

Research Article

Consumer Demand Behavior Mining and Product Recommendation Based on Online Product Review Mining and Fuzzy Sets

Jia Zhuo 

School of Digital Economics, Luzhou Vocational and Technical College, Luzhou, Sichuan 646000, China

Correspondence should be addressed to Jia Zhuo; zhuojia@lzy.edu.cn

Received 9 August 2022; Revised 21 September 2022; Accepted 23 September 2022; Published 11 October 2022

Academic Editor: Zaoli Yang

Copyright © 2022 Jia Zhuo. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Consumer demand is the need for product characteristics expressed in their own words, which is the basis for producers to develop product recommendations. The extraction and analysis of consumer demands is the most critical input information in quality function deployment (QFD), which has a significant impact on the final prioritization of product technical features, product optimization, and subsequent configuration decisions in QFD, and is directly related to the success of product development. However, the traditional QFD approach to demand analysis lacks reliability and feasibility, and its application often requires time and labor costs that exceed the company's actual capabilities. Therefore, this paper uses online reviews as the data source and constructs a latent Dirichlet allocation (LDA) topic model based on fuzzy sets to explore the consumer demand information reflected in user reviews. We also introduce the concept of word vector to improve the LDA topic model and compare it with the traditional topic model to verify the performance of the model, so as to explore the consumer demand behavior more accurately and efficiently.

1. Introduction

In an increasingly competitive market, the consumer occupies an increasingly active position in the comparison of buyer and seller relationships. Only those companies that are able to capture consumer demand and improve consumer satisfaction while reducing production costs will be able to gain a larger consumer market and improve their competitiveness. In order to develop products that satisfy consumers, the quality function development model, which is based on the idea of configuring product functions to meet consumer demands, has been widely disseminated and applied. How to apply the theoretical quality function deployment (QFD) to actual production and how to effectively improve the popularity of new products in the market is the key to the successful application of QFD [1] and the basic requirement for developing the corresponding product recommendation strategies. However, due to the ambiguity, uncertainty, and imprecision of consumer demand information and product technical attributes, QFD often faces a fuzzy and uncertain environment. In the practical

application of QFD, it is difficult to handle the ambiguity and uncertainty of various input information due to the different judging criteria and the difficulty in testing consistency. These shortcomings seriously hinder the grasp of consumer demand behavior and also cause great trouble on how to make effective product recommendation.

In the context of big data era, different scholars have many different definitions of big data, but it is generally agreed that big data is a general term for information that is large in quantity, variety, speed, and value [2]. Big data generally originates from human beings and is the information about their behavior such as logging in, shopping, and commenting when they consume or browse websites online. Big data also serves human beings by data mining and analysis of massive data, so that the valuable information contained in the data can be extracted from the disorganized data, thus making data available for people. Data mining technology refers to the process of finding valuable hidden information from data by building algorithmic models [3]. Data mining is fact-oriented, and marketing, finance,

intelligent detection, and other fields can use data mining tools to study demand information and to obtain the real consumer relationship and behavior strategy, which is helpful for enterprises to grasp the market trend and make targeted strategies [4].

Considering the background of data mining, this paper attempts to use consumer comments in radio stations as the data source and use fuzzy set-based data mining technology to deeply analyze consumer behavior, so as to obtain the key input information in QFD and build QFD models to enrich the application and improvement of related research, thus providing technical support for consumer demand behavior mining and product recommendation.

2. Literature Review

We review the literature in terms of both consumer demand behavior and data mining techniques, with a view to providing theoretical support for the application of data mining techniques in online product review analysis.

2.1. Consumer Demand Behavior: Perspective from Quality Function Deployment. QFD plays an important role in understanding consumer demand behavior, which is defined as “the systematic calculation of the relationship between consumer demands and product characteristics, the decomposition of consumer demands into corresponding product technical characteristics, and then the distribution of the product production process to meet consumer demands through functional deployment, while achieving quality standards of the product” [5]. The most distinctive feature of QFD is that it requires companies to design and integrate technical features of the product while analyzing consumer demands, so that the product is preferred by consumers.

The house of quality (HoQ) is the core of QFD, which integrates consumer demands with technical characteristics of the product and calculates the degree of correlation between consumer demands and technical characteristics by creating a matrix to determine consumer satisfaction and improve the technical characteristics of the product. Normally, the HoQ is divided into left wall, roof, ceiling, room, and floor, and the details of the five sections are as follows. The left wall contains information about consumer demands, including consumer demand categories and consumer demand weights [6]. The corresponding consumer demands and the importance of each consumer demand can be filtered out through consumer research. The ceiling section is the technical characteristics of the product, which are related to consumer demands and can be described in verbal and quantitative form. The roof section is built on top of the ceiling and represents the autocorrelation between the technical characteristics of the product. Due to the differences in the level of technology and design methods in the product development process, the technical characteristics of the product may be included in each other, which needs to ensure the scientific rationality of the calculation. The room section represents the correlation matrix between the

consumer demands and the technical characteristics of the product. By establishing the correlation matrix, the final importance of each technical characteristic can be obtained, which provides important information for the extension of the QFD. The flooring part is calculated by integrating the abovementioned house of quality information to assign the importance of technical characteristics from the perspective of maximizing consumer satisfaction, which can be used as a reference for further product improvement [7].

2.1.1. Analysis of Consumer Demands and Determination of Weights. In the QFD model, consumer demand information is the primary key input information, which has a decisive influence on the subsequent product function development research. In order to accurately obtain the importance of consumer demands, companies can purposefully develop and design new products to improve consumer satisfaction by introducing new products, so as to occupy the market and improve their competitiveness.

In order to obtain consumer demands before product development, most of the studies use market research methods to collect consumer demand information. Usually, the initial consumer demands collected are fragmented and not easy to analyze. Hierarchical analysis or the Kano model is often used to classify and refine consumer demands. In determining the weight of consumer demands, scholars often use “necessity” or “importance” to express consumer demands, which is then converted into a precise value to calculate the importance of consumer demands. Traditionally, scaling is used to calculate the importance of consumer demands, where consumers are asked to use a numerical scale between 1 and 7 to determine the importance of their needs, and then arithmetic averaging is done. Other studies have used the AHP method to calculate consumer demand weights. When determining the importance of consumer demands, the traditional proportional scaling method is less accurate and less consistent, and the overall satisfaction is more inaccurate, while the AHP method can hierarchize the consumer demands and improve the accuracy of the calculation results, which is widely used in QFD studies [8].

It can be seen that the current research mainly focuses on the ambiguity of user descriptions as an important output of QFD, and most of the data sources are from survey methods.

2.1.2. Analysis of Technical Characteristics and Determination of Importance. After the information about consumer demands is input into the QFD, information about the product technical characteristics (TCs) is entered next. It is necessary to analyze the relationship between consumer requirement specification (CRs) and TCs. In the traditional quality function development method, the expert evaluation information is often used to express the correlation between CRs and TCs and each TC in a proportional scale. Similarly, researchers developed a multiobjective optimization model to obtain the functional relationship between CRs and TCs and build the correlation matrix [9]. However, the correlations between CRs and TCs and the autocorrelations

among TCs are often uncertain and ambiguous. Therefore, some studies developed an integrated two-step model based on fuzzy-AHP and TOPSIS methods to determine the correlations, first analyzing the structural components of the correlations using fuzzy hierarchical analysis and determining the magnitude of the correlations between CRs and the corresponding correlated TCs. The second part uses the TOPSIS method to calculate the final ranking of each TC. In addition, some studies used fuzzy linear regression for the calculation of the relationship matrix, using asymmetric triangular fuzzy numbers to describe the relationship coefficients and the optimal linear regression model to find the relationship matrix [10]. However, this linear regression method does not solve the problem of ambiguity in the relationship, and the relationship is more nonlinear. The neural network model is trained to adjust the activation function of the hidden layer neurons to obtain the mapping function that can reflect the relationship between consumer demand and product technical characteristics.

The analysis of the final priority of technical characteristics includes the following [11]. Firstly, we construct the HoQ model and input the consumer demands and their weights, the types of relevant technical characteristics that can satisfy these demands, the correlation matrix between consumer demands and product technical characteristics, and the autocorrelation between product characteristics. Then, the correlation between consumer demands and technical characteristics is further explored, taking into account the product positioning, enterprise resource allocation, and market competition. Then, we can classify the basic importance of TCs and consider the correlation mining methods such as TOPSIS, so that the QFD model can produce the final priority of TCs that can satisfy the consumer demands and improve the performance of the company to guide the production and development of new products [12]. It can be seen that the analysis of technical characteristics is mainly focused on the study of the relationship between consumer demands and technical characteristics and the study of the autocorrelation of technical characteristics.

2.1.3. QFD Methods in Fuzzy Environments. QFD can make effective planning in early product design to avoid rework and repetitive work in late product development, but most of the data available to companies are vague and imprecise [13]. Specifically, the following are manifested. (1) When users express their needs for products, the descriptions are generally less professional and more emotional, such as “nice,” “suitable,” “like,” and so on. (2) The importance of consumer demands is usually expressed in the natural language form of “must,” “average,” and “indifferent.” (3) The relationship matrix and the correlation matrix are obtained through subjective evaluation by the QFD expert team, and the results of the evaluation are mostly expressed as “strong relationship,” “weak relationship,” and so on. (4) The evaluation of the product by users and the evaluation of the technical characteristics of the product by engineers are generally

expressed in terms of “good,” “very good,” “better,” and so on. Most of the commonly used methods quantify these evaluation information by precise numbers, which do not reflect the complexity and ambiguity of natural language. Many of these expressions are qualitative, making it difficult to be precise about the evaluation. It is not reasonable to use only a definite number to represent such evaluation. Fuzzy theory can be applied to natural language information, and it can quantify the subjective natural language, so that the uncertainty of evaluation information can be more reasonably expressed.

It is difficult to precisely characterize text-based linguistic information due to the complexity of consumer demand. Therefore, most of the current studies use a quantitative form to numerically analyze the degree of user preferences and calculate the importance of different demands based on the results of users’ scoring of different types of demands [14]. Therefore, the semantic analysis method to deal with the ambiguity in QFD has become a new feasible idea [15]. Therefore, introducing the theory of fuzzy sets into QFD, i.e., quality function configuration in a fuzzy environment, has far-reaching significance for the theoretical research and practical application of QFD.

2.2. Data Mining and Online Review. In the context of big data era, discovering and utilizing competitive intelligence from massive amounts of data to support decision making is an important issue for businesses and individuals. The huge amount of online review data accumulated on the Internet is an important data source for consumers and businesses to obtain information. However, online reviews present a multisource heterogeneous distribution, huge data volume, and redundancy and do not support retrieval, which brings great challenges to the organization and utilization of online review information in the big data environment; moreover, users have different information needs, some users are concerned about price, some users are concerned about performance, and the current research related to online review display methods and review usefulness ranking is consistent with the default consumer demands without considering users’ personalized needs; in addition, the demand for intelligent information services in the era of artificial intelligence has become more and more urgent, and the existing research on online reviews is not deep enough and lacks systematic representation and modeling of online review.

The overall characteristics of online reviews are as follows: massive data, unstructured, and much mineable information. Different types and domains of comments have their own characteristics, and different data mining models need to be tailored to specifically address different comment corpora. Generally speaking, there are several text mining techniques as follows.

2.2.1. Word Splitting. Word splitting is the process of rearranging words to form word order according to certain norms [16]. Chinese word separation refers to the process of

dividing consecutive combinations of Chinese characters into words according to certain norms and guidelines. Chinese word separation is the basic work of text mining and one of the key technical steps of natural language processing. The result of word separation determines the results and quality of the subsequent analysis and is therefore extremely important. For Chinese, it is more difficult than for English because of the basic grammatical peculiarities, as well as the blurring of word boundaries, word ambiguities, unregistered words, missing separators, and so on. For example, a cell phone product review with the content "Huawei mate40 battery drains super fast" can be divided into "Huawei mate40," "battery," "drains," "super," and "fast."

Currently, there are three basic word separation algorithms: string matching-based word separation, comprehension-based word separation, and statistical-based word separation. According to whether the word separation process is combined with the lexical annotation process, they can be divided into simple word separation and combined word separation and annotation methods [17].

2.2.2. Text Clustering. Text clustering refers to the process of dividing text data into different classes or clusters. Specifically, text clustering can be explained as assuming that there are N texts in a dataset, and the dataset containing N texts is divided into M clusters according to a certain clustering rule, and the documents in the same cluster are similar, while the documents in different clusters have significant differences. The principle of text clustering is that the similarity of similar documents is greater and the similarity of different classes of documents is smaller. To determine the category of a text based on the relationship between content and structure of the text, it can help to explore the internal information of the text. Text clustering is divided into five steps: text pre-processing, text model representation, calculation of text similarity, selection of clustering algorithm, and information organization navigation. There are many kinds of algorithms for text clustering, and LDA topic model is a more commonly used clustering algorithm [18].

LDA topic model is a Bayesian generative model with three levels: feature words, topics, and documents. Among them, topics can be discovered through documents and individual documents can be annotated with topic tags. In general, the way the LDA topic model generates documents follows the following steps [19]. First, a topic is selected with a specific probability, and then a word is selected again with a specific probability under that topic, so that the first word of the document can be generated, and the above selection process is repeated N times to obtain the complete document, which contains N words. When applied in practice, the generation process of the LDA topic model is not taken seriously. Instead, its inverse process is more important, i.e., finding the topics contained in a complete document and the words contained under each topic.

In this paper, we adopt a semi-supervised learning method for feature extraction based on the LDA topic model, which is a bag-of-words model that considers a

paragraph to be composed of a set of words, where the order of words is arbitrary. The LDA topic model is to give the topic of each sentence with certain probability.

3. Methodology

The extraction and analysis of user demands is the most critical input information in QFD, which has an important impact on the final prioritization of product technical features, product optimization, and subsequent configuration decisions in QFD, and is directly related to the success of product development. We propose an integrated LDA-Word2vec to explore consumer demands by building a data mining model and feature engineering, using a combination of data mining and sentiment dictionary modeling methods to explore users' needs. In this sense, the topic clustering results are more accurate, the topic words are more comprehensive and the topic confusion is lower compared with the traditional topic model, and the above improvements are demonstrated by simulation studies.

3.1. Construction of LDA Topic Model. We adopt the LDA topic model to explore consumer demands, which is a document topic generation model that can effectively semantically analyze online review data to build a topic model containing various feature information, and then perform feature analysis to obtain user need information [20].

The topic model is the most used generative model in the field of text topic mining, which considers each document as a combination of probability distributions of multiple topics, each of which contains keywords that reflect the topic information. The model, consisting of three Bayesian models, builds two generative models, text-topic and topic-word, and describes the process of text-topic-word generation in the form of probability distributions. Figure 1 shows the structure of the topic model.

$\alpha \rightarrow \theta_m \rightarrow Z_{m,n}$ denotes that when the m th document is generated, the text-topic distribution m is first extracted, and then the topic number $Z_{m,n}$ is generated for the n th word in the document.

$\beta \rightarrow \Phi_k \rightarrow W_{m,n}$ denotes the selection of the topic-word probability distribution with number $k = Z_{m,n}$ to generate the word $W_{m,n}$, which is the n th word of the m th document.

$\alpha \rightarrow \theta_m$ is the Dirichlet distribution, $\theta_m \rightarrow Z_{m,n}$ is the polynomial distribution, and the overall composition of the Dirichlet-multinomial conjugate distribution, through the characteristics of the conjugate distribution, can be calculated to obtain the text-subject probability distribution as

$$p(\vec{z}|\vec{\alpha}) = \prod_{m=1}^M p(\vec{z}_m|\vec{\alpha}) = \prod_{m=1}^M \frac{\Delta(\vec{n}_m|\vec{\alpha})}{\Delta(\vec{\alpha})}. \quad (1)$$

In addition, since $\beta \rightarrow \varphi_k \rightarrow W_{m,n}|k = Z_{m,n}$ at this time, $\beta \rightarrow \varphi_k$ and $\varphi_k \rightarrow w_k$ are also a Dirichlet-multinomial conjugate structure. Thus, the probability distribution of the subject terms is obtained as follows:

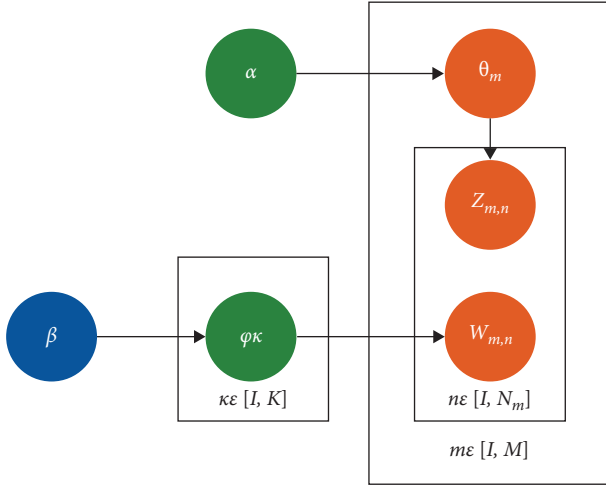


FIGURE 1: LDA topic model.

$$p(\vec{w}|\vec{z}, \vec{\beta}) = p(\vec{w}'|\vec{z}', \vec{\beta}),$$

$$\prod_{k=1}^K p(\vec{w}_{(k)}|\vec{\alpha}_{(k)}, \vec{\beta}) = \prod_{k=1}^K \frac{\Delta(\vec{n}_k + \vec{\beta})}{\Delta(\vec{\beta})}. \quad (2)$$

Combining the above equations, we obtain the joint distribution:

$$p(\vec{w}, \vec{z}, \vec{\alpha}, \vec{\beta}) = p(\vec{w}|\vec{z}, \vec{\beta})p(\vec{z}|\vec{\beta}),$$

$$p(\vec{w}|\vec{z}, \vec{\beta})p(\vec{z}|\vec{\beta}) = \prod_{k=1}^K \frac{\Delta(\vec{n}_k + \vec{\beta})}{\Delta(\vec{\beta})} \prod_{m=1}^M \frac{\Delta(\vec{n}_m + \vec{\alpha})}{\Delta(\vec{\alpha})}. \quad (3)$$

The Gibbs sampling algorithm is an algorithm used to construct a Markov chain that converges to a certain distribution and belongs to the MCMC algorithm. Suppose that the subject of the i_{th} word in corpus z is z_i , where $i = (m, n)$ corresponds to the n_{th} word of the m_{th} document, and $-i$ denotes the word with subscript. According to the Gibbs algorithm, we want to compute the conditional distribution corresponding to any coordinate axis i :

$$p(z_i = k|\vec{z}_{-i}, \vec{w}),$$

$$p(z_i = k, w_i = i|\vec{z}_{-i}, \vec{w}_{-i}). \quad (4)$$

The posterior distribution (θ_m and φ_k) is further obtained:

$$p(\vec{\theta}_m|\vec{z}_{-i}, \vec{w}_{-i}) = \text{Dir}(\vec{\theta}_m|\vec{n}_{m,-i} + \vec{\alpha}),$$

$$p(\vec{\varphi}_k|\vec{z}_{-i}, \vec{w}_{-i}) = \text{Dir}(\vec{\varphi}_k|\vec{n}_{k,-i} + \vec{\beta}). \quad (5)$$

We finally obtain the Gibbs formula for the LDA model:

$$p(z_i = k|\vec{z}_{-i}, \vec{w}) = \frac{n_{m,-i}^{(k)} + \alpha_k}{\sum_{k=1}^K (n_{m,-i}^{(k)} + \alpha_k)} \cdot \frac{n_{m,-i}^{(i)} + \beta_t}{\sum_{t=1}^V (n_{k,-i}^{(t)} + \beta_t)}. \quad (6)$$

According to the above formula, the LDA model can be trained by Gibbs sampling algorithm. ①Random initialization: each word w is randomly assigned a topic number z . ②For each word w , the topic is re-sampled and its topic number is updated. ③The above process is repeated until the predefined parameters converge. ④The topic word distribution matrix of the corpus is output. The final LDA topic model is obtained, which contains several topics and the corresponding topic words.

3.2. Improvement of Topic Model Based on Word2vec. LDA models have been applied to topic clustering and feature extraction; however, because there are a large number of similar words in Chinese, the traditional word sense expression models ignore the similarity between close words. On the other hand, the “bag-of-words model” in LDA considers that each word exists separately and ignores the association between words, which is an area where LDA models need to be improved. In order to improve the LDA model, this paper introduces the concept of word vectors and uses Word2vec, a neural network model, for word vectorization. By combining the LDA model and word vector model, the W-LDA model is constructed to extract topics and subject words by integrating the probability information of subject words and word sense correlation.

Unlike the LDA model, the Word2vec model allows two semantically similar words to have a higher similarity value in mathematical expression, takes into account the connection between sentence contexts, and provides more accurate semantic analysis. The network structure of Word2vec model is shown in Figure 2.

In order to further clarify the implied semantic relationships among the topic corpus, this paper uses Word2vec to transform the T-W distribution matrix into a T-WV (topic-word vector) distribution matrix by constructing word vectors, so as to obtain a more accurate topic distribution according to the quantized column order. The completed word vector model can be further used to expand the topic words and improve the distribution of topic words in each topic.

We constructs a W-LDA model integrating LDA topic model and Word2vec word vector model to perform topic clustering, and the whole process is shown in Figure 3.

The objective function of the CBOW model is as follows:

$$L = \sum_{w \in C} \log p(w|\text{Context}(w)), \quad (7)$$

where $p(w|\text{Context}(w))$ denotes the index of word w in the dictionary D .

Finally, Word2vec can be used to expand the subject words. The algorithm mainly uses the Word2vec model to obtain the vectorized output of the T-W matrix and sets a reasonable vector dimension according to the need. After obtaining the word vectors of all the corpus, the subject words in each topic are matched by the similarity calculation of the word vectors to obtain more subject words. This approach is feasible because it can retain the important semantic information in the topics obtained from the LDA

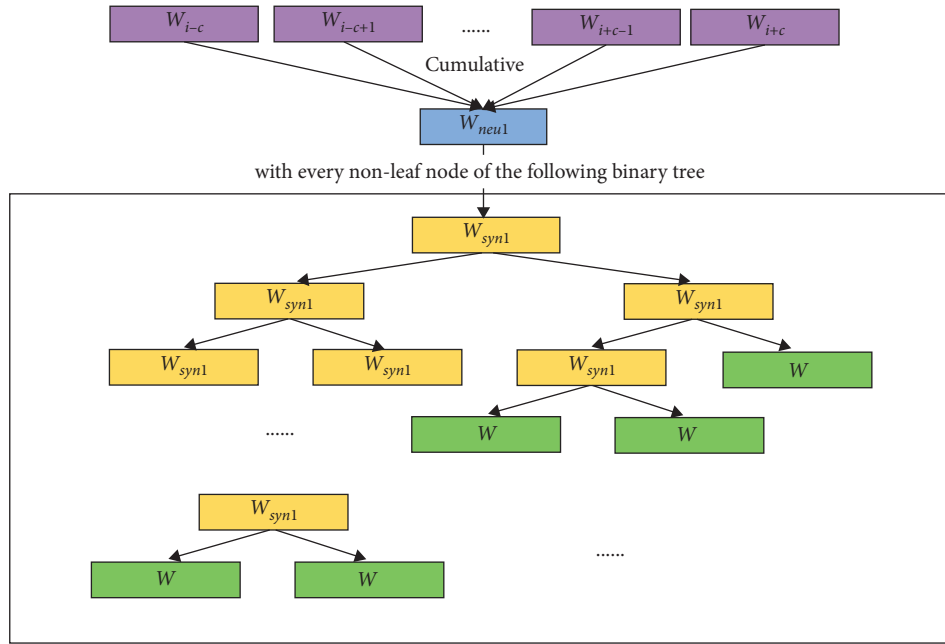


FIGURE 2: The network structure of Word2vec.

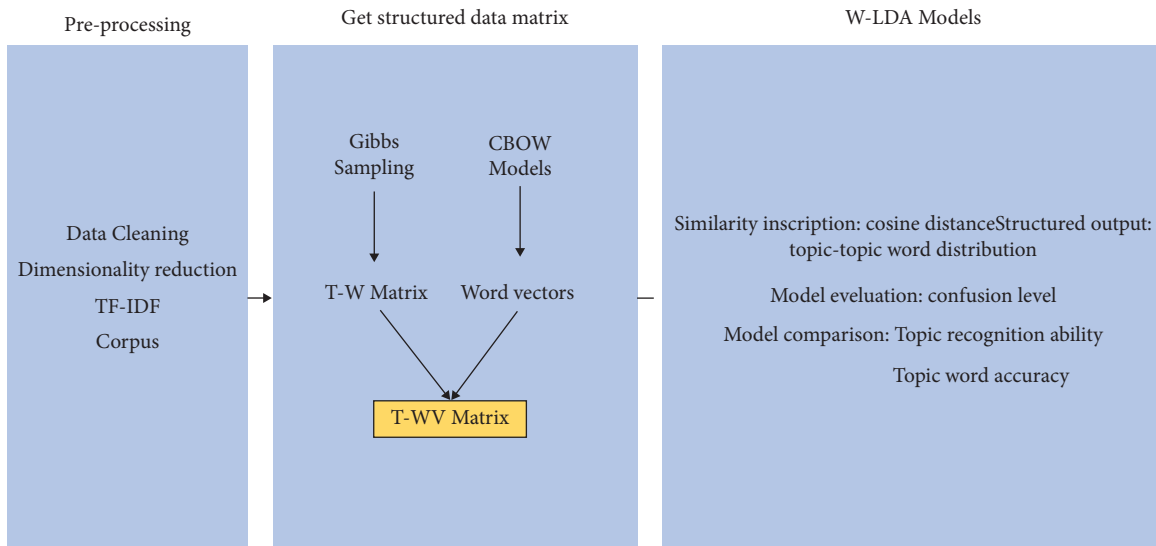


FIGURE 3: W-LDA algorithm flow.

model, keep the topic content unchanged, and significantly expand the topic information. In this way, more accurate topic clustering results are obtained.

We also use the perplexity index to verify the reliability of the model. The perplexity value of the model is the entropy value of the model, and the lower the perplexity of the model, the better the prediction of the sample. The perplexity of the LDA thematic model is calculated as follows:

$$\text{Perplexity}(D) = \exp \left\{ -\frac{\sum_{d=1}^M \log p(w_d)}{\sum_{d=1}^M N_d} \right\}, \quad (8)$$

where D denotes the test set in the corpus with M documents, N_d denotes the number of words in each document d , w_d denotes the words in document d , and $p(w_d)$ is the probability that word w_d is generated in the document. A comparison of the confusion level using the improved model and the traditional topic model is considered to analyze the differences between them.

This paper introduces the word vector model Word2vec on the basis of the traditional LDA topic model in order to overcome the problems of low classification accuracy and large topic word clustering error of the traditional theme

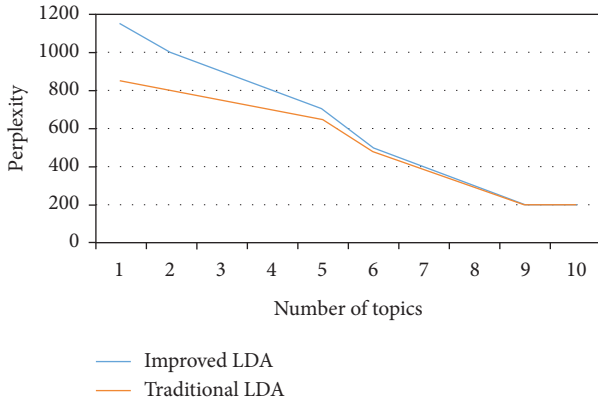


FIGURE 4: Topic model confusion comparison.

model, but the performance is subject to further experimental analysis.

4. Discussion and Results

We collected a total of more than 40,000 product reviews in the electronics category on <https://jingdong.com>, and after data preprocessing, we construct an improved W-LDA topic model to obtain topics and related topic words and compare it with the traditional LDA model to verify the effectiveness of the W-LDA model.

In the topic model, there are two key parameters: α and β . The choice of α is related to the number of topics; generally, $\alpha = 50/K$, K is the number of topics, and β is a fixed value of 0.01. By setting different K values, the optimal settings of α and β can be determined. Next, the lexicon required for the topic model is constructed for the text data, and the text data are vectorized and transformed using the annotation method. As many topics as possible can be selected to provide more space for the subsequent selection of topics. The initial topic distribution is obtained through model training by setting $K = 50$ and iterations = 200.

To verify the scientific validity of the improved model, the magnitude of perplexity is calculated separately for the traditional topic model and the improved topic model. The confusion values are calculated separately, and Figure 4 shows the comparison of the confusion degree between the improved LDA model and the traditional model.

The perplexity of the improved LDA model is always lower than that of the traditional LDA, indicating that the W-LDA model constructed in this paper has superior performance. In practical applications, the number of topics is usually chosen between 10 and 30.

In short, the Word2vec model can better express the correlation between words and reflect the uniqueness and professionalism of the corpus; more importantly, it can model the consumer preferences of different types of products and calculate the corresponding weights, which provides technical support for accurate product recommendation strategies.

In order to analyze the sentiment of consumer reviews to obtain quantitative consumer satisfaction and to determine the fuzzy set classification according to the change of

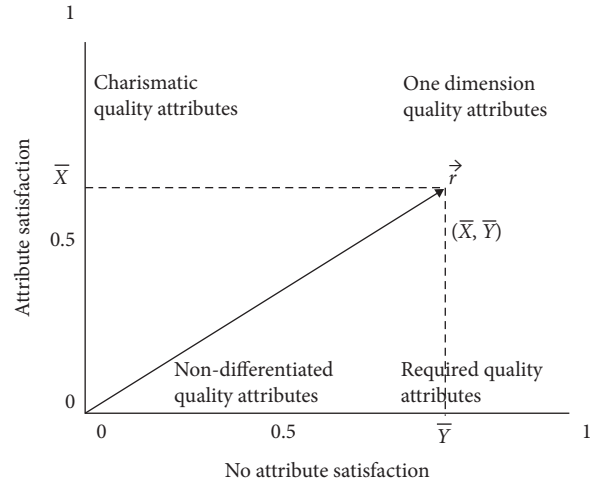


FIGURE 5: Consumer satisfaction/dissatisfaction vector based on fuzzy sets.

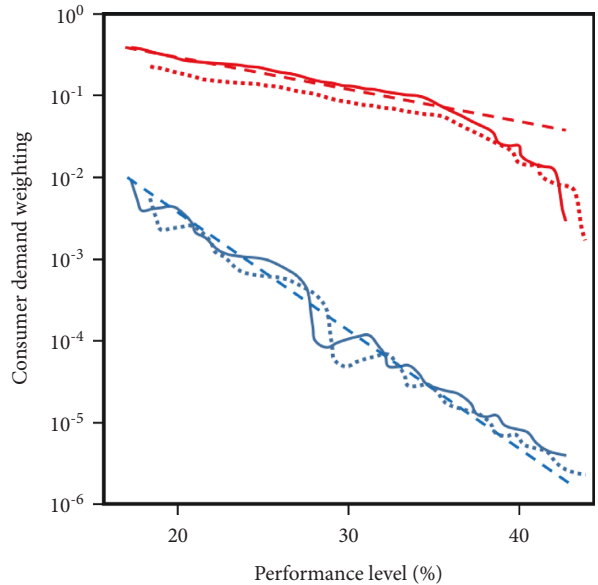


FIGURE 6: Low performance level.

satisfaction of different consumer demands, this paper constructs a fuzzy set model from the perspective of text-based consumer reviews. The vector model of consumer demand satisfaction based on fuzzy sets is shown in Figure 5.

Finally, we still used the same corpus to compare the fuzzy set-based consumer demand satisfaction vector model with the traditional vector model based on the consumer satisfaction model. Given the performance requirements of the fuzzy set-based model for the running system environment, we also examine the simulation of different models for consumer demand weights at different performance levels.

Figures 6 and 7 show the simulation of consumer demand weights at different performance levels. In this case, the red line represents the model proposed in this paper, while the blue line represents the traditional model. It can be found that the model proposed in this paper exhibits a

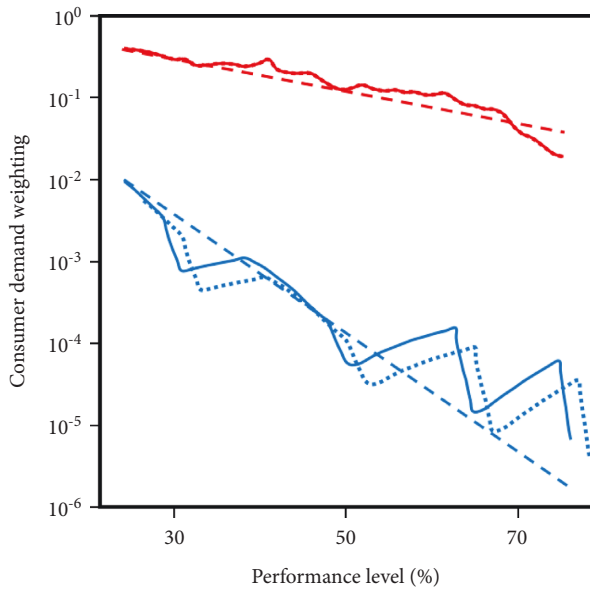


FIGURE 7: High performance level.

higher degree of fit, as well as stability, both at low performance levels and at high performance levels.

5. Conclusion

The development of artificial intelligence technology in the context of the big data era corresponds to the application of data mining technology. The innovation of data mining is achieved through deep learning of machines, which is mainly manifested in deep mining of complex, random and irregular data information, and then obtaining information with certain concealment and a large amount of valuable information clues. From the technical point of view, data mining technology can discover and explore the value and potential of information itself; from the business point of view, it can bring huge profits and economic benefits to enterprises, thus achieving efficient and stable development of enterprises. However, at present, the data source of consumer demand behavior is single, the amount of data is small, and the collection cost is high, which leads to the traditional data mining technology being more limited in this field.

Therefore, this paper analyzes consumer demand based on online reviews: in order to obtain qualitative demand information in QFD, this paper firstly takes online reviews as the data source from the perspective of consumers, introduces the concept of word vector, and constructs an LDA topic model based on Word2vec to extract user review information from the perspective of machine learning. Secondly, this paper adopts the fuzzy set method to construct a consumer satisfaction model and compare it with the traditional topic model to verify the performance improvement of the model and extract the customer demand information more accurately and efficiently. The findings not only extend the research related to QFD but also enrich the application of data mining technology in the field of

consumer demand behavior prediction and provide technical support for the formulation of product recommendation strategies.

In conclusion, this paper provides a new algorithm and model for the application of data mining technology in the field of product reviews, which is of great value for capturing changes in consumer demand behavior and formulating corresponding product recommendation strategies.

However, it should be noted that this paper also has the following shortcomings. First, the model proposed in this paper has only been tested on a small sample, and the sample scope can continue to be expanded in the future. Secondly, how to identify the usability and usefulness of online product reviews is an important direction for future research. Finally, the fuzzy set theory in the field of data mining with applications still has a lot of room for improvement, and future research can further consider the data mining potential of the sub-theory fuzzy lattice of fuzzy sets.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The author declares that there are no conflicts of interest.

References

- [1] L. Yang and S. Dong, "Rebate strategy to stimulate online customer reviews," *International Journal of Production Economics*, vol. 204, pp. 99–107, 2018.
- [2] X. Xu, "Examining the role of emotion in online consumer reviews of various attributes in the surprise box shopping model," *Decision Support Systems*, vol. 136, Article ID 113344, 2020.
- [3] X. Xu, "How do consumers in the sharing economy value sharing? Evidence from online reviews," *Decision Support Systems*, vol. 128, Article ID 113162, 2020.
- [4] C. Lee, X. Xun, and L. Chia-Chun, "Using online user-generated reviews to predict offline box-office sales and online DVD store sales in the O2O era," *Journal of Theoretical and Applied Electronic Commerce Research*, vol. 14, no. 1, pp. 0–83, 2019.
- [5] S. Geng, W. Li, X. Qu, and L. Chen, "Design for the pricing strategy of return-freight insurance based on online product reviews," *Electronic Commerce Research and Applications*, vol. 25, pp. 16–28, 2017.
- [6] P. Engström and E. Forsell, "Demand effects of consumers' stated and revealed preferences," *Journal of Economic Behavior & Organization*, vol. 150, pp. 43–61, 2018.
- [7] Y. Geum, R. Kwak, and Y. Park, "Modularizing services: a modified HoQ approach," *Computers & Industrial Engineering*, vol. 62, no. 2, pp. 579–590, 2012.
- [8] M. Nilashi, R. A. Abumalloh, A. Alghamdi et al., "What is the impact of service quality on customers' satisfaction during COVID-19 outbreak? New findings from online reviews analysis," *Telematics and Informatics*, vol. 64, Article ID 101693, 2021.
- [9] Y. Papanastasiou and N. Savva, "Dynamic pricing in the presence of social learning and strategic consumers," *Management Science*, vol. 63, no. 4, pp. 919–939, 2017.

- [10] F. R. Lima-Junior and L. C. R. Carpinetti, "A multicriteria approach based on Fuzzy QFD for choosing criteria for supplier selection," *Computers & Industrial Engineering*, vol. 101, pp. 269–285, 2016.
- [11] K. Song, H. Kim, J. Lee, and Y. G. Kim, "Dissecting movie performance across multiple distribution channels: an elastic justification theory perspective," *Telematics and Informatics*, vol. 35, no. 1, pp. 159–167, 2018.
- [12] K. Fritz, V. Schoenmueller, and M. Bruhn, "Authenticity in branding – exploring antecedents and consequences of brand authenticity," *European Journal of Marketing*, vol. 51, no. 2, pp. 324–348, 2017.
- [13] A. Ghose, P. G. Ipeirotis, and B. Li, "Designing ranking systems for hotels on travel search engines by mining user-generated and crowdsourced content," *Marketing Science*, vol. 31, no. 3, pp. 493–520, 2012.
- [14] M. Salehan and D. J. Kim, "Predicting the performance of online consumer reviews: a sentiment mining approach to big data analytics," *Decision Support Systems*, vol. 81, pp. 30–40, 2016.
- [15] H. Zhang, H. Rao, and J. Feng, "Product innovation based on online review data mining: a case study of Huawei phones," *Electronic Commerce Research*, vol. 18, no. 1, pp. 3–22, 2018.
- [16] S. S. Jia, "Behind the ratings: text mining of restaurant customers' online reviews," *International Journal of Market Research*, vol. 60, no. 6, pp. 561–572, 2018.
- [17] L. S. Guzman, V. Saquicela, E. A. Ordóñez, J. Vandewalle, and D. Cattrysse, "Literature review of data mining applications in academic libraries," *The Journal of Academic Librarianship*, vol. 41, no. 4, pp. 499–510, 2015.
- [18] A. Dutt, M. A. Ismail, and T. Herawan, "A systematic review on educational data mining," *IEEE Access*, vol. 5, pp. 15991–16005, 2017.
- [19] J. Qi, Z. Zhang, S. Jeon, and Y. Zhou, "Mining customer requirements from online reviews: a product improvement perspective," *Information & Management*, vol. 53, no. 8, pp. 951–963, 2016.
- [20] S. Moro, J. Esmerado, P. Ramos, and B. Alturas, "Evaluating a guest satisfaction model through data mining," *International Journal of Contemporary Hospitality Management*, vol. 32, no. 4, pp. 1523–1538, 2019.