Product design research based on online reviews and quality features

Haibin Feng, Chuanfei Li, Keqing Fan, Zhihao Wang

Shandong University of Technology, Zibo, 255022, China

Abstract: With the gradual rise of social living standards, users' shopping needs are increasingly diversified, differentiated, personalized and other characteristics. Therefore, how to accurately grasp the focus of user needs has become the key to enterprise product design innovation and product quality management. The development of quality functions is one of the core tools for user demand-driven product design innovation. Based on this, a product design method based on online review and quality function expansion is proposed. This paper uses BTM topic model to extract user topics from online review texts, uses Word2vec word vector similarity and cosine similarity to map topics to requirements, and determines the weight of each demand. Finally, each demand and the corresponding weight are input into the House of Quality model to determine the importance ranking of each technical feature. Use quality function to expand the model to achieve online review-driven product design.

Keywords: Online Reviews, BTM Theme Modeling, Quality Function Expansion, Product Design, Quality House

1. Analysis of related concepts

1.1 Online comments

Online reviews, also known as online customer reviews, which serve as one of the most prominent forms of online word-of-mouth communication, are text-based evaluations of products, and are information submitted by consumers via the Internet to review a product or company[1]. Compared with product reviews collected offline, online reviews on the Internet can reflect consumers' views on products in a more realistic and timely manner. Therefore, in the context of big data, how to mine the demand information that users pay attention to from the online reviews of products published on the Internet and obtain the degree of importance of users' demands becomes the focus of the article's research.

1.2 Quality function development

Quality Function Deployment (QFD), as a customer needs oriented product design methodology, provides a model for the transformation between customer needs and functional requirements and enables the derivation of key modules in product design.QFD is a quality assurance methodology for the product design phase and a methodology for coordinating the various functions of the product development department [2]. Nowadays, QFD is widely used in the process of enterprise product innovation as well as academic research practice.

2. Product Design Ideas Based on Online Reviews and Quality Functionality Expansion

2.1 User demand acquisition based on online review data

2.1.1 Data acquisition and pre-processing

The article mainly uses python software to accomplish the acquisition of online review data. Since the input data of the BTM topic model used subsequently should be the words after word splitting, the main work related to cleaning, de-weighting, word splitting and de-deactivation of the data is carried out in the data preprocessing stage to form a corpus consisting of words [3].

2.1.2 BTM Topic Modeling to Extract Comment Topics

Compared to the traditional LDA topic model, the BTM topic model chooses to use word pairs extracted from the entire corpus for topic modeling instead of using individual words, as a way to solve the sparsity problem associated with short texts consisting of only a few words. The word pairs of the word co-occurrence model are able to reveal the topic better than the individual words [4-5].

The number of topics K needs to be entered in the modeling process of topic model , the optimal number of topics K will have a great impact on the results of the BTM model run, the article uses the confusion degree to determine the optimal number of topics, the confusion degree is calculated by equation (1):

$$Perplexity = \exp\left\{-\frac{\Sigma \log(p(b))}{B}\right\} \qquad \text{* MERGEFORMAT (1)}$$

where $P(b)$ is the conditional probability of word pair b and B is the corpus of word pairs.

After determining the optimal number of topics K , the values of parameters Φ and θ need to be inferred. The main steps of parameter inference are as follows:

Inputs: optimal number of topics K , a priori parameters α and β , set of word pairs B , number of iterations N .

Output: polynomial parameters Φ and θ .

① Perform N iterations for word pairs b belonging to the set B of word pairs.

② Calculate the conditional probability $P(z|z_{-b}, B, \alpha, \beta)$ of word pair b and use the conditional probability to assign a topic Z to each word pair.

$$P(z|z_{-b}, B, \alpha, \beta) \propto (n_z + \alpha) \frac{(n_{wi|z} + \beta)(n_{wj|z} + \beta)}{(\Sigma_w n_{w|z} + M\beta)^2} \qquad \text{* MERGEFORMAT (2)}$$

The values of parameters Φ and θ can be calculated by equations (3) and (4).

$$P(w|z) = \Phi_{w|z} = \frac{n_{w|z} + \beta}{\Sigma_w n_{w|z} + M\beta} \qquad \text{* MERGEFORMAT (3)}$$

$$P(z) = \theta_z = \frac{n_z + \alpha}{|B| + K\alpha} \qquad \text{* MERGEFORMAT (4)}$$

After the inference calculation for polynomial parameters Φ and θ , the document-subject distribution can be further inferred, and the main calculation process is as follows:

① Calculate the distributional probability of word pairs - themes from equation (5):

$$P(z|b) = \frac{P(z)P(w_i|z)P(w_j|z)}{\sum_z P(z)P(w_i|z)P(w_j|z)} = \frac{\theta_z \Phi_{wi|Z} \Phi_{wj|z}}{\sum_z \theta_z \Phi_{wi|z} \Phi_{wj|z}} \qquad \text{* MERGEFORMAT (5)}$$

② Calculate the distribution probability of document-word pairs from equation (6):

$$P(b|d) = \frac{n_d(b)}{\Sigma_b n_d(b)} \qquad \text{* MERGEFORMAT (6)}$$

where n is the frequency of word pair b in the document.

③ Calculate the probability of the document-subject distribution from equation (7):

$$P(z|d) = \sum_b P(z|b)P(b|d) \qquad \text{* MERGEFORMAT (7)}$$

2.1.3 Generating user requirement sets

The article tries to use word2vec word vectors to train the word vectors of topic words under each topic, calculate the similarity of each word, and select the emotion words with high similarity to match two by two to complete the mapping from the first-level topics of online reviews to the second-level user needs, and finally we can get the set of user needs of the related products $\{CR_1, CR_2, \cdots CR_i\}$. and the cosine similarity between two word vectors is calculated by equation (8):

$$\cos\theta = \frac{\Sigma_{i=1}^{n}(A_i * B_i)}{\sqrt{\Sigma_{i=1}^{n}(A_i)^2} * \sqrt{\Sigma_{i=1}^{n}(B_i)^2}} \qquad \text{* MERGEFORMAT (8)}$$

2.1.4 Determining user requirement weights

Based on the results obtained from the BTM model, the product of the topic-word pair distribution probabilities of the two similar keywords that make up a user need under a given topic and the document-topic distribution probability for that topic is used as the weight λ_k for that user need k .

$$\lambda_k = M_l \sum_{i=0}^{1} N_{li} \quad \text{* MERGEFORMAT (9)}$$

Where M_l is the document-topic distribution probability of the l th comment topic, N_{li} is the topic-word pair distribution probability of the k similar keywords that make up the user demand, and finally the user demand weight A is normalized to get the final user demand importance.

2.2 Construction of Quality House Model

2.2.1 Determining the technical characteristics of the product

The technical characteristics of the product are also crucial to the construction of the quality house, the article mainly through the reading of the literature of the relevant products, as well as consulting the technical manuals of the relevant products, and at the same time, communicating with the relevant product technicians, combining with the collection of the identified user requirements to form the collection of the relevant product's technical characteristics of the collection $\{EC_1,\ EC_2,\ \cdots,\ EC_j\}$.

2.2.2 Constructing a quality house model

Quality House (QH), as an intuitive matrix framework expression, is an important tool for realizing quality function unfolding. After constructing the Quality House model, the absolute importance of product technical characteristics is calculated by equation (10) and the relative importance of product technical characteristics is calculated by equation (11).

$$W_j = \sum_{i=1}^{q} W_i P_{ij} \quad \text{* MERGEFORMAT (10)}$$

$$W_k = \frac{W_j}{\sum_{i=1}^{q} X_j} \quad \text{* MERGEFORMAT (11)}$$

In the above formula, W_j represents the absolute importance of product technical characteristics, W_i represents the importance of product user needs, P_{ij} represents the correlation coefficient between user needs and technical characteristics, and W_k represents the relative importance of product technical characteristics [6].

3. Product design practice based on online review and quality function development

3.1 Data Acquisition and Preprocessing

The article takes laptop computers as the products in the example study, and chooses Apple's MacBook pro 13-inch laptop as the specific product object of the study. A total of 2,000 relevant product reviews were crawled from the Jingdong e-commerce platform, and 1,942 valid reviews were eventually left after eliminating duplicates and irrelevant reviews. Subsequently, data preprocessing and lexical processing are carried out on the effective comment data to obtain the lexical results.

3.2 Access to user comment threads

When running the BTM topic model, the number of subjects needs to be set. Article after generating the perplexity image of the related comment data to determine the number of subjects, as shown in Fig. 3, it can be found that the optimal number of subjects k for the article comment data is 7. After setting the optimal number of subjects, and at the same time, the number of keywords under each subject is set to 8. Running the BTM subject model can get the keywords under the 7 user comment subjects and their subject- word pair distribution probabilities, as shown in Table 1:

Table 1 Online review data themes Table

Topic1	Topic2	Topic3	Topic4	Topic5	Topic6	Topic7
Screen	Heat dissipation	Ultrathin	Appearance	Quality	Battery	Operation
*0.304	*0.212	*0.337	*0.331	*0.131	*0.151	*0.151
Distinct	Performance	Carry	Modelling	Price	Endurance	Smooth
*0.071	*0.173	*0.06	*0.281	*0.072	*0.039	*0.044
Colour	Fan	Portable	Beautiful	Brand	Durable	Ultrafast
*0.057	*0.047	*0.033	*0.068	*0.045	*0.037	*0.044
Delicate	Temperature	Easy to carry	Simple	Excellent	Stand-by time	Rapid
*0.045	*0.035	*0.02	*0.023	*0.037	*0.023	*0.037
Great	Amazing	Light and handy	Concise	Cost-efficient	Practical	Perfect
*0.032	*0.022	*0.02	*0.019	*0.03	*0.023	*0.034
Resolution ratio	Silence	Great	Fashion	Cost performance	Electric quantity	Configura-
*0.023	*0.021	*0.017	*0.016	*0.021	*0.021	tion*0.03
Amazing	Stronger	Tiny	Quality	Great	Performance	Function
*0.018	*0.014	*0.014	*0.015	*0.019	*0.017	*0.024
Definition	Powerful	Appearance	Outline	Quality of goods	Powerful	Excellent
*0.018	*0.011	*0.013	*0.007	*0.01	*0.016	*0.02

3.3 Acquisition of user requirements

Using word2vec to calculate the word vectors of related words, and then calculate the similarity of related words, screen out the topic keywords with high similarity and can be combined as the user's needs, and finally get the set of user's needs of the related products C {clear screen, fine color, good heat dissipation, thin and light, portable, beautiful appearance, simple styling, cost-effective, battery durability, strong endurance, running silky smooth}, and then the importance of each user requirement can be calculated by equation (9). As shown in Table 2:

Table 2 User requirements and weighting table

Comment subject layer	User requirement layer	Weight
Display performance	Clear screen	0.2552
	Fine color	0.1406
Heat dispersion	Good heat dissipation	0.0468
Portablity	Light and portable	0.0729
Appearance and shape	Pretty appearance	0.3072
	Simple shape	0.0885
Cost performance	Inexpensive cost	0.0156
Battery performance	Battery durability	0.0416
System speed	Running silky	0.0312

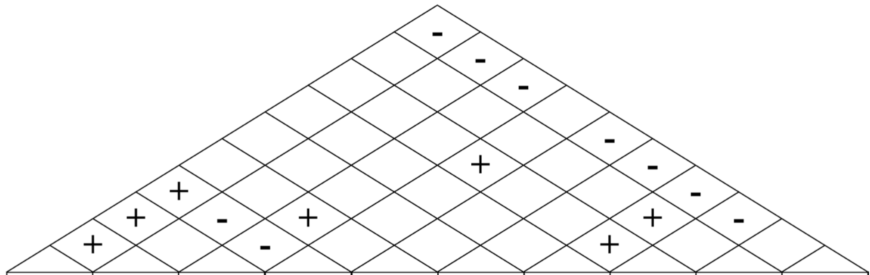
3.4 Formation of product technical characteristics

Combined with the related product literature research and product technical manuals, corresponding to the existing product user requirements can form a collection of product technical characteristics, which are chip, CPU, GPU, memory, battery energy efficiency, XDR display, battery capacity, size, battery material, body material, appearance form, product pricing.

3.5 Construct product quality house model

After obtaining the user's demand, importance and technical characteristics of the product, when constructing the quality house, it is necessary to input the user's demand and importance of the product into the left wall of the quality house, and input the technical characteristics of the product into the ceiling of the quality house. In addition, when constructing the body of the quality house, it is necessary to judge and assign values to the correlation relationship between user needs and technical characteristics, and construct the correlation relationship matrix between user needs and technical characteristics. The correlation between user requirements and technical characteristics can be expressed by strong correlation, medium correlation, weak correlation, and no correlation respectively, and the corresponding scores are 5, 3, 1, and 0.

After completing the construction of the left wall, ceiling and body of the quality house, it is necessary to calculate the absolute and relative importance of the technical characteristics of the product using equation (12) (13) to construct the basement of the quality house. In addition, it is also necessary to judge the correlation between the technical characteristics, mainly positive correlation and negative correlation two kinds of correlation, in the technical characteristics of the correlation between the judgment, as the basis for the construction of the ceiling of the quality house. The final construction of the completed quality house is shown in Figure 1:



		Microchip	CPU	GPU	Battery Energy Consumption	XDR Display	Battery capacity	Dimension	Fuselage Material	Appearance	Product Pricing
Clear screen	0.2552	5	5	3		3		1			1
Color finesse	0.1406	5	5	3		3					1
Good heat dissipation	0.0468	5	5		1				3		1
Light and portable	0.0729							3	3	1	1
Nice appearance	0.3072							5	5	3	1
Simple styling	0.0885							5		3	1
good price	0.0156										3
Battery Durability	0.0416		1		3		3				1
Runs silky smooth	0.0312	3	3	1							1
Absolute importance of technology needs*100		148.38	152.54	201.02	25.48	197.9	20.8	180.68	152.01	205.14	103.12
Relative importance of technology needs(%)		10.6	10.9	14.5	1.8	14.2	1.5	13	10.9	14.8	7.4

Fig. 1 macbook pro laptop product quality house model diagram

By constructing a quality house model of Macbook pro laptops and observing the basement part of the quality house, it can be found that in the design innovation process of Macbook pro laptops, the appearance form of the laptop (14.8), the GPU (14.5), and the XDR display (14.2) are of the most important, and should be given the primary attention; secondly, the laptop's size (13), body material (10.9), CPU (10.9), and chip (10.6) are also relatively important and should receive corresponding attention; furthermore, the product pricing of laptop computers (7.4), battery energy consumption (1.8), and battery capacity (1.5) are sub-important and should receive moderate attention.

4. Conclusion

Online review data, which is rich in consumers' emotions, has become a key resource for enterprises to iterate their products and innovate their services. How to effectively and adequately utilize online review data to realize product design innovation has become the key for enterprises to obtain core competitiveness. The article proposes a method to utilize online reviews as a data source, use natural language processing technology to obtain user needs and importance, and then use quality function expansion technology to construct a quality house to realize product design innovation. However, there are still some shortcomings in the study. First of all, online reviews do not fully reflect the real needs of consumers, and in future research, we can try to obtain user requirements from both online and offline channels. In addition, the article still uses a relatively subjective approach in the process of converting from user requirements to technical characteristics, so the resulting technical characteristics may be vague and not clear enough. In future research, attention should also be paid to avoiding the influence of subjective factors on the conversion of technical characteristics.

References

[1] Du X M,Ding J Y,Xie Z H. A study on the impact of online reviews on consumers' purchase intention[J]. Management Review,2016,28(03):173-183.

[2] Hu Y, **ao R B, Zhang W X. QFD customer needs mining driven by product review data[J]. Computer Integrated Manufacturing Systems, 2022, 28(1): 184-196.

[3]Wang Q,Liu Y. Product Improvement Design Based on Sentiment Analysis of Online Reviews and QFD [J]. Packaging Engineering,2021,42(24):169-174.

[4]Yan X, Guo J, Lan Y, et al. A biterm topic model for short texts[C]//Proceedings of the 22nd international conference on World Wide Web. 2013: 1445-1456.

[5] Wang Y, Hu Y. Hotspot detection in microblog public opinion based on biterm topic model[J]. Journal of Intelligence, 2016, 35(11): 119-124.

[6]Jin J, Ji P, Liu Y, et al. Translating online customer opinions into engineering characteristics in QFD: A probabilistic language analysis approach[J]. Engineering Applications of Artificial Intelligence, 2015, 41: 115-127.