbs and adjectives respectively. r_n, r_v and r_a are correction coefficients, which are determined by the correlation coefficient between the weight and frequency of nouns, verbs and adjectives.

The importance of initial CRs is obtained:

$$W'_{CR_j} = \sum_{i=1}^{n} W'_{T_i} \tag{4}$$

where W'_{CR_j} is the weight of customer requirement CR_j; $\sum_{i=1}^{n} W'_{T_i}$ is the sum of weights of all feature items in customer requirement CR_j.

Existing importance ranking method ignores the costs and benefits of technical characteristics of product in the future. In order to improve the accuracy and completeness of the calculation results, the proportion of product improvement and key points of sales are integrated to define the basic importance of CRs:

$$W''_{CR_j} = W'_{CR_j} \bullet r \bullet p \tag{5}$$

where W''_{CR_j} is the basic importance of customer requirement CR_j; r is the enterprise's expectation for product improvement, which is determined by market, competitors, capital, etc. The formula for r is:

$$r = \frac{improvement\ plan}{current\ status} \tag{6}$$

The key point of sales p is the extent to which technical measures affect customers purchase, which can be divided into three categories:

- The first one is that customers must want to own and are willing to pay, the weight is 1.5;
- The second one is that customers are interested, but the price should be considered, the weight is 1.2;
- The third one is that the product is not creativity, the weight is 1.

3.2.2 Improved importance of customer requirements

Customer requirement preference is an important factor affecting the effect of product improvement. Therefore, this research proposes a method to measure the degree of customer requirement preference by integrating Kano model and sentiment polarity. Then, uses the degree of customer requirement preference to improve basic importance of CRs to get the final importance. The specific steps are as follows:

Step 1: Identify the Kano category of the customer requirements

For each customer requirement, the kano questionnaire was designed by setting forward and backward questions; conduct kano questionnaire survey, calculate the frequency of various CRs in different requirement types according to the evaluation results, and get the membership degree of each customer requirement to the Kano category. In kano model, O stands for expected requirement, M stands for basic requirement, A stands for attractive requirement, R stands for reverse requirement, I stands for no difference requirement, and Q stands for suspicious results.

Step 2: Determine the satisfaction influence and dissatisfaction influence of the customer requirements

According to the Kano classification and membership degree of CRs, the satisfaction influence and dissatisfaction influence of each customer requirement are analyzed to judge the sensitivity of customers to the improvement of these CRs. The higher sensitivity means that customers have the higher preference degree to the customer requirement, and the improvement to this customer requirement can inspire more customer satisfaction. Where the formula for calculating the satisfaction influence (SI) and dissatisfaction influence (DSI) is:

$$SI = (A+O)/(A+O+M+I) \tag{7}$$

$$DSI = -1 \bullet (O+M)/(A+O+M+I) \tag{8}$$

Step3: Sentiment polarity analysis

This research adopts the method of text sentiment polarity analysis based on sentiment dictionary, and calculates the sentiment score by using negative word dictionary and degree adverb dictionary. In Chinese comments, each sentence is consisted of CRs, negative word, adverb of degree, and sentiment word, such as “the screen is very nice”. Therefore, the sentiment score of each customer requirement is determined by the weight of adverb, negative word and sentiment word score in the sentence:

$$Q_{CR} = \frac{\sum_{k=1}^{K} (-1)^{NOT} \bullet k \bullet P_k \bullet S_k}{K} \tag{9}$$

where Q_{CR} is the sentiment score of customer requirement CR, K is the number of sentence comments, NOT is the number of negative words in the sentence k, P_k is the weight of adverb, and S_k is the score of sentiment word.

Step4: Ranking of importance of customer requirements

Through Kano score and sentiment score to improve the basic importance of CRs to get the final importance:

$$W'''_{CR_j} = W''_{CR_j} \bullet \sqrt{SI^2 + DSI^2} \bullet Q_{CR_j} \tag{10}$$

Then the relative importance of each customer requirement is:

$$W_{CR_j} = \frac{W'''_{CR_j}}{\sum_{j=1}^{N} W'''_{CR_j}} \tag{11}$$

where W_{CR_j} is the relative importance of the customer requirement CR_j; $\sum_{j=1}^{N} W'''_{CR_j}$ is the sum of the importance of all customer requirements.

3.3 An Integrated System for the Transformation of Customer Requirements

QFD is a multi-level deductive analysis method that transforms CRs into design needs, component characteristics, process needs, and production needs [50]. QFD can reduce the time and cost of product design and improve products quality and customer satisfaction [51]. The house of quality (HOQ) is the core of QFD and an important tool to transform CRs to technical characteristics of product [52]. It consists of CRs, technical characteristics of product, importance of CRs and technical characteristics, planning matrix, relationship between CRs and technical characteristics of product, and target value.

This research builds an integrated system to transform CRs to technical characteristics of product by integrating the model of customer requirement acquisition and method of importance ranking to HOQ, as show in Figure 6.

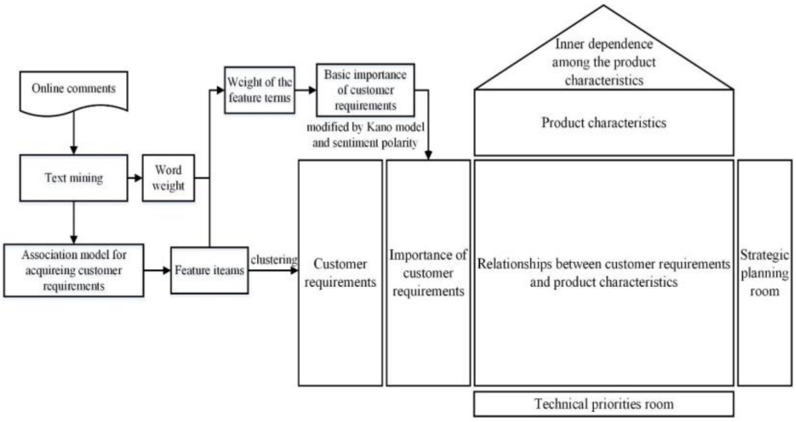

Figure 6: An integrated system for the transformation of customer requirements.

We integrate the method of customer requirement acquisition and importance analysis based on online comment mining with HOQ together to design an integrated system from customer requirement acquisition to transformation. In this system, the word association model is used to acquire the feature items of CRs, the word weight and feature items are used to calculate the basic importance of each customer requirement. Kano model and sentiment polarity are employed to improve the basic importance to get the final importance ranking.

In addition, we can configure the technical characteristics of product accurately and reasonably according to the CRs and its importance, and analyze the inner dependence among the technical characteristics of product to determine the priority of each technical characteristics of product. Competitor analysis can also be implemented to analyze the advantages and disadvantages of products, and set reasonable improvement goals to improve product functions and product competitiveness continuously.

The integrated system is designed to achieve an overall process from CRs to product design, which can guide product developers to redesign product, and improve the integrity and efficiency of product iterative design and innovation. Most of the work in the integrated method is computer-aided, such as acquisition and importance ranking of CRs, technical priorities, etc., which greatly reduces the workload and cost of product design.

The integrated method is a general approach, it works for most product types that can get customers' online comments. Table 1 shows some products that apply to the integrated system.

Product classification	*Products*
Electronic products	mobile phones, computers, TV, etc.
Software products	APPs, computer software products, information systems, etc.
Dress	clothes, shoes, hats, etc.
Furniture products	office furniture, hotel furniture, folk furniture, etc.
Beverage	wine, drinks, etc.
Stationery and sporting goods	paper, pens, balls, etc.
Living goods	personal hygiene products, kitchen supplies, bedding, etc.

Table1 1: Products applicable to the integrated method.

4 CASE STUDY

In this section, a case study is implemented to acquire and prioritize the CRs of smart phones. This case study considers smart phones because there are a large number of online comments in online shopping platforms. In this research, the Octoparse crawler software was used to collect 5,000 customer comments on Huawei P40pro from JD.com. Python was used to preprocess online comment text data. ICTCLAS lexical analysis system was used for word segmentation, POS tagging, word frequency and weight statistics.

4.1 Acquisition of Customer Requirements

4.1.1 Data collection and pretreatment

(1) Data collection

A total of 5000 online comments data of Huawei P40pro were crawled from JD.com by using the Octoparse software. The 5000 online comments as the original corpus for customer requirement acquisition and importance ranking. Table 2 presents a sample data in online comments, it includes customer ID, time of comments made, comment text, and feature items of CRs corresponding to the comment text.

Customer ID	jd_13889***
Time of comments made	2020-11-04 05:58
Comment text	This is a great phone, probably has the best camera on the market right now. Battery life is awesome and it gets charged very fast. And the phone works perfectly. But it is a little less intuitive to set up, possibly due to restrictions. Almost all of the apps I did use on my old phone are available through Huawei. The replacement for Maps, in my opinion, is better.
Feature items of CRs	best camera; batter life awesome; charged fast; work perfectly; little less intuitive set up; apps available; replacement for Maps better

Table 2: Sample data.

As shown in Table 1, there are a lot of feature items of CRs, such as the best camera, charged fast, work perfectly, etc. All feature items are made up of key words with different POS, and they are obtained using Eq. (1). But we found that there are some noise words in the online comments, such as the "Huawei", "phone", and so on, they reduce the quality of comment texts and increase the difficulty of text analysis. In addition, we crawl the latest customer comments according to the comment time, which improves the timeliness and efficiency of product design.

(2) Text pretreatment

The comment text is preprocessed according to Algorithm 1 in section 3.1.1, and the preprocessed comment texts is got after synonym replacing and stop word removal. The pretreatment results of some online comment texts are shown in Table 3.

Raw textual comments	*Pretreatment effect*
After using it for a month, the brand-new machine is different. The original genuine product has perfect screen and perfect camera. Taking photos is amazing, especially intelligent HDR. There is no stuck delay, the signal is good, the battery is durable, the wireless charging attached is really great, charging is fast.	Original genuine product, perfect screen, perfect camera, amazing photo taking, especially intelligent HDR. No stuck delay, good signal, durable battery, great wireless charging, charging fast.

Table 3: Partial text pretreatment results.

(3) Word segmentation

Using the ICTCLAS lexical analysis system to segment the comments text. The word segmentation results are classified into three according to the POS: nouns, verbs, and adjectives. The frequency and weight of each keyword are counted. The keywords with a word frequency less than 10 are filtered. Table 4 shows the partial word segmentation results, it includes the words, POS, word frequency and weight.

Words	*POS*	*Frequency*	*Weight*
screen	n	1385	155.48
handle	n	1091	164.61
fingerprint	n	655	86.44
photograph	v	883	135.12
use	v	597	112.02
charge	v	394	70.47
pretty good	a	1677	29.15
smooth	a	1109	37.06
nice	a	781	19.19

Table 4: Partial word segmentation results and word frequency and weight.

The correlation coefficient between the weight and frequency of each POS are calculated, and the results are shown in Table 5. The results show that the correlation coefficient between the weight and frequency of each POS is greater than 0.7, which has a significant correlation.

POS	*Noun*	*Verb*	*Adjective*
Correlation coefficient	0.87	0.77	0.72

Table 5: Correlation coefficient table of each POS.

4.1.2 Customer requirements

The filtered keywords are classified into the word association model in section 3.1.2 according to their POS to acquire the initial secondary CRs. Then, the primary CRs and their feature items are acquired by using Eq. (1-2), the results are shown in Table 6. In the end, there are a total of 12 first-level CRs and a total of 78 feature items, which included 57 positive sentiment feature items and 21 negative sentiment feature items.

No.	*Customer requirements*	*Feature items*	
		Positive affective	*Negative affect*
1	screen	high resolution screen, smooth screen, bright screen, clear picture quality, clear display	poor picture quality
2	appearance	Beautiful color, curved surface, unique appearance, light and thin, young, comfortable feel, exquisite workmanship, appropriate size, gorgeous color, fashion, beautiful back cover	earthy appearance, heavy color, protruding camera
3	unlocking speed	quick face recognition, quick fingerprint recognition, sensitive reaction	slow response
4	running speed	system runs smoothly, starts up quickly, game is fluent, processor speed is fast, system updates quickly, no jam, software	slow start-up, serious system heating, slow heat dissipation

		runs smoothly	
5	battery life	good battery life, fast charging	slow charging, poor battery life, poor standby
6	photo effect	effect of taking photos is clear, photos are real, resolution is good, night scenery is good	slow speed, poor resolution
7	functionality	powerful, full-featured, waterproof, NFC, humanized, good optimization, practical function, fast software installation, good configuration	complicated functions
8	individualization	novel design, high-end technology, intelligence, personality	no innovation
9	sound quality	clear sound quality, clear voice, clear call, HiFi	poor sound quality, poor communication
10	quality of network communication	fast internet speed, fast download speed, good signal	bad signal
11	operation difficulty	easy to operate, comfortable to use	complex operation
12	storage	enough storage, large storage	not enough storage

Table 6: Customer requirements.

As show in Table 6, the customers have stronger positive sentiments than negative sentiments for this mobile phone. Most customers believe that the mobile phone can meet most of their requirements. And they have a high degree of satisfaction to the mobile phone. During the clustering of requirement feature items, it is found that customers' focus points for the product are basically similar, such as appearance, hardware and software configuration, sense of technology, personalization, functionality, and so on. In addition, customers' comments on the product tend to be positive, with fewer negative comments. Most of the negative comments have individual differences, with fewer common negative comments. Therefore, this research believes that enterprise should fully understand and subdivide customer groups when acquiring CRs. And the potential requirements that are difficult to acquire can be deeply explored by cooperation with customers, so as to continuously improve product functions and quality.

4.2 Importance Ranking of Customer Requirements

We calculated the basic importance of each customers' requirement using Eq. (3-6). And then the Kano score and the sentiment score are calculated using Eq. (7-9). Finally, we calculated the final importance using Eq. (10-11) to obtain the ranking of each customers' requirement (Table 7). It can be seen that the most important customer requirement is running speed, and its relative importance is 0.31, which is significantly higher than other customer requirements. The customers are most concerned about whether the mobile phone meet their own requirements, and has a strong preference for running speed. On the other hand, the relative importance of appearance, photo effect and functionality are all greater than 0.1, which are the requirements that customers are more concerned about. The results accord with the current trend of CRs diversification and fashion.

Customer requirements	*Preference degree*		*Basic importance*		*Final importance*		
	Kano score	*Sentiment score*	*Importance*	*Rank*	*Importance*	*Relative importance*	*Rank*
running speed	0.85	1.60	1387.64	1	1881.75	0.31	1
appearance	0.80	1.80	634.07	4	917.69	0.15	2
photo effect	0.83	1.23	745.24	2	759.87	0.12	3

functionality	0.76	1.70	509.44	5	658.09	0.11	4
screen	0.61	1.31	659.38	3	531.93	0.09	5
battery life	0.80	1.14	397.24	6	359.99	0.06	6
unlocking speed	0.64	1.80	227.89	8	260.71	0.04	7
network and communication quality	0.79	0.90	360.43	7	256.90	0.04	8
operation difficulty	0.64	1.40	207.71	9	185.10	0.03	9
sound quality	0.88	1.23	159.81	10	171.56	0.03	10
individualization	0.82	1.05	149.16	11	128.19	0.02	11
storage	0.67	1.40	48.49	12	45.78	0.01	12

Table 7: Importance ranking of customer requirements.

Through the modification of customer preference, the importance of appearance, functionality and unlocking speed has risen. Although the frequency of these three requirements in customer comments is low, but customers have a higher degree of preference for these three demands, with high Kano scores and sentiment scores. The importance of photo effect, screen, quality of network and communication has declined, although the frequency of these three requirements in customer comments is high, but they have lower customer preference, and the sentiment score is significantly lower than other customer requirements.

The customers have a low preference for operation difficulty, which means the operation interface, operation system, and operation methods of this mobile phone have relatively low impact on customer satisfaction. Therefore, the enterprise should pay more attention to the improvement of running speed, emphasize the advantages of running speed, and make differentiated investments in product functions when improving product and marketing.

4.3 Transformation of Customer Requirements to Technical Characteristics

According to the design method of QFD, the 12 CRs and their importance are integrated into the HOQ using the integrated system in section 3.3. The technical characteristic indicators are set according to the relationship between CRs and technical characteristics of product. The technical characteristics of the mobile phone include operating system, processor, screen resolution, battery, pixel, size, material, network, weight, shape, fingerprint and face recognition technology. And then the roof of the HOQ is built by analyzing the relationship between the various technical characteristics. Finally, a competitor analysis is completed to clarify the advantages and disadvantages of the product. The specific results of the QFD construction are shown in Figure 7.

Figure 7 shows that the processor is the most important technical characteristics of all the 12 technical characteristics. It has a strong relationship with running speed, functionality and unlocking speed. In addition, the processor is the importance technical characteristics that affects CRs such as running speed, functionality, unlocking speed, and so on. Therefore, the processor is a key points of product improvement in the future.

The operating system has a high degree of relative importance. It is an important foundation for supporting functionality, unlocking speed, difficulty of operation and sound quality. The importance weight of other technical characteristics is not much different, and it does not have a universally significant influence among CRs.

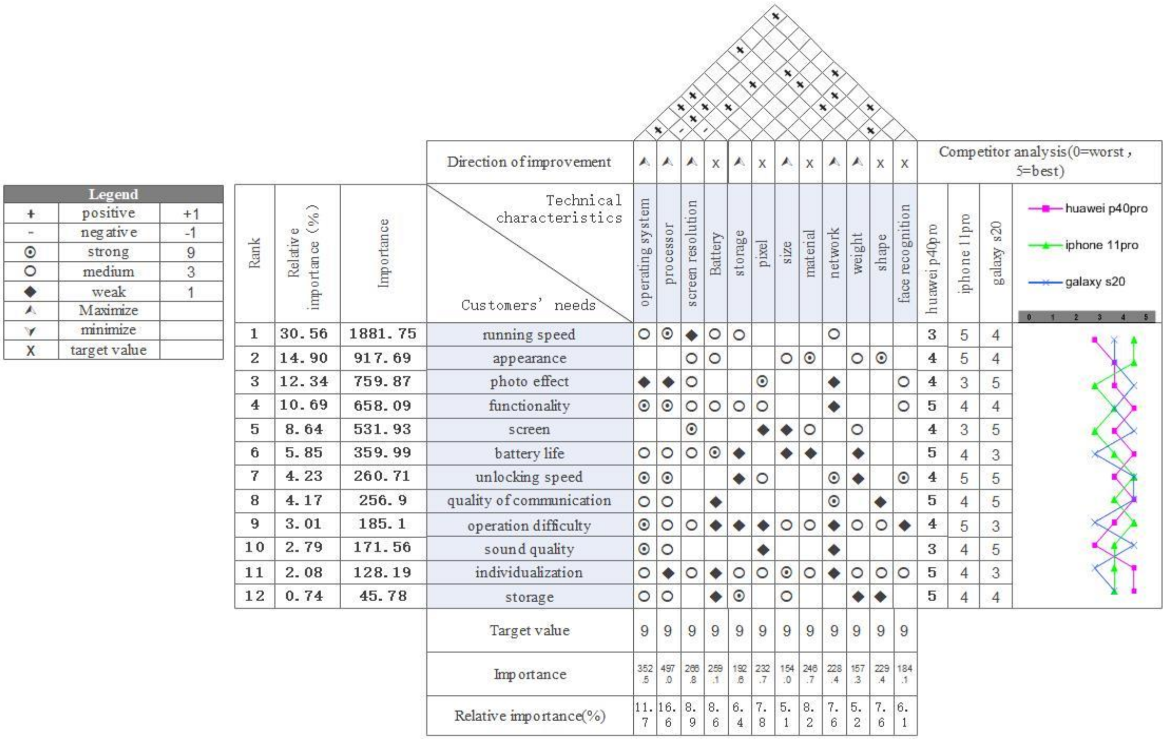

Figure 7: House of quality.

The results of competitors analysis shows that the product has advantages among competitors in follow CRs: functionality, endurance, network and communication quality, personalization, and storage. Compared with competitors, the product has a higher degree of satisfaction to the above-mentioned CRs. Especially in the industry-leading 5G technology, which can be used as the future marketing focus of this product. The main competitive disadvantages of this product include: running speed, face value, photo effect, screen, unlocking speed, difficulty of operation and sound quality, which requirements should be further strengthened in future.

The results of correlation analysis between the various technical characteristics shows that there is a negative correlation between processor and screen resolution, and higher screen resolution requires relatively higher processor performance. The battery and screen resolution has a negative correlation. High screen resolution consumes more battery power. Using a higher screen resolution will reduce battery usage and endurance. On the other hand, operating system has a positive correlation with processor, battery, memory, network communication technology, fingerprint and face recognition technology, and it is important to meet product framework and CRs.

5 DISCUSSIONS

On the basis of case analysis, the effectiveness and superiority of the proposed method are discussed.

In this research, a computer-aided approach for acquisition and importance ranking of CRs from the online comment mining is proposed, and an integrated system for the transformation of CRs is built. The proposed approach could help product development departments and marketing departments to know in advance the customer requirement state evolution trends, and satisfy the requirement. In this way, right resource could be put into the right CRs at the right time. The validation of the proposed method in the requirement analysis of mobile phone shows that it is a more effective tool. In addition, the method we proposed in this paper is suitable for most products that can get customers' online comments, with the goal of improving existing products and discovering new customer requirements. For the completely novel product design, we can use the method to find potential customer requirements by analyzing historical comments data. And then through the integrated system to determine the key technical characteristics meet the potential customer requirements. Based on this, we can quickly develop a usable product prototype, and then quickly enter the iterative update of products by continuously collecting and analyzing customers' comments, so as to implement a completely novel product design.

Compared to the models of Zhang et al. [14] and He et al. [53], this model is more in line with the language expression habits of customers in Chinese online comments. It can effectively distinguish and identify positive CRs and negative CRs in online comments, which improves the accuracy and comprehensiveness of customer requirement acquisition. And the integrated method based on QFD can continuously transform CRs into technical characteristics of products, and provide a customer-driven systematic design tool for enterprise to determine the key points of product iterative improvement.

The importance ranking of CRs is an important indicator for enterprise to determine the priority of technical characteristics of product. Existing methods for importance ranking of CRs are mainly based on traditional methods such as Kano model and AHP, which always inefficient and inaccurate. Online comments are often used to acquire CRs and analysis customers' sentiment, but there are few researches use the word frequency and weight of keywords in online comments for importance ranking. In addition, existing research mainly consider factors such as customer desire for requirements and requirements satisfaction, but rarely consider the influence of customer requirement preference, key points of enterprise sales, and the proportion of technological improvement. This research designs a method for importance ranking of CRs, which based on the frequency and weight of the keywords that make up each feature item in the online comments. And then we integrate the key points of sales and proportion of technology improvement into the importance ranking method, and uses the Kano score and sentiment score to improve the efficiency and accuracy of importance ranking.

To sum up, the approach reveals the following strengths:

- The proposed computer-aided approach for acquiring CRs solves the problems of differentiation in expression of CRs. It considers the customers' expression habit, and therefore, provides a combination mode suitable for different expression habits.
- The proposed computer-aided approach provides more reasonable and reliable importance ranking of CRs because it takes the product improvement ratio, key sales points, Kano model and sentiment polarity into account to improve the efficiency and accuracy of the importance ranking.
- An integrated system is built to improve product innovation performance and product development efficiency. With the above integrated system, enterprise can clarify their own advantages and disadvantages, and improve the product continuously.

However, this research also has certain limitations, which needs further research. First of all, in the section of data cleansing, a dictionary-based method is used for synonym replacement and stop word removal. This method is limited by the accuracy and comprehensiveness of the dictionary, which leads to the decline of the data cleansing effect. Building exclusive thesaurus and stop words dictionary for different products will be a potential solution. In the future research, we will develop a decision support system that integrates customer requirement acquisition model, importance ranking method and QFD, so that designers can quickly carry out product iterative design and improve the automation level of related activities.

6 CONCLUSIONS

Customer requirement acquisition and importance ranking is an important part of customer requirements analysis, and it is a crucial factor for enterprise to design and upgrade products. However, the existing analysis method of customer requirements is inefficient and does not form a complete system from customer requirements to product design. In order to tackle the problem, this research proposes a computer-aided approach for acquiring customer requirements and ranking their importance based on online comment mining, and taking into account the proportion of product improvement, key points of sales, and customer requirement preference to improve the efficiency and accuracy of the importance ranking. Furthermore, the customer requirement acquisition model and importance ranking method are integrated into the QFD to build a integrated system to improve product innovation performance and product development efficiency. With the above integrated system, enterprise can clarify their own advantages and disadvantages, and improve the product continuously.

Li Xiangdong, https://orcid.org/0000-0002-0159-9866
He Youlong, https://orcid.org/0000-0003-4613-6045
Qingjin Peng, https://orcid.org/0000-0002-9664-5326
Liu Wei, https://orcid.org/0000-0001-6111-4149

Acknowledgements
Funded by Chinese Industrial Internet innovation and development project (TC19083W8)

REFERENCES

[1] Dou, R.; Li, W.; Nan, G.: An integrated approach for dynamic customer requirement identification for product development, Enterprise Information Systems, 13(4), 2018, 1-19. https://doi.org/10.1080/17517575.2018.1526321
[2] Yang, Q.; Jiao, H.; Song, F.; Pan, G.; Wei, D.: Customer requirement acquisition system and requirement expression guidance based on ant colony optimization, Advances in Mechanical Engineering, 9(6), 2017, 1-9. https://doi.org/10.1177/1687814017704412
[3] Song, W.; Ming, X.; Xu, Z.: Integrating Kano model and Grey–Markov chain to predict customer requirement states, Journal of Engineering Manufacture, 227(8), 2013, 1232-1244. https://doi.org/10.1177/0954405413485365
[4] Jiao, J. R.; Chen, C. H.: Customer requirement management in product development: A review of research issues, Concurrent Engineering, 14(3), 2006, 173-185. https://doi.org/10.1177/1063293X06068357
[5] Hu, F.: The relationship analysis between online reviews and online shopping based on B2C platform technology, Cluster Computing, 22, 2018, 3365–3373. https://doi.org/10.1007/s10586-018-2182-3
[6] Changchit, C.; Klaus, T.; Lonkani, R.: Online reviews: What drives consumers to use them, Journal of Computer Information Systems, 2020, 1-10. https://doi.org/10.1080/08874417.2020.1779149
[7] Jo, Y.; Oh, A. H.: Aspect and sentiment unification model for online comment analysis, In Proceedings of the fourth ACM international conference on Web search and data mining, Hong Kong, 2011, 815-824. https://doi.org/10.1145/1935826.1935932
[8] Li, M.; Huang, L.; Tan, C.; Wei, K.: Helpfulness of online product comments as seen by consumers: Source and content features, International Journal of Electronic Commerce, 17(4), 2013, 101-136. https://doi.org/10.2753/JEC1086-4415170404
[9] Weathers, D.; Swain, S. D.; Grover, V.: Can online product comments be more helpful? Examining characteristics of information content by product type, Decision Support Systems, 79, 2015, 12-23. https://doi.org/10.1016/j.dss.2015.07.009

[10] Chen, H.; Yan, Q.; Xie, M.; Zhang, D.; Chen, Y.: The sequence effect of supplementary online comments in book sales, IEEE Access, 7, 2019, 115650-115658. https://doi.org/10.1109/ACCESS.2019.2948190
[11] Chen, S.; Lv, X.; Gou J.: Personalized recommendation model: An online comment sentiment-based analysis, International Journal of Computers, Communications & Control (IJCCC), 15(1), 2020, 1-14. https://doi.org/10.15837/ijccc.2020.1.3764
[12] Zhao, X.; Huang, C.; Pan, H.: Online comments of multi-category commodities based on emotional tendency analysis, Cluster Computing, 22(3), 2019, 6345-6357. https://doi.org/10.1007/s10586-018-2086-2
[13] Liu, H.; Cui, T.; He, M.: Product optimization design based on online comment and orthogonal experiment under the background of big data, In Archive Proceedings of the Institution of Mechanical Engineers Part E Journal of Process Mechanical Engineering, 2020. https://doi.org/10.1177/0954408920943690
[14] Zhang, L.; Chu, X.; Xue, D.: Identification of the to-be-improved product features based on online comments for product redesign, International Journal of Production Research, 57(7-8), 2019, 2464-2479. https://doi.org/10.1080/00207543.2018.1521019
[15] Dellarocas, C.; Zhang, X.; Awad, N.: Exploring the value of online product comments in forecasting sales: The case of motion pictures, Journal of Interactive Marketing, 21(4), 2007, 23-45. https://doi.org/10.1002/dir.20087
[16] Franke, N.; Piller, F. T.: Key research issues in user interaction with user toolkits in a mass customisation system, International Journal of Technology Management, 26(5-6), 2003,578-599. https://doi.org/10.1504/IJTM.2003.003424
[17] Li, N.; Jin, X.; Li, Y.: Identification of key customer requirements based on online reviews, Journal of Intelligent and Fuzzy Systems, 39(1984), 2020, 1-14. https://doi.org/10.3233/JIFS-200057
[18] Qi, J.; Zhang, Z.; Jeon, S.; Zhou, Y.: Mining customer requirements from online reviews: A product improvement perspective, Social Science Electronic Publishing, 53(8), 2016, 951-963. https://doi.org/10.1016/j.im.2016.06.002
[19] Akao, Y.: Quality function deployment: Integrating customer requirements into product design, Productivity Press, New York, NY, 1990.
[20] Davis, A.; Dieste, O.; Hickey, A.; Juristo, N.; Moreno, A. M.: Effectiveness of requirements elicitation techniques: Empirical results derived from a systematic review, 14th IEEE International Requirements Engineering Conference (RE'06), Minneapolis/St. Paul, MN, USA, 2006, 179-188. https://doi.org/10.1109/RE.2006.17
[21] Abbasi, M. A.; Jabeen, J.; Hafeez, Y.; Batool, D.; Fareen, N.: Assessment of requirement elicitation tools and techniques by various parameters, Software Engineering, 3(2), 2015, 7-11. https://doi.org/10.11648/j.se.20150302.11
[22] Mudambi, S. M.; Schuff, D.: What makes a helpful online comment? A study of customer comments on Amazon.com, MIS Quarterly, 34(1), 2010, 185-200. https://doi.org/10.1007/s10107-008-0244-7
[23] Liu, Y.; Jin, J.; Ji, P.; Harding, J. A.; Fung, R. Y.: Identifying helpful online comments: A product designer's perspective, Computer-Aided Design, 45(2), 2013, 180-194. https://doi.org/10.1016/j.cad.2012.07.008
[24] Jiang, W., Zhang, L.; Dai, Y.; Jiang, J.; Wang, G.: Analyzing helpfulness of online comments for user requirements elicitation, Chinese Journal of Computers, 36(1), 2013, 120-131. https://doi.org/10.3724/SP.J.1016.2013.00119
[25] Jiang, W.; Ruan, H. B.; Zhang, L.: Analysis of economic impact of online reviews: An approach for market-driven requirements evolution, Communications in Computer & Information Science, 432, 2014, 45-59. https://doi.org/10.1007/978-3-662-43610-3_4
[26] Peng, Z.; Wang, J.; He, K.; Dong, M.: A requirements elicitation approach based on feature model and collaborative filtering, Journal of Computer Research and Development, 53(9), 2016, 2055-2066. https://doi.org/10.7544/issn1000-1239.2016.20150426

[27] Li, N.; Jin, X.; Li, Y.: Identification of key customer requirements based on online reviews, Journal of Intelligent and Fuzzy Systems, 39(1984), 2020, 1-14. https://doi.org/10.3233/JIFS-200057
[28] Yu, H.; Shi, M.; Xiao, X.: Study on the VOC analyzing methods based on concentrated web data, 2008 International Symposium on Knowledge Acquisition and Modeling, Wuhan, China, 2008, 23-28. https://doi.org/10.1109/KAM.2008.20
[29] Zheng, L.; Wang, H.; Gao, S.: Sentimental feature selection for sentiment analysis of Chinese online reviews, International Journal of Machine Learning and Cybernetics, 6(1), 2018, 75-84. https://doi.org/10.1007/s13042-015-0347-4
[30] Hemmatian, F.; Sohrabi, M. K.: A survey on classification techniques for opinion mining and sentiment analysis, Artificial Intelligence, 52, 2019, 1495–1545. https://doi.org/10.1007/s10462-017-9599-6
[31] Bakar, N. H.; Kasirun, Z. M.; Salleh, N.: Feature extraction approaches from natural language requirements for reuse in software product lines: A systematic literature review, Journal of Systems and Software, 106, 2015, 132-149. https://doi.org/10.1016/j.jss.2015.05.006
[32] Soni, A.; Saxena, A.; Bajaj, P.: A methodological approach for mining the user requirements using apriori algorithm, Journal of Cases on Information Technology (JCIT), 22(4), 2020, 1-30. https://doi.org/10.4018/JCIT.2020100101
[33] Shen, Y.; Chen, G.: Study of feature word extraction and cluster in Chinese product reviews, 2016 IEEE International Conference on Network Infrastructure and Digital Content (IC-NIDC), Beijing, 2016, 447-451. https://doi.org/10.1109/ICNIDC.2016.7974614
[34] Ma, B.; Zhang, D.; Yan, Z.; Kim, T.: An lda and synonym lexicon based approach to product feature extraction from online consumer product reviews, Journal of Electronic Commerce Research, 14(4), 2013, 304-314. https://doi.org/10.3846/16111699.2012.665383
[35] Chen, G.; Wan, Y.; Xu, X.: An analysis of the sales and consumer preferences of e-cigarettes based on text mining of online reviews, 2016 3rd International Conference on Systems and Informatics (ICSAI), Shanghai, China, 2016, 1045-1049. https://doi.org/10.1109/ICSAI.2016.7811105
[36] Blei, D. M.; Jordan, M. I.: Modeling annotated data, In Proceedings of the 26th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, New York, NY, 2003, 127-134. https://doi.org/10.1145/860435.860460
[37] Hu, M.; Liu, B.: Mining and summarizing customer reviews, In Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Washington, 2004, 168-177. https://doi.org/10.1145/1014052.1014073
[38] Kim, S.; Zhang, J.; Chen, Z.; Oh, A.; Liu, S.: A hierarchical aspect-sentiment model for online comments, In Proceedings of the 27th AAAI Conference on Artificial Intelligence, Palo Alto, 2013, 526-533. https://dl.acm.org/doi/10.5555/2891460.2891533
[39] Griffin, A., Hauser, J.: The voice of the customer, Marketing Science, 12(1), 1993, 1-27. https://doi.org/10.1287/mksc.12.1.1
[40] Saaty, T. L.: How to make a decision: The analytic hierarchy process, European journal of operational research, 48(1), 1990, 9-26. https://doi.org/10.1016/0377-2217(90)90057-i
[41] Kwong, C. K.; Bai, H.: A fuzzy AHP approach to the determination of importance weights of customer requirements in quality function deployment, Journal of Intelligent Manufacturing, 13, 2002, 367–377. https://doi.org/10.1023/A:1019984626631
[42] Roy, J.; Adhikary, K.: A weighted interval rough number based method to determine relative importance ratings of customer requirements in QFD product planning, Journal of Intelligent Manufacturing, 30(1), 2019, 3-16. https://doi.org/10.1007/s10845-016-1224-z
[43] Chan, L. K.; Kao, H. P.; Wu, M. L.: Rating the importance of customer requirements in quality function deployment by fuzzy and entropy methods, International Journal of Production Research, 37(11), 1999, 2499-2518. https://doi.org/10.1080/002075499190635

[44] Kwong, C. K.; Bai, H.: Determining the importance weights for the customer requirements in QFD using a fuzzy AHP with an extent analysis approach, IIE Transactions, 35(7), 2003, 619-626. https://doi.org/10.1080/07408170304355
[45] Fung, R. Y.; Popplewell, K.; Xie, J.: An intelligent hybrid system for customer requirements analysis and product attribute targets determination, International Journal of Production Research, 36(1), 1998, 13-34. https://doi.org/10.1080/002075498193912
[46] Zheng, P.; Xu, X.; Xie, S.: A weighted interval rough number based method to determine relative importance ratings of customer requirements in QFD product planning, Journal of Intelligent Manufacturing, 30, 2019, 3–16. https://doi.org/10.1007/s10845-016-1224-z
[47] Guo, C.; Liu, Y.; Hou, S.; Wang, W.: Innovative product design based on customer requirement weight calculation model, International Journal of Automation and Computing, 7, 2010, 578–583. https://doi.org/10.1007/s11633-010-0543-3
[48] Octoparse, https://www.octoparse.com/, Octopus Data Inc.
[49] Zhang, H.; Liu, Q.; Cheng, X.; Zhang, H.; Yu, H.: Chinese lexical analysis using hierarchical hidden Markov model, In Proceedings of the second SIGHAN workshop on Chinese language processing - Volume 17 (SIGHAN' 03). Association for Computational Linguistics, USA, 2003, 63–70. https://doi.org/10.3115/1119250.1119259
[50] Chan, L. K.; Wu, M. L.: Quality function deployment: A literature review, European Journal of Operational Research, 143(3), 2002, 463-497. https://doi.org/10.1016/S0377-2217(02)00178-9
[51] King, B.: Listening to the voice of the customer: Using the quality function deployment system, National Productivity Review, 6(3), 2010, 277-281. https://doi.org/10.1002/npr.4040060312
[52] Hauser, J. R.; Clausing, D.: The house of quality, Harvard Business Review, 66(3), 1988, 63-73. https://hbr.org/1988/05/the-house-of-quality
[53] He, L.; Wu, Z.; Xiang, W.; Goh, M.; Xu, Z.; Song, W.; Ming, X.; Wu X.: A novel Kano-QFD-DEMATEL approach to optimise the risk resilience solution for sustainable supply chain, International Journal of Production Research, 2020. https://doi.org/10.1080/00207543.2020.1724343