Customer Requirements Mining and Product Design Analysis Based on E-commerce Comment Data

Pei-yu Shi(✉) and Jun-he Yu

Mechanical Engineering and Mechanics, Ningbo University, Ningbo, Zhejiang, China
xun125286@sina.com, yujunhe@nbu.edu.cn

Abstract. Accurate mining of customer requirements (CR) can better improve the design of e-commerce products, and the Kano model is an approach to quantify CR satisfaction. Yet related parameters are mainly obtained by questionnaires. This study preliminarily extracts CR from online comment data by keyword frequency and then develops a Kano model-based method to quantitatively evaluate the degree to which CR affects the improvement of product attributes to obtain indices that scientifically contribute to the design and upgrade of corporate products. By applying online comment data mining, this method makes up defects such as biased information and expensive labor and time costs of questionnaires, and expands the possibility for enterprises to engage in e-commerce in the digital economy era. CR for air conditioners varies a lot, so taking air conditioners as an example, this study verifies the feasibility of replacing questionnaire surveys with data-driven requirements information mining.

Keywords: Requirements Mining · Online Comments · Kano Model · Customer Satisfaction

1 Introduction

The fiercer enterprise-level vying in the demand market has spurred scholars and designers to identify the customer requirement (CR) and its application in product design. It is traditionally common to obtain CR through customer interviews. But given more diversified CR coupled with shorter product updates caused by distinctive time-varying features, companies need faster and more accurate identification means. Timoshenko et al. identified CR from customer-generated content to offer decision-making assistance in product design [4]. Hu et al. developed an iterative, innovative product design method from dynamic CR [1]. Thus, CR can be applied to production practice, and the larger the amount of information and the stronger the timeliness, the greater the assistance to product design. That is why research is shunting towards quantitative analysis of CR. Peng et al. put forward a method to evaluate the quality of user-based product design plans [3]. After building a customer satisfaction model based on comment data, Yan obtained satisfaction indicators by measuring keywords and calculated consumer satisfaction values combined with emotional analysis [6].

Z. Zeng et al. (Eds.): ECIT 2022, AHE 11, pp. 14–25, 2023.
https://doi.org/10.2991/978-94-6463-005-3_3

Structured CR identification methods and technologies are required to better align CR information from massive data with production practices. To obtain customer satisfaction, Zhou combined the Kano model and function fitting and finally developed a customer satisfaction model for customized products [9]. Zhang et al. proposed integrating fuzzy clustering analysis into the Kano model to obtain the comprehensive weight of CR [7]. It is not difficult to conclude that the Kano model can quantitatively analyze CR. Lina et al. quantitatively analyzed the nonlinear relationship between consumer satisfaction (CS) and CR achievement [2]. Zhao et al. applied the said quantitative method to integrate CR into product design and verified its applicability by taking the requirement for configuring the auto parts production line as an example [8]. However, the implementation of the above method often uses questionnaires to obtain the data needed for Kano classification and quantitative evaluation of CR. Yet it requires tremendous time and human costs and cannot ensure the universality of the collected samples. This study resolves the universality of samples with comment data and obtains data needed for quantitative evaluation combined with natural language processing technology.

2 Comment Data-Based CR Identification Framework

As shown in the figure, the study process consists of three parts: data acquisition, data processing and data application. To break through limitations and address the high cost of questionnaire surveys in obtaining data, online comment text data was used as the object of this study in the data acquisition stage. So, in the data processing stage, unstructured and colloquial problems of online comment data need to be solved. Combining Bert preprocessing language model with aspect-based sentiment analysis can summarize consumer sentiment tendency according to qualities of different requirements in consumer comments. At the data application level, the Kano model was used to intuitively express the relationship between the extent to which products affect CR and consumer experience upon quantitative analysis of the functional relationship between requirement achievement and consumer satisfaction. Then we analyzed the current market situation of product requirements with the concept of priority in the Kano model.

A method of obtaining CR for Kano classification was presented to gain relevant parameters of Kano CR classification [5]. Among them, requirement intensity refers to the intensity of consumers for a certain requirement and can be mined by extracting sentiment words from comment sentences; requirement percentage means the frequency of mention of a certain requirement in a comment set. Comment data can be employed to determine two evaluation indicators: CR proportion p_i and CR intensity f_i. For example, of all comments on the same product in a period, if ten consumers say that "this product performs well in intelligent control," then the sentimental words used and the number of users can be used as the quantitative basis. Based on the Kano model and combined with these two indicators, the Must-be Quality (M), One-dimensional Quality (O) and Attractive Quality (A) in the Kano model are redefined, as shown in Fig. 2.

(1) Must-be Attributes (M): Most consumers mention must-be attributes in comments (i.e., a high proportion of CR), and they dominate the product comment group (i.e., high CR intensity f_i), so manufacturers produce goods with must-be quality up to CR.

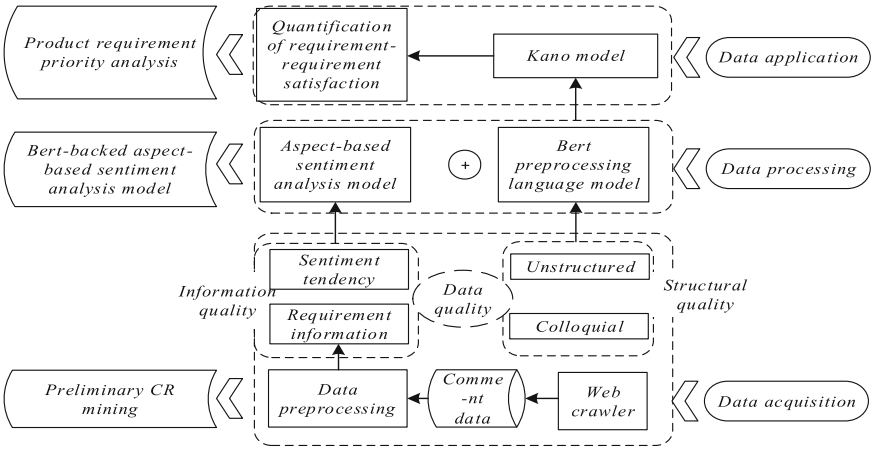

Fig. 1. Comment Data-based Requirement Information Mining Method.

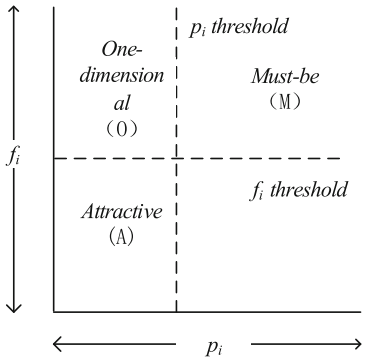

Fig. 2. Sentiment Analysis-based CR Kano Classification.

(2) One-dimensional Attributes (O): Not all consumers care for one-dimensional attributes, so some may not mention them in comments (i.e., low proportion of CR), but they still dominate the product comment group as an important part of products (i.e., high CR intensity f_i).
(3) Attractive Attributes (A): Unless attractive attributes perform beyond expectations, otherwise consumers may not mention them in comments (i.e., low proportion of CR) and not all products have them, so they do not dominate the product comment group as an important part of products (i.e., low CR intensity f_i).

To apply the above method to data processing quickly and conveniently, this paper proposed to analyze the weights of words describing product attributes in comment sentences with TF-IDF which can express the emphasis of words in documents. The

average weighted value of TF-IDF of requirement i in the text j is used to quantify requirement intensity, as shown in Formula 1–4.

$$tf_{i,j} = \frac{n_{i,j}}{\sum_k n_{k,j}} \tag{1}$$

$$idf_{i,j} = \lg \frac{|n|}{\left|\{j : t_i \in d_j\}\right|} \tag{2}$$

$$tfidf_{i,j} = tf_{i,j} \times idf_i \tag{3}$$

$$f_i = \overline{tfidf_{i,j}} \tag{4}$$

where, $n_{i,j}$ represents the frequency of given quality i in the document j; $n_{k,j}$ represents the frequency of all words in the document j; $tf_{i,j}$ represents the frequency of quality i in the document j; idf_i represents the inverse document frequency; and f_i represents the average value of requirement intensity at the level of quality i.

In addition, the threshold of CR ratio represents the proportion of product attributes widely mentioned in the product comment group, and that of CR intensity f_i represents the proportion of product attributes dominated in the product comment group. Both thresholds depend on the target products concerned, market characteristics and other factors.

3 Kano Model-Based CR Quantitative Analysis

3.1 Customer Satisfaction (CS) Coefficient and Customer Dissatisfaction (DS) Coefficient

CS coefficient and DS coefficient refer to customer satisfaction and dissatisfaction, which measure the percentage of CR satisfied, respectively. Assuming that a comment is representative of a customer, CS value can be calculated by dividing the sum of comments with satisfaction attributes (attractive attributes f_A and one-dimensional attributes f_O) by the sum of comments with attractive attributes f_A, one-dimensional attributes f_O, must-be attributes f_M and indifferent attributes f_I, as shown in Formula 5.

Similarly, the DS value can be calculated by dividing the sum of comments with dissatisfaction attributes (must-be attributes f_M and one-dimensional attributes f_O) by the sum of comments with f_A, f_O, f_M and f_I, as shown in Formula 6.

$$CS = \frac{(f_A + f_O)}{(f_A + f_O + f_M + f_1)} \tag{5}$$

$$CS = \frac{(f_O + f_M)}{(f_A + f_O + f_M + f_I)} \tag{6}$$

where, f_A, f_O, f_M and f_I represent the number of attractive attributes, one-dimensional attributes, must-be attributes and indifferent attributes, respectively.

To define CS and DS and their quantitative achievement of CR, we assume CS as 1 if product attributes fully meet CR and 0 if product attributes fail to fully meet CR.

According to this assumption, CS and DS points are defined. The CS_i point of CR_i is $(1, CS_i)$, so its achievement level is 1. The DS_i of CR_i is $(0, DS_i)$, so its achievement level is 1. Integrating them into the Kano model can accurately quantify the relationship between product attributes and customer satisfaction at the CR level.

3.2 Curve of Relationship between Product Attribute Achievement and Customer Satisfaction

After CS and DS points are determined, the curve of the relationship between product attribute achievement and customer satisfaction at the CR level can be obtained. Among them, variable x represents the achievement of product attributes at the CR level, ranging from 0 to 1 and variable y represents customer satisfaction, ranging from -1 to 1, as shown in Fig. 3.

The relationship between consumer satisfaction S and CR achievement CR can be roughly quantified by appropriate functions. Generally, the S-CR relational function can be expressed as $S = f(x, a, b)$, of which S represents customer satisfaction and x ranges from 0 to 1, representing product-related CR achievement.

The S-CR curve for one-dimensional CR can be estimated by setting the function $S = a_1 x + b_1$, in which parameters $a_1 = CS_i - DS_i$ and $b_1 = DS_i$:

$$S_i = (CS_i - DS_i)x_i + DS_i \tag{7}$$

Similarly, the S-CR curve for attractive CR can be estimated by setting the function $S = a_2 e^x + b_2$:

$$S_i = \frac{CS_i - DS_i}{e - 1} e^x - \frac{CS_i - eDS_i}{e - 1} \tag{8}$$

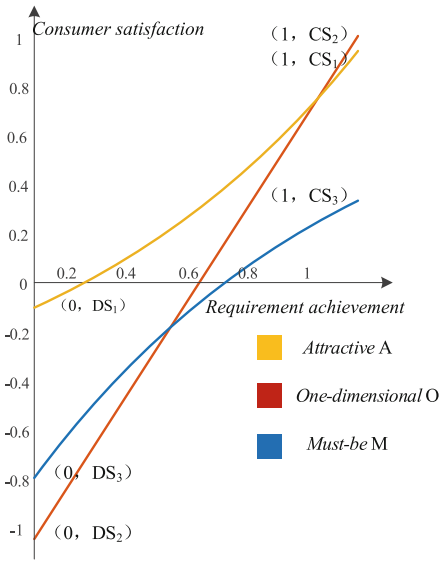

Fig. 3. Curve of Relationship between Consumer Satisfaction and CR Achievement.

The S-CR curve for must-be CR can also be estimated by setting the function $S = a_3(-e^{-x}) + b_3$:

$$S_i = -\frac{e(CS_i - DS_i)}{e-1}e^{-x} - \frac{eCS_i - DS_i}{e-1} \tag{9}$$

3.3 CR Achievement Indicator

To obtain CR achievement, the aspect-based sentiment analysis (ABSA) method is introduced, as shown in Fig. 4.

Where, $\{E_1, E_2, \ldots, E_N\}$ is the given aspect and $\{E_1, E_2, \ldots, E_N\}$ is the corresponding sentence, which are concatenated by a special token called "[SEP]" and another special token "[CLS]" is placed at the sequence header. The input sequence is encoded by BERT, and the output vector corresponding to "[CLS]" is represented as an aspect-specific sentence. Finally, the aspect-oriented sentence representation is input into the classifier for sentimental classification.

The sentimental polarity of each CR group is obtained with the above method to calculate the achievement of requirement i, namely:

$$T_i = \sum_{j=1}^{f(u_i)} \frac{T'_{i,j}}{f(u_i)} \tag{10}$$

where, j is a comment subset and $j \in u$, T_i is the polarity of requirement, $T'_{i,j}$ is the sentimental polarity of requirement i in the comment j, and $f(u_i)$ is the frequency of requirement i in the comment j.

Consumer comments are the evaluation of the achievement of requirements; thus, the analysis of CR achievements is to obtain CR achievement through normalization of the calculated polarities of consumer comments. For the convenience of statistics, the percentile rule is applied to the calculated polarity Ti after weighted processing to

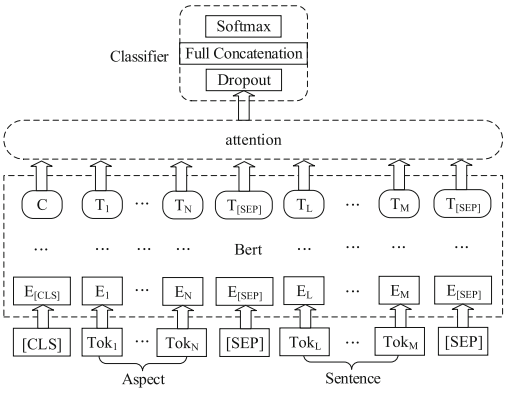

Fig. 4. Bert-backed Aspect-based Sentiment Analysis Model.

obtain the CR evaluation value ti, which is then normalized to obtain the requirement achievement value x.

$$x = \frac{t_i - \min(t_i)}{\max(t_i) - \min(t_i)} \tag{11}$$

4 Case Analysis

4.1 Kano Model-Based CR Classification

The web crawler technology was employed to obtain 33,023 comments on 146 air-conditioning products from e-commerce platforms. Customer requirements {energy saving, temperature control, appearance, sound, intelligent control, blowing system} were developed based on preliminary cluster analysis and the design experience of this kind of product. CR proportions were obtained according to frequency statistics of CR and semantic words in comments, and CR intensity values were calculated using formulas 1–4. The results are shown in Table 1.

After taking CR intensity threshold as $(\max f_i + \min f_i)/2$ and CR proportion threshold as 20%, the visualized Kano classification results of CR percentages and integrity values in Table 1 can be drawn according to Fig. 1, as shown in Fig. 5.

Among them, the abscissa is CR percentage, the ordinate is the CR intensity, and the dotted line is the thresholds of both.

Results show that among CR for air-conditioning products, sound and appearance are must-be requirements, the blowing system is a one-dimensional requirement, and temperature control, energy-saving and intelligent control are attractive requirements.

4.2 Sentiment Analysis-Based CR Achievement Calculation

To obtain the sentimental tendency of CR, comments were marked according to sentimental tag values, and the sentimental polarity analysis method was employed to predict the sentimental tendency of data, with results shown in Table 2.

Positive and negative sentiments of different CR for air-conditioning products and percentages of CR sentimental polarities were obtained to calculate CR intensity according to formulas 10–11. The results are shown in Table 3.

Table 1. CR Frequency Statistics.

Requirement	Frequency $f(u_i)$	Percentage	Requirement Intensity
Energy-saving	1014	8.11%	0.6998
Temperature control	1717	13.74%	0.7416
Sound	4553	27.85%	0.9097
Intelligence	3480	5.11%	0.6144
Blowing	638	8.75%	0.8273
Appearance	1094	36.44%	0.7941

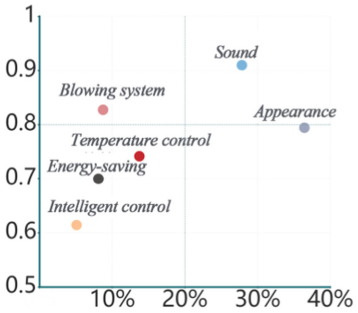

Fig. 5. Kano Classification of CR.

Table 2. CR Sentiment Tendency.

Precision	Recall	Accuracy
0.8504	0.8571	0.8483

Table 3. CR Achievement.

Requirement	t_i	x
Energy-saving	0.7438	0.9457
Temperature control	0.6782	0.8510
Sound	0.7076	0.8934
Intelligent control	0.7648	0.9760
Blowing system	0.0888	0
Appearance	0.7814	1

4.3 CR Satisfaction Calculation

According to the above calculation results, $CS = 0.4107$ and $DS = -0.6368$.The visualized curve can be drawn based on Table 4, as shown in Fig. 6. In Table 4, if CR satisfaction S is positive, the customer group tends to be satisfied with this requirement; if negative, the customer group tends to be dissatisfied with this requirement. In particular, "satisfaction" and "dissatisfaction" are both evaluations in a relative sense.

According to results, in terms of the blowing system, consumer satisfaction and achievement are low and remain to be developed; in terms of appearance, consumer satisfaction and achievement approach their highest levels; in terms of the other four aspects, with the increments of CR achievement being equal, the increments of customer satisfaction rank by intelligent control > energy-saving > sound > temperature control.

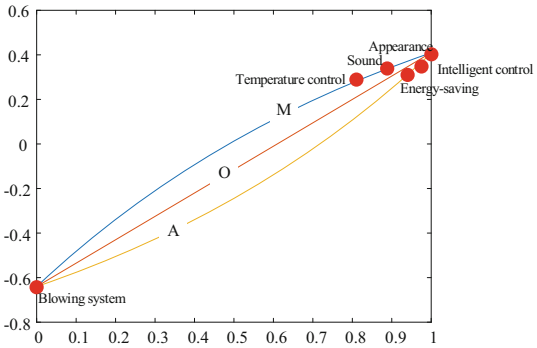

Fig. 6. S-CR Curve of CR.

Table 4. CR S-CR Curve of Air-conditioning Products.

Requirement	a	b	g(x)	S = ag(x) + b	x	S
Energy-saving	0.6096	−1.2464	e^x	$S = 0.61e^x - 1.25$	0.9457	0.3205
Temperature control	1.6571	1.0203	$-e^{-x}$	$S = -1.66e^{-x} + 1.02$	0.8510	0.2984
Appearance	1.6571	1.0203	$-e^{-x}$	$S = -1.66e^{-x} + 1.02$	1	0.3983
Sound	1.6571	1.0203	$-e^{-x}$	$S = -1.66e^{-x} + 1.02$	0.8934	0.3283
Intelligence	0.6096	−1.2464	e^x	$S = 0.61e^x - 1.25$	0.9760	0.3688
Blowing	1.0475	−0.6368	x	$S = 1.05x - 0.64$	0	−0.64

4.4 Consumer Satisfaction-Based Requirement Priority

The improvement of product attributes enhances consumer satisfaction to varying degrees because of different consumer requirements for product attributes. It is costly for enterprises to factor in all consumer requirements when designing and optimizing product attributes. Manufacturing or improving prioritized product attributes with the biggest influences on consumer satisfaction can not only save costs, but also enhance inventory operation ability and improve corporate operation efficiency.

According to analysis results of the Kano model and CR satisfaction, the priority principles for manufacturing and improving product attributes are determined as follows:

(1) Product attributes determined by the Kano model as must-be and with lower importance shall be improved in priority;
(2) Product attributes determined by the Kano model as one-dimensional and with lower importance shall be improved in priority on the premise that must-be requirements are up to consumer satisfaction;
(3) Product attributes determined by the Kano model as attractive and with lower importance shall be improved in priority on the premise that must-be and one-dimensional requirements are up to consumer satisfaction;

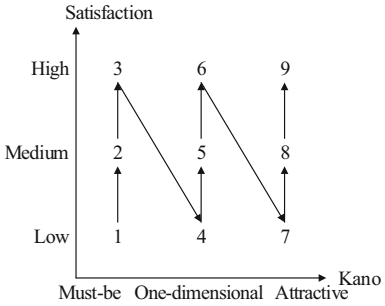

Fig. 7. Kano Model-based CR Priority.

Table 5. CR Dominance of Product A

Requirement	type	$F_{dominance\text{-}overall}$
Energy-saving	A	−2.28%
Temperature control	M	−5.9%
Sound	M	−104.17%
Intelligent control	M	−1.37%
Blowing system	A	10.98%
Appearance	O	0

Based on the above principles, it is concluded that the way to obtain CR priorities is shown in Fig. 7.

CR importance needs to be compared in the competitive market before it is more practical. Therefore, whether CR should be valued needs to refer to the competitive environment of products. In this case, product A is taken as an example to compare with the product group data, whose dominance calculation method is shown in Formula 12.

The overall CR satisfaction S_{overall} was obtained based on the Kano model, and the dominance $F_{\text{dominance-overall}}$ of this product over its peers was calculated according to Formula 12, as shown in Table 5.

As shown in Table 5, through a comparison between product A and the overall industry in consumer satisfaction, the Kano model-based CR priority is ranked by {appearance > temperature control > sound > blowing system > energy-saving > intelligent control}. Based on the overall industry analysis in the above chapters, assuming that this product is user-centered, the priority ranking can be referred to for product improvement.

$$F_{\text{dominance}} = \frac{\text{satifaction with this requirement} - \text{satisfaction with all requirements in the product set}}{|\text{Satisfaction with all requirements in the product set}|} \tag{12}$$

5 Conclusions

This paper combined big data research with the classical Kano model to analyze the nonlinear relationship between requirement achievement and consumer satisfaction, and proposed a data-driven method to analyze CR satisfaction quantitatively. Based on experiments, some reasonable suggestions were put forward through the quantitative evaluation of current customer requirements. There are also some shortcomings in this paper. In the process of combining big data research with classical theoretical models, all relevant parameters were obtained from the summary and analysis of parameter definitions with the help of natural language processing technology. Compared with questionnaire surveys, this method features wider sample universality and stronger timeliness, but its accuracy is lower.

References

1. Hu S, Liu J, Wang YQ, Guo WQ, Wang N (2020) Research on product iterative innovation design method based on user dynamic demand. Mod Manuf Eng 12:41–48
2. Lina H, Xinguo M, Miao L, Zheng MK, Zhitao X (2017) Understanding customer requirements through quantitative analysis of an improved fuzzy Kano's model. Proc Inst Mech Eng Part B: J Eng Manuf 231(4)
3. Peng DH, Huang ZH, Peng B (2021) User demand-oriented product design quality evaluation model. J Chin Comput Syst 42(01):218–224
4. Timoshenko A, Hauser JR (2019) Identifying customer needs from user-generated content. Market Sci 38(1):1–20
5. Xu B, Yu JH, Sha H, Shi PY, Wu Y (2020). Research on Kano model analysis method based on product comment data. Product Res (09):10–15+52+61
6. Yan YH (2021) Research on satisfaction model design based on comment data. Jiangxi Sci 39(01):143–149
7. Zhang DJ, Hou ZX, Huang L, Zhao YH (2020) Research on product requirement configuration method combining with consumer satisfaction. J Graph 41(04):649–657
8. Zhao S, Zhang Q, Peng Z, Fan Y (2019) Integrating customer requirements into customized product configuration design based on Kano's model. J Intell Manuf 31(3):597–613. https://doi.org/10.1007/s10845-019-01467-y
9. Zhou AQ, Zhao XH, Zhang MY, Shi YB (2020) Two-stage customized product design based on customer satisfaction. J Xi'an Polytech Univ 34(01):105–112