Title:
A computer-aided Approach for Acquisition and Importance Ranking of Customer Requirements from the Online Comment Mining

Authors:
Li Xiangdong, lee96xd@163.com, Hebei University of Technology
He Youlong, hyl199508@163.com, Hebei University of Technology
Peng Qinjin, qingjin.peng@umanitoba.ca, University of Manitoba
Liu Wei, lwofhebut@126.com, Hebei University of Technology

Keywords:
Online Comment Mining, Customer Requirements, Requirement Acquisition, Importance Ranking, Quality Function Deployment

DOI: 10.14733/cadconfP.2021.267-271

Introduction:

Nowadays, product design has developed from product-oriented design to customer-oriented design [1], therefore, the analysis of customer requirements (CRs) is more important for enterprise to improve customer satisfaction [2]. With the development of computer technologies, ways to acquire CRs from a large number of online comments has been facilitated by text mining. Liu et al. [3] and Jiang et al. [4] 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. [5] proposed a demand-centered approach for requirements evolution by mining and analyzing online reviews. Peng et al. [6] proposed a method for acquiring CRs based on feature models and collaborative filtering, in order to improve the accuracy and efficiency of the software requirement acquisition process.

After acquiring the CRs, it is necessary to rank the importance of CRs to determine the key points for product improvement. There are many methods for the importance ranking of CRs, such as score ranking method, analytic hierarchy process (AHP), rough set, fuzzy set. But the customers are always ambiguous or contain multiple semantics when expressing their requirements. The problem of low accuracy inevitably occurs when the traditional methods is used to calculate the importance weights of CRs. In order to solve this problem, the fuzzy sets are employed to enhance the expression of CRs [7]. The fuzzy sets convert the linguistic assessment of CRs to triangular fuzzy numbers, which are used to build the pairwise comparison matrix of AHP [8]. 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.

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.

To solve the above problems, 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 [9]. 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. Second, we propose 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, we employ the Kano model and sentiment analysis to improve the importance ranking of CRs. Finally, the CRs and importance analysis results are integrated into the QFD to build an integrated system for the transformation of CRs. And a case study implemented the acquisition and analysis of CRs for a smartphone to demonstrate the feasibility and effectiveness of the proposed methodology.

Main Idea:

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.

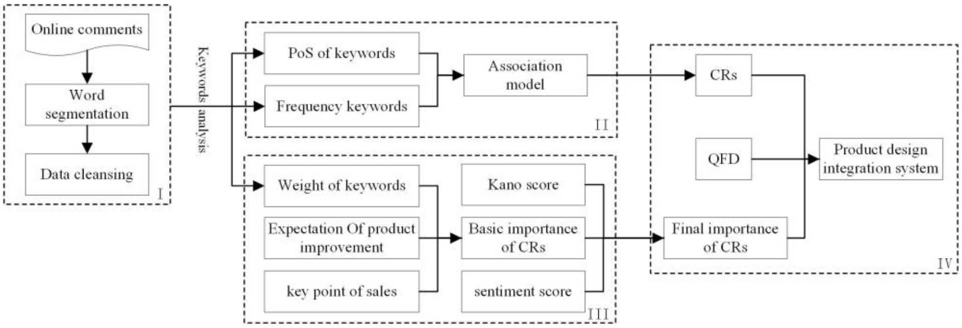

Fig. 1: Flow chart of the method.

Data Collection and Pretreatment

Online comments from a wide range of sources, such as e-commerce platforms, forums and online virtual communities, which store a large number of customer comments data. In this research, the Octoparse [10], a crawler software for data collection from China, was used to crawl customers' comments.

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 stop-words removal, which aims to eliminate noise words unrelated to CRs.

In this research, the ICTCLAS [11], a Chinese lexical analysis system, is used for semantic analysis. It is a Chinese lexical analysis based on the Cascaded Hidden Markov Model (CHMM), which includes five levels of atomic segmentation, simple unknown word recognition, recognition of nested unknown words, class-based hidden horse segmentation and hidden horse tagging of POS.

Integrated 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. Product attributes are usually expressed in two ways in the comment text. The first way is expressed by nouns, such as "system", "appearance", "battery", etc. The second way is implicit in the comments, such as "(appearance) nice", "long (battery) standby time".

We proposed a model to express CRs by combinations of feature words which include noun, verb and adjective. The first combination is "noun + verb + adjective", such as "system/n run/v fluent/a". The second combination is "noun + adjective", such as "screen/n beautiful/a". The third combination is

Proceedings of CAD'21, Barcelona, Spain, July 5-7, 2021, 267-271

"verb + adjective", such as "charge/v fast/a". These three combinations summarize the majority of product attributes, and fully reflect the customer sentiment expression of product attributes.

Therefore, CRs can be expressed in feature terms as combinations 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.

The word association model for customer requirement acquisition is shown in Fig. 2. In this model, all feature items are clustered to obtain CRs.

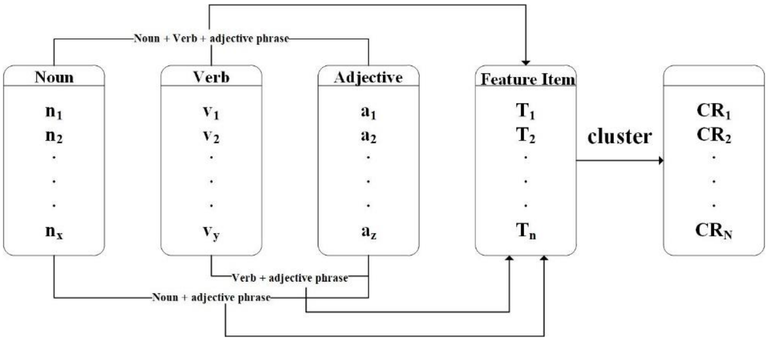

Fig. 2: Word association model.

Method for Ranking Importance of Customer Requirements

Assume each customer requirement can be expressed by n feature terms:

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

where CR_j is the j^{th} customer requirement; T_i is the i^{th} feature item that constitutes the customer requirement. The weight of the i^{th} feature term is:

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

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 word 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

In order to improve the accuracy and completeness of the calculation results, weight correction is carried out by using the improved ratio and key points of sales 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 product improvement expectation determined by the enterprise according to the market condition, capital condition, competition condition, etc., and the formula is:

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

p is the key point of sales, representing the extent to which technical characteristics of product affect customers purchase decisions. It can be divided into three levels: level 1 means the customers want to own and are willing to pay, the weight is 1.5; level 2 means the customers are interested but will consider the price, the weight is 1.2; level 3 means the customers is not interested, the weight is 1.

According to the basic importance of CRs, we propose a method to improve the accuracy of basic importance by integrating the Kano model and the sentiment polarity. The determination process of final importance is shown in the following steps.

Step 1. Identify the Kano category of the CRs

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 (SI) and dissatisfaction influence (DSI) of the CRs

The formulas for calculating the satisfaction influence and dissatisfaction influence are:

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

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

Step 3. Sentiment polarity analysis

This research adopts the method of sentiment polarity analysis based on sentiment dictionary, and calculates the sentiment score by using negative word dictionary and degree adverb dictionary.

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

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.

Step 4. Ranking of importance of CRs

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

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

An Integration Method for the Transformation of Customer Requirements

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 QFD.

Case Study

In this section, a case study is implemented to acquire and prioritize the CRs of smart phones. The QFD is employed to redesign the target characteristics that need to be improved in order to verify the effectiveness of the proposed method. This case study considers smart phones because there are a large number of online comments in online shopping platforms. In this research, Octopus 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.

Conclusions:

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. 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.

Funds Information:
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, 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] 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

[4] 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

[5] 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

[6] 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

[7] 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

[8] 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

[9] Akao, Y.: Quality function deployment: Integrating customer requirements into product design, Productivity Press, New York, NY, 1990.

[10] Octoparse, https://www.octoparse.com/, Octopus Data Inc.

[11] 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