ISSN: 2664-9640

Research on User Demand Based on E-commerce Consumer Negative Reviews

Pingchuan Wen[1] and Xianbiao Li[1]

[1]School of Economics and Management, Chongqing University of Posts and Telecommunications, China.

Abstract

This article creates the ultimate goal for enterprise products. First, this article uses the theme web crawler technology to collect and preprocess user negative review data, and uses text analysis technology to identify user needs and lay the foundation for user needs for enterprise product innovation. Second, the axiomatic design theory is used to achieve user needs. Mapping to product functions, using cognitive map theory to implement user demand-driven product innovation solutions generated by user participation, and providing an idea or method for enterprise product innovation. Finally, based on the smart speaker on Jingdong Mall as an example, Verify the feasibility of the research and make predictions for the future development of smart speakers.

Keywords

E-commerce users Negative comment User demand.

1. Introduction

The widespread use of e-commerce platforms has brought unprecedented opportunities to enterprise user demand analysis while also facing many challenges. First, the real-time, shared, and interactive features of Internet information dissemination require companies to be able to respond to and process user information quickly and efficiently. Secondly, the characteristics of freer expression of opinions and views under online social networks and more diversified user requirements also pose challenges for companies to conduct scientific and effective demand analysis. At present, the analysis of user needs based on e-commerce platforms has become a hot issue for scholars and has achieved certain research results. After reading the existing research, we find that there is less research on user needs mining based on user negative reviews. It is possible to better grasp the user's needs in the negative comments of the users; it is less researched to improve or develop new products based on user needs. How to quickly and effectively convert user needs into new products is an urgent need for enterprise product development. solved problem. This research aims at the deficiencies in the research and improves and perfects. Not only can it enrich the theory of user demand analysis, but its research methods and results are conducive to the rapid collection of user information, the effective analysis of user needs, and the rational formulation of product upgrades Or new product development plans, can effectively play a role in the field of enterprise product management.

2. Literature Review

2.1. Research Status of User Demand Acquisition

The first step in user requirements analysis is requirements acquisition. According to the way of communication, the traditional methods of user demand acquisition can be divided into four categories: talk, observation, analysis, and comprehensive Interviews are the most common method for acquiring user requirements. It mainly involves two or more people directly

ISSN: 2664-9640

acquiring user requirements through conversation. Parasuramanet et al. Designed a conceptual model that considers user needs to include ten dimensions, which are reliability, responsiveness, ability, access, courtesy, communication, integrity, security, understanding and tangible assets.

2.2. Research Status of Traditional User Needs Identification

User demand identification was first proposed by Scott in a report from Northwestern University in 1901. In the report, he pointed out that advertising is a science, and user demand plays an important role in advertising. It is a question worthy of scholars' research. During the period from 1920 to the Second World War, due to poor product sales, many companies gradually began to realize that more products were sold through sales and gradually began to transition from the "seller market" to the "buyer market". Driven by this trend, more companies have begun to focus on sales. In the fierce market competition, in order to obtain as many users as possible, companies have focused on identifying user needs. In order to help enterprises solve the problem of identifying needs, scholars have begun to pay attention to user needs and carry out systematic research.

2.3. Research Status of New-type User Demand Identification

Levy et al. Analyzed 1946 one-star scoring reviews from 10 popular e-commerce platform review sites, and found that front desk employees, bathrooms, cleanliness of rooms, and noise of guest rooms are the most common demand dimensions that users are not satisfied with. Fu et al. Collected reviews of Android system applications from GooglePlay, analyzed the one-star and two-star scoring reviews using the LDA method, and found the ten most unsatisfactory user requirements dimensions, including stability, compatibility, Costs etc.

3. User Needs Identification

This paper builds a demand management framework based on the data analysis method of demand collection. The data in the data analysis method includes a variety of sources, such as product usage logs, user information in user management systems, and user webpage visits. These data have some value, but it is difficult to express the user's true thoughts. By reviewing the literature, it is found that there are many studies on the use of online reviews to mine user needs. Most of them have studied user needs by researching the positive, middle, and negative reviews of products. The value is low, and the value of mining user needs from negative reviews is high. Therefore, this article will only use user negative reviews as a data set to mine user needs. The research steps are shown in Figure 1. Data preprocessing, negative text feature extraction, and negative user .identification

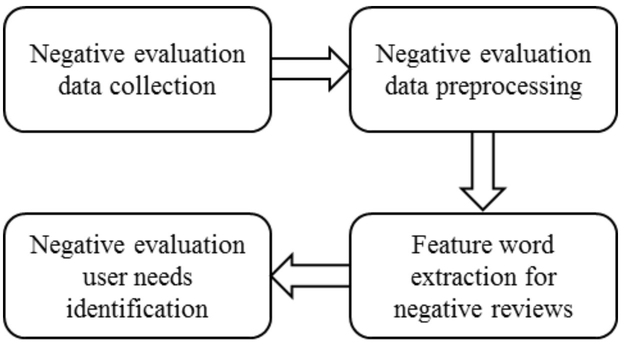

Figure 1. Demand Identification Process Based On User Negative Evaluation

(1) Data collection for negative reviews

Data collection refers to the use of data grabbing programs or software to meet the requirements of laws and webpages to collect target data to form a database. Individuals can write data grabbing programs using Python, Java, R and other languages to grab data. You can also use the data capture software GooSeeker, octopus, train collector, etc. to capture data. This article mainly collects target negative evaluation data by writing Python code.

(2) Pre-processing of bad evaluation data

The data prediction processing mainly includes preliminary preprocessing and text preprocessing. The preliminary preprocessing refers to deleting missing values and duplicates in the collected data. The text preprocessing generally includes word segmentation and part-of-speech tags. It mainly includes review data field filtering and data cleaning. The purpose of field filtering is to select review text content from reviewers, review time, review stars, and review text. The main purpose of data cleansing is to remove duplicate, false and irrelevant reviews, and let review data Authentic.

(3) Feature word extraction for negative reviews

The purpose of text feature word extraction is to represent the text as a data structure that the algorithm can process, extract useful information from the text, convert text information that cannot be recognized by the computer into feature vectors, and use the computer to find some high-value information; common feature words Common extraction methods include TF-IDF, Word2Vec, and LDA for feature word extraction. In this paper, TF-IDF method is used for feature word extraction.

(4) Expression of user needs for negative reviews

The expression of user needs is based on the calculation of text feature word vectors, the similarity calculation of feature vectors, classification and association analysis, etc. to find the bad user needs; the calculation of feature vectors is mainly based on artificial intelligence, neural networks and data mining. The related technology realizes the classification, clustering, and association analysis of text data, and mines valuable information in text data, which is further expressed as user needs.

3.1. Collection of User Negative Comments Text Data

Scrapy is an application framework written to crawl website data and extract structured data. The Scrapy framework can help users easily collect network resources, reduce some repetitive operations when writing crawler code, and to some extent help users crawl structured data. The Scrapy framework mainly includes components such as engine, scheduler, downloader, crawler, and pipeline. This article uses these components to customize its own crawler program. The flow of the Scrapy framework is shown in Figure 2.

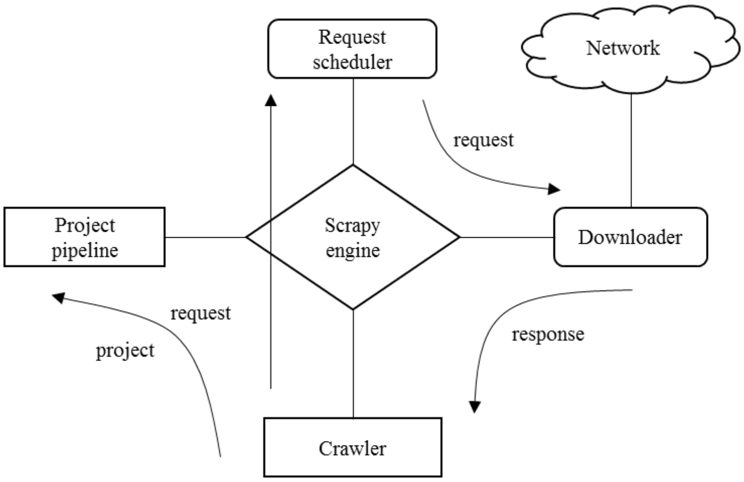

Figure 2. Flowchart Of The Scrapy Framework

ISSN: 2664-9640

The engine in the Scrapy framework is the brain of the framework. It is responsible for all request and response operations, and each step of the process will go through the engine for the next procedure. The engine will communicate with each other module and listen to each module. When a module completes its own task, the engine will notify the next module to perform the next operation. Each module has its own division of labor. This coupling is relatively low, which provides users with great convenience in terms of use and configuration. In the project, the initialized URL will be issued to the dispatcher and wait for dispatching. First, the engine will select the URL from the dispatcher to access. After accessing, the engine will be notified and the engine will then notify the downloader to download the content of the webpage for download. The next process is used. After the web page is downloaded, the engine will notify the crawler to obtain the data fields it wants. We can use XPath technology to obtain the data of the title and body of the web page. Finally, the engine will notify the project pipeline to send the required data. Download to save locally.

The theme crawler system of this article is a set of theme crawler system based on Scrapy. When the crawler requests the topic crawler system, the Scrapy framework will encapsulate the request to call the corresponding network resource. After downloading the network resource, it will fetch and parse the returned HTML page through xpath, and then obtain the content of the field we want, calling us in advance The trained classifier judges the relevance of the topic, stores the content fields that meet the requirements of the topic into the database, and finally parses the URL information in the web page to the URL queue to continue crawling.

3.2. User Negative Comment Data Preprocessing

3.2.1 User Negative Comment Screening

The rapid development of the Internet has profoundly changed the speed and influence of information generation. In the field of e-commerce, the "China E-Commerce Development Report (2018-2019)" states that in 2018, China's total e-commerce transactions totaled 31.63 trillion yuan, an increase of 8.5% year-on-year; in the first half of 2019, online retail sales of physical goods increased by as much as 21.6%, accounting for 19.6% of the total retail sales of consumer goods. Consumers are increasingly turning to online purchases. Valuable online reviews have largely provided a reference for consumers, avoiding the risks and risks of the purchase process. uncertain. Under the trend of economic benefits, more and more retailers are aware of the word-of-mouth effect of online reviews, and have begun to manipulate and participate in the practice of word-of-mouth, and influence consumer purchase behavior by manipulating online product reviews The malicious users whose main goal is advertising and marketing are increasingly active. By publishing or reposting product reviews that contain false content, they influence users' purchasing decisions, thereby promoting the sale of their own products or suppressing competitors. False reviews began to flood the Internet, disrupting consumer purchase judgments, affecting the reference value of online reviews, causing unnecessary economic losses to consumers, and increasing the difficulty for companies to discover user needs and improve products through user reviews. Deng Shengli and others divided reviews into true reviews, fake reviews, and meaningless reviews based on whether the theme content wanted to be relevant to the theme. According to the types of fake reviews, three people with more than five years of online shopping experience were invited. The master screened false reviews, and developed three criteria for false negative reviews: 1) negative reviews where the reviewer, review time, and review content were repeated at the same time; 2) negative reviews where the review content did not match the number of reviews; 3) The screening process of negative reviews that are not related to the product is shown in Figure 37. When two reviewers think that the review is true, the review is a valid review, otherwise it is an invalid review.

3.2.2 Segmentation of negative user comments

ISSN: 2664-9640

At present, there are many relatively complete typical word segmentation systems and algorithms in China, such as: ICTCLAS Chinese word segmentation system, HTTP-based open source Chinese word segmentation system (HTTPCWS), Simple Chinese word segmentation system (SCWS), Pangu word segmentation, and stuttering word segmentation algorithms. According to reports, ICTCLAS 3.0 word segmentation speed is 996KB / s, stand-alone accuracy is 98.45%, API does not exceed 200KB, and various dictionary data is compressed to less than 3M. It is currently the best Chinese lexical analyzer in the world; its main function is new word recognition, part-of-speech tagging, Chinese word segmentation, and named entity recognition. In the part-of-speech tagging, "/ n" represents a noun, "/ v" represents a verb, and "/ a" represents an adjective. At the same time, the ICTCLAS Chinese word segmentation system also supports The user-defined dictionary is imported, and users can add word segmentation rules based on their own research needs; therefore, this article intends to use ICTCLAS word segmentation software to perform Chinese text word segmentation on user negative comments.

After the Chinese text is segmented, it is usually necessary to delete the stop words of the text. Stop words are words or words that contain little or no information, such as "about", "present", "of", etc., and also contain some special symbol. Common stopwords include: Chinese Academy of Sciences, Harbin Institute of Technology, Baidu stopword list, etc. In order to maintain consistency, this article chooses the Chinese Academy of Sciences stopword list.

3.2.3 User negative comment keyword extraction

The main idea of TF-IDF is: If a word or phrase appears frequently in a text and has a high frequency of TF, and it rarely appears in other articles, it is considered that the word or phrase has a good classification ability and is suitable for classification.

In the TF-IDF technology, TF represents the word frequency of a specific word in the text. Generally, the higher the word frequency, the more important the word is. Assuming that the specific text is j, the TF calculation formula of feature word i in text j is as follows:

$$tf_{ij} = \frac{n_{ij}}{K}$$

Among them, n_{ij} represents the number of times the feature word i appears in the text j, and K represents the total number of feature words, that is, the sum of all feature words.

IDF stands for inverse document frequency, that is, the exclusivity or uniqueness of a particular word in a particular text, and distinguishing topics between texts has a core influence. The formula for calculating IDF of feature word i is as follows:

$$idf_{ij} = \log\frac{N}{n_i + 1}$$

Among them, N represents the total number of texts, and n_i represents the number of texts including the feature word i. The reason for adding 1 to n_i is to prevent the denominator in the fraction of the IDF value from being calculated to increase the robustness of the algorithm.

Based on the above analysis, the TF-IDF of the feature word i in the text j is:

$$w_{ij} = tf_{ij} \times idf_{ij}$$

According to the above analysis, TF-IDF considers the word frequency and specificity of the feature words, and can effectively extract keywords in the text. The disadvantage of the algorithm is to consider the position where the words appear, and for review texts, reviews are mostly In one sentence or several phrases, the position of the feature words has a small impact on the core meaning of the review; the total text amount can be controlled by combining the reviews to ensure the efficiency of the calculation, so TF-IDF can effectively mine the feature information of the user's negative reviews .

3.3. Identification of User Needs Based on Negative Reviews

3.3.1 Vector Space Model of User Negative Comments

ISSN: 2664-9640

The computer cannot recognize the text information and needs to convert the text information into a vector space model (VSM) that the computer can recognize. The VSM model contains two elements, one is the feature word, and the other is the weight of the feature word. The choice of the feature word is based on the noun, verb, and adjective found by the TF-IDF method. The weight is the TF-IDF value of the feature. The vector form of the negative comment text $d_j$is as follows:

$$d_j = (w_{1,j}, w_{2,j}, w_{3,j}, \dots, w_{t,j})$$

The text-feature word matrix of negative review text can be expressed as follows:

$$\begin{bmatrix} - & w_1 & w_2 & \cdots & w_i \\ d_1 & w_{11} & w_{21} & \cdots & w_{i1} \\ d_2 & w_{12} & w_{22} & \dots & w_{i2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ d_j & w_{1j} & w_{2j} & \dots & w_{ij} \end{bmatrix}$$

The weight of each feature word in different texts is usually different. W_{ij} represents the weight of feature word i in text j. The weight value calculated in the TF-IDF method is used to lay the foundation for subsequent cluster analysis.

3.3.2 Cluster Analysis of User Negative Comments

Based on the user's negative comments, only the content of the comments, and no analysis criteria for negative comments, in order to be able to mine user needs, this paper uses the K-meams algorithm in SPSS to perform cluster analysis on the review text. K-means clustering algorithm is an iterative clustering algorithm. Its steps are to randomly select k objects as the initial cluster center, and then calculate the distance between each object and each seed cluster center. An object is assigned to its nearest cluster center. The cluster centers and the objects assigned to them represent a cluster. For each sample assigned, the clustering center of the cluster is recalculated based on the existing objects in the cluster. This process is repeated until a certain termination condition is met. The termination condition may be that no (or minimum number) objects are reassigned to different clusters, no (or minimum number) cluster centers change again, and the squared error and local minimum.

According to the core idea of K-means, the steps to transform into K-means algorithm are as follows:

In the first step, select the number k of categories to be clustered (such as k = 3 in the example above), and select k center points.

In the second step, for each sample point, find the closest center point (find the organization), and the point closest to the same center point is a class. This completes a cluster.

The third step is to determine whether the classification of the sample points before and after the clustering is the same. If they are the same, the algorithm terminates, otherwise it proceeds to the fourth step.

In the fourth step, for the sample points in each category, calculate the center points of these sample points as the new center points of the class, and continue to the second step.

Due to the randomness of the initial clustering center, the results may deviate severely from the global optimal classification. Therefore, in actual operations, often running the K-Means clustering algorithm multiple times to achieve different cluster centers as the initial center, thereby improving the accuracy of clustering.

3.3.3 Identification of user needs based on cluster analysis

Through cluster analysis of user negative comments, we can find the main factors of user complaints and mine user needs from them. After the cluster analysis, according to the feature words with a high sum of TF-IDFs of each type of reviews, and in conjunction with reading some reviews in this category at random, user needs can be inferred. At the same time, by comparing the number of reviews owned by each type of user needs, it is possible to infer the ranking of the degree of attention of these user needs, and the results can assist decision makers to make better rational allocation of resources when making plans and strategies.

4. Conclusion

This article describes the collection of e-commerce user reviews, mining user needs, and new product development processes centered on user needs at the theoretical and practical levels. The main conclusions of this article are as follows:

(1) It is feasible to mine user needs based on negative comments from e-commerce users. In this paper, crawler technology and text analysis technology are used in the field of user demand collection and identification. The feasibility of this method is proved by the principles in Chapter III and the practice in Chapter 5. On the other hand, most of the users who published negative reviews were innovative customers or leading customers. The negative reviews to some extent expressed the direction users want the product to develop, which is of great significance for companies to grasp and judge their future needs.

(2) It is feasible to drive new product development based on user needs. This article uses axiomatic design to map user needs to product functional requirements, and then uses cognitive map theory to allow users to participate in mapping product functional requirements into new product development schemes, which can provide enterprises with ideas for new product development.

References

[1] Zhang Z. Effective requirements development-A comparison of requirement selicitation techniques[J]. Software Quality Management XV: Software Quality inthe Knowledge Society, E. Berki, J. Nummenmaa, I. Sunley, M. Ross and G. Staples (Ed.) British Computer Society,2007:225-240.

[2] Leite J CS D P, Gilvaz A P P. Requirements elicitation driven by interviews: theuse ofviewpoints[C]. Proceedings of the 8th International Workshop on SoftwareSpecification and Design,1996:85.

[3] Paetsch F, Eberlein A, Maurer F. Requirements engineering and agile softwaredevelopment[C]. Enabling Technologies: Infrastructure for Collaborative Enterprises, 2003.WET ICE 2003. Proceedings. Twelfth IEEE InternationalWorkshops on,2003:308-313.

[4] Manski C F, Molinari F. Skip sequencing: A decision problem in questionnairedesign[J]. The annals of applied statistics,2008,2(1):264.