Research Article

A Method of Product Selection Based on Online Reviews

Xia Liang,1 Jie Guo,1 Yan Sun,1 and Xiaoxiao Liu2

1School of Management Science and Engineering, Shandong University of Finance and Economics, Jinan 250014, China
2Faculty of Business, City University of Macau, Macau 999078, China

Correspondence should be addressed to Xia Liang; susanliangxia@163.com and Yan Sun; sunyanbjtu@163.com

Received 6 May 2021; Revised 26 July 2021; Accepted 14 August 2021; Published 3 September 2021

Academic Editor: Mario Muñoz-Organero

Copyright © 2021 Xia Liang et al. 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.

With the rapid development of information technology and market economy, global e-commerce platform develops rapidly. Recently, online reviews are widely available on e-commerce platforms to express customers’ experience of products. When ranking alternative products based on online reviews, how to make full use of the information in online reviews to represent the sentiment analysis results of online reviews is an important prerequisite for decision analysis. To this end, we propose a method for measuring the time utility and support utility of online reviews. Then a method for representing the sentiment analysis results of online reviews in the form of linguistic distribution is proposed. In addition, in view of the attributes and their weights being unknown, we propose a method for extracting product attributes from online reviews by using the Term Frequency-Inverse Document Frequency (TF-IDF) algorithm; and the objective weights of attributes are determined through the Criteria Importance through Intercriteria Correlation (CRITIC) method. Additionally, in order to highlight the differences between the alternatives, the roulette wheel selection algorithm is first used to randomly select product attributes. Then the alternative products can be ranked by the extended Multi-Attributive Border Approximation area Comparison (MABAC) method with mixed information. Finally, we illustrate the applicability of the proposed method through a case study of selecting a 5G mobile phone and simulation experiment.

1. Introduction

With the continuous updating and iteration of network technology, information technology, and communication technology, we have gradually stepped into the era of big data. Against this background, the e-commerce has also developed rapidly. Many e-commerce platforms increasingly support consumers to make online reviews on the products they have purchased [1–3]. Decision analysis based on online reviews has generated considerable recent research interest. The existing researches have shown that online reviews have important reference value for product selection, especially when consumers are buying products with higher prices [4, 5]. Online reviews contain a large amount of valuable information, such as customer sentiment, customer preferences, and customer opinions [6–10]. Compared with the questionnaire survey, online reviews have many advantages. For example, online reviews are generated spontaneously and are easy to obtain. Furthermore, online reviews have a wide coverage and a large number of advantages. Online reviews have been successfully applied to various management decision analyses: product sequencing [11–15], product recommendation [16], product improvement [17], market structure analysis [18], brand analysis [19, 20], customer preference analysis, customer satisfaction survey [7, 20–23], and so on. Therefore, when the consumer has a desire to purchase a product, they can browse online reviews of alternative products on pertinent websites. For instance, if one consumer plans to purchase a 5G mobile phone, the consumer can thumb through the online reviews of acceptable 5G mobile phones through the shopping website to learn about them. However, due to the Internet technology and e-commerce platforms rapidly developing in recent years, the number of online reviews on each e-commerce platform has increased exponentially. Hence, it is very difficult for consumers to browse total online reviews and make decisions. Therefore, in order to provide valuable
suggestions for consumers to select products, it is essential to propose a product selection method based on online reviews.

The complete product ranking research on the basis of online reviews mainly consists of two parts. One is the analysis and processing of online reviews; another part is the product ranking method. The analysis and processing of online reviews includes three tasks: product attributes extracted from online reviews, sentiment analysis of online reviews, and expression of sentiment analysis results of online reviews. In the past few years, sentiment analysis has received a lot of scholarly attention due to its wide range of applications. As a result, many sentiment analysis techniques have been proposed [24–34]. By analyzing the existing sentiment analysis techniques, it can be found that sentiment analysis methods mainly include machine learning-based sentiment analysis techniques [24–28] and lexicon-based sentiment analysis techniques [29–34]. Lexicon-based sentiment analysis techniques are more suitable for online reviews sentiment analysis, while lexicon-based emotional analysis is divided into dictionary-based emotional analysis and corpus-based emotional analysis [35]. In this paper, the dictionary-based sentiment analysis technology is selected to conduct sentiment analysis of online reviews.

Since the different forms of information expression reflect different accuracy of information, the expression form of sentiment analysis results for online reviews is another key to select product. With the continuous introduction of information forms, the sentiment analysis results for online reviews are gradually being expressed in the forms of fuzzy numbers [11, 12, 36], linguistic distribution (LD) [37], hesitant fuzzy sets (HFS) [38], and so forth. After analyzing and processing online reviews, the alternatives were ranked according to the sentiment analysis results of online reviews. Frequently used methods are preference ranking organization methods for enrichment of evaluations II (PROMETHEE II) [12], technique for order preference by similarity to an ideal solution (TOPSIS) [36], VIKOR [39], TODIM [40], and other methods to rank products [41]. In addition, some scholars have offered some new perspectives on the issue of product selection based on online reviews. Fu et al. put forward a method of product ranking by combining feature opinion mining with interval ratio Pythagorean fuzzy sets [42]. Li et al. put forward a product selection method which considers the combination of consumer expectation and online reviews [43].

Judging from the existing research results, although the research on sentiment analysis of online reviews and alternative products ranking has achieved preliminary results, there are still some problems that deserve further research. First, most of the existing product selection methods based on online reviews set the product attributes and their weights in advance. There are fewer methods for product selection based on online reviews with unknown attributes and weights. Then, when dealing with the online reviews, the reference values of online reviews are seldom fully utilized. In addition, it is easy to overlook subtle differences between products in existing methods, yet such subtle differences are likely to be critical in product selection. Therefore, it is necessary to carry out further research.

To solve the above problems, the aim of this study is to come up with a product selection method applicable to unknown product attributes and weights. This method not only maximizes the reference value of online reviews but also provides reasonable ranking results. The main contributions of this study are as follows:

1. To take full advantage of the reference value of online reviews, this paper considers the time utility and support utility of online reviews. Based on statistical theory, the sentiment analysis results of online reviews are processed into linguistic distribution (LD).
2. The TF-IDF algorithm is used to extract the attributes that consumers care about from online reviews. Based on this, we used the CRITIC method to determine the attribute objective weights.
3. The roulette wheel selection algorithm is used to randomly select the attribute values of the products so that the differences between the alternatives are retained. Then, the ranking result of the alternative products is obtained by using extended MABAC method with mix information.

The structure of this paper is arranged as follows. In Section 2, we introduce the relevant theoretical knowledge of linguistic distribution, roulette wheel selection algorithm, and MABAC method. In Section 3, the specific steps of the product ranking method through online reviews are presented in detail. In Section 4, the method proposed in this paper is used to select 5G mobile phones so as to illustrate the practicability of this method. Section 5 is presents conclusion of the study.

2. Related Works

In this part, some theoretical foundations used later are introduced, such as linguistic distribution (LD), roulette wheel selection algorithm, and MABAC method.

2.1. Linguistic Distribution. Since online reviews data is textual data, converting textual data to real numbers is difficult. Linguistic terms are more in line with human language expression. Hence, linguistic terms are accurate expression of the emotional information contained in consumer online reviews. As the emotions contained in online reviews are complex, it is difficult to express them accurately in a linguistic term. Hesitant fuzzy linguistic term set (HFLTS) allows decision-makers to use more than one linguistic term to express their views. The concept of HFLTS was first put forward by Rodriguez et al. [44]. However, the weight of each linguistic term is the same in HFLTS. Decision-makers may have different preferences for different linguistic terms in realistic evaluation. To this end, the concept of distribution assessments in a linguistic term set was proposed by Zhang et al. [45]. LD can describe the distributed preference of decision-makers as distributed evaluation by providing certain symbol proportion
information in linguistic terms [46]. Additionally, LD can be used to indicate the evaluation of the evaluation object. Due to the fact that each product contains a lot of consumer online reviews, different consumers have different emotional tendencies for each attribute of the product. Since LD can contain many messages, LD can convey actual information expressed by consumer to the maximum extent [37].

Xu [47] proposed a linguistic symbol calculation model for virtual linguistic terms. Then Dong et al. [48] called the linguistic symbol calculation model a continuous linguistic model.

Definition 1 (see [47]). Let \( S = \{S_\alpha \mid \alpha = -\tau, \ldots, 0, 1, \ldots, \tau \} \) be a linguistic term set. There are some characteristics for the linguistic term set which should satisfy the following:

1. The set is ordered: \( S_\alpha > S_\beta \) only if \( \alpha > \beta \)
2. There is a negation operator: \( \text{Neg}(S_\alpha) = S_{-\alpha} \)

Considering any two linguistic terms \( S_\alpha, S_\beta \in S \) and \( \mu, \mu_1, \mu_2 \in [0, 1] \), Xu introduces some operating rules as follows:

\[
\begin{align*}
(1) & \quad S_\alpha \oplus S_\beta = S_{\alpha \oplus \beta} \\
(2) & \quad S_\alpha \otimes S_\beta = S_{\alpha \otimes \beta} \\
(3) & \quad \mu S_\alpha = S_{\mu \alpha} \\
(4) & \quad (\mu_1 + \mu_2) S_\alpha = \mu_1 S_\alpha \oplus \mu_2 S_\alpha \\
(5) & \quad (\mu_1 \cdot \mu_2) S_\alpha = \mu_1 S_\alpha \otimes \mu_2 S_\alpha
\end{align*}
\]

Definition 2 (see [45]). Let \( S \) be a linguistic term set, and \( L = \{ (S_\alpha, P_\alpha) \mid \alpha = -\tau, \ldots, 0, \ldots, \tau \} \), where \( S_\alpha \in S \), \( P_\alpha \geq 0 \), \( \sum_\alpha P_\alpha = 1 \), and \( P_\alpha \) is the proportion of \( S_\alpha \). Then, \( (S_\alpha, P_\alpha) \) is a proportional linguistic pair, so \( L \) is called a distribution assessment of \( S \).

\[
d(L_1, L_2) = \frac{1}{2(\max_{ij,y} y_{ij} - \min_{ij,y} y_{ij})} \left[ P(L_1) - P(L_2) \right]^T \left[ P(L_1) - P(L_2) \right],
\]

where \( P(L_k) = [P_{-\tau,k}, \ldots, P_{0,k}, \ldots, P_{\tau,k}]^T \), \( k = 1, 2 \), and matrix \( Y \) is defined as \( Y = [y_{ij}]_{T \times T} \), \( y_{ij} = [\Delta^{-1}(S_i)]^T[S_j] - \Delta^{-1}(S_i) \), and \( \Delta^{-1}(S_i) = i + 1 \). Matrix \( Y \) is a positive symmetric matrix. Apparently, the greater the distance between \( L_1 \) and \( L_2 \), the smaller the value of \( y_{ij} \).

Example 1. Let \( S \) be the linguistic term set: \( S = \{S_{-3} = \text{very bad}, S_{-2} = \text{bad}, S_{-1} = \text{slightly bad}, S_0 = \text{fair}, S_1 = \text{slightly good}, S_2 = \text{good}, S_3 = \text{very good} \} \). The football coach evaluated a player’s performance in ten matches. The coach thought that the player’s performance in 3 matches was slightly good, his performance in 5 matches was good, and his performance in 2 matches was very good. Therefore, it can be expressed in LD as \( L = \{(S_1, 0.3), (S_2, 0.5), (S_3, 0.2)\} \).

Example 2. There are 100 consumers to evaluate the engine performance of a certain car. There are 55 consumers thinking the engine performance of the car is slightly good; there are 20 consumers thinking the engine performance of the car is good; and the remaining consumers did not express their opinions. Hence, this can be expressed in LD as \( L = \{(S_0, 0.25), (S_1, 0.55), (S_2, 0.2)\} \).

Definition 3 (see [45]). Let \( L = \{(S_\alpha, P_\alpha) \mid S_\alpha \in S, \alpha = -\tau, \ldots, 0, \ldots, \tau \} \) be an LD and let \( \alpha \) be the subscript of linguistic term \( S_\alpha \). Then the expectation of linguistic distribution \( L \) is defined as follows:

\[
E(L) = \sum_{\alpha = -\tau}^{\tau} \alpha P_\alpha.
\]

For two LDs \( L_1 \) and \( L_2 \), if \( E(L_1) > E(L_2) \), then \( L_1 \) is superior to \( L_2 \), expressed as \( L_1 > L_2 \); if \( E(L_1) < E(L_2) \), then \( L_1 \) is inferior to \( L_2 \), expressed as \( L_1 < L_2 \). However, if \( E(L_1) = E(L_2) \), then we cannot distinguish two LDs if they have the same expectation.

Definition 4 (see [49]). Let \( L_1 \) and \( L_2 \) be two LDs; the distance between \( L_1 \) and \( L_2 \) is defined as follows:

\[
\begin{align*}
W(L_1, L_2, \ldots, L_n) &= \left\{ S_{-\tau}, \ldots, S_0, \ldots, S_{\tau} \right\} \\
&= \left\{ \left( S_{-\tau} \sum_{k=1}^{n} \frac{1}{n} P_{-\tau,k} \right), \ldots, \left( S_0 \sum_{k=1}^{n} \frac{1}{n} P_{0,k} \right), \ldots, \left( S_{\tau} \sum_{k=1}^{n} \frac{1}{n} P_{\tau,k} \right) \right\}
\end{align*}
\]

(3) Degeneration: if \( L_1 \) and \( L_2 \) are two special LDs, that is, two linguistic terms, then the distance between \( L_1 \) and \( L_2 \) will be reduced to Euclidean distance of subscripts of linguistic terms.
2.2. Roulette Wheel Selection Algorithm. Roulette wheel selection algorithm is an algorithm for individual selection according to fitness [50]. The probability of an individual being selected is proportional to its fitness. This mechanism prevents individuals with small fitness values from being eliminated directly. When the probability of an individual is higher, it is easier to be selected.

Assume that there are m individuals; \( X = \{X_1, X_2, \ldots, X_m\} \), the probability of each individual appearing is \( P = \{P_1, P_2, \ldots, P_m\} \), and \( \sum_{i=1}^{m} P_i = 1 \). The roulette wheel selection model diagram is shown in Figure 1.

In Figure 1, \( P_i \) is the probability that individual \( X_i \) will be selected. The greater probability \( P_i \) that an individual will be selected is, the larger the area of the fan-shaped area obtained on the roulette wheel is. When the wheel rotation stops, the area pointed to by the red pointer is the selected area. So, the selection process of the wheel disc is only related to the probability of each individual appearing. The detailed algorithm steps are as follows:

1. **Step 1**: the cumulative probability \( s_i \) of individual \( X_i \) is calculated by the following equation:
   \[
   s_i = \sum_{k=1}^{i} P_k, \quad i = 1, 2, \ldots, m. \tag{4}
   \]

2. **Step 2**: randomly generate a random number \( r \) between 0 and 1.
3. **Step 3**: let \( i = 1 \). Compare the values of \( r \) and \( s_i \); if \( r \leq s_i \), individual \( X_i \) is selected and the selection process is ended; otherwise, go to the next step.
4. **Step 4**: let \( i = i + 1 \); go to step 3.

2.3. Multiattributive Border Approximation Area Comparison Method. The MABAC method is one of the most accurate multiattributive decision-making (MADM) methods, which makes the information as accurate as possible by calculating the close coefficient value [51]. The MABAC method has been introduced by Pamučar and Ćirović [52], the main advantages of which are simple calculation and lasting solution. Considering the potential gains and losses, the final results are extensive. It has simple calculation process, systematic processing process, and reasonable logic. Much research in recent years has focused on solving decision-making problems by the MABAC method [53–57]. The special steps of the MABAC method are as follows:

1. **Step 1**: the initial decision matrix \( X = (x_{ij})_{m \times n} \) is obtained, where \( x_{ij} \) is the evaluation value of the \( i \)th alternative related to the \( j \)th attribute: \( i = 1, 2, \ldots, m \) and \( j = 1, 2, \ldots, n \).
2. **Step 2**: the initial decision matrix \( X = (x_{ij})_{m \times n} \) should be normalized into the normalized matrix \( N = (n_{ij})_{m \times n} \). The normalization formula of attributes with different properties is different.

When the attribute is benefit-type attribute, it means that the bigger the attribute value is, the better the alternative is. The normalization formula is

\[
N_{ij} = \frac{x_{ij} - x_{\max}}{x_{\max} - x_{\min}}, \quad i = 1, 2, \ldots, m, j = 1, 2, \ldots, n. \tag{5}
\]

When the attribute is a cost-type attribute, it means that the smaller the attribute value is, the better the alternative is. The normalization formula is

\[
N_{ij} = \frac{x_{\min} - x_{ij}}{x_{\max} - x_{\min}}, \quad i = 1, 2, \ldots, m, j = 1, 2, \ldots, n. \tag{6}
\]

3. **Step 3**: the element of weight matrix \( V = (v_{ij})_{m \times n} \) can be calculated by the following formula:

\[
v_{ij} = w_j \left( n_{ij} + 1 \right), \quad i = 1, 2, \ldots, m, j = 1, 2, \ldots, n. \tag{7}
\]

where \( w_j \) is the attribute weight.

4. **Step 4**: the border approximation area matrix \( G = [g_1, g_2, \ldots, g_n] \) needs to be determined, where the element of the border approximation area matrix \( G \) can be calculated by the following equation:

\[
g_{ij} = \left( \prod_{i=1}^{m} v_{ij} \right)^{1/m}, \quad j = 1, 2, \ldots, n. \tag{8}
\]

5. **Step 5**: we can construct the distance matrix \( Q = (q_{ij})_{m \times n} \) between alternative and the border approximation area, where the element of the distance matrix can be calculated by the following formula:

\[
q_{ij} = v_{ij} - g_{j}, \quad i = 1, 2, \ldots, m, j = 1, 2, \ldots, n. \tag{9}
\]

As shown in Figure 2, alternative products \( X_i \) may belong to the lower approximation area (\( G^{-} \), the
border approximation area \( (G^-) \), or the upper approximation area \( (G^+) \). The attribution of the approximate area \( (G^+, G, G^-) \) to the alternative product \( X_i \) is determined according to the following equation:

\[
X_i \in \begin{cases} 
G^+, & \text{if } q_{ij} > 0, \\
G, & \text{if } q_{ij} = 0, \quad i = 1, 2, \ldots, m, \; j = 1, 2, \ldots, n. \\
G^-, & \text{if } q_{ij} < 0,
\end{cases}
\]

In order for alternative product \( X_i \) to be selected as the best one in alternative products set, it is necessary for alternative product \( X_i \) to belong to the upper approximation area \( (G^+) \) under as many attributes as possible.

Step 6: by the value of \( S_i \), the ranking result of the alternatives is gained.

\[
S_i = \sum_{j=1}^{n} q_{ij}, \quad i = 1, 2, \ldots, m, \; j = 1, 2, \ldots, n.
\]

\( S_i \) is the comprehensive score of alternative products \( X_i \); the larger the value of \( S_i \) is, the better the alternative product \( X_i \) is.

### 3. Product Ranking Method Based on Online Reviews

This section mainly gives the concrete operation of the method for product selection based on online reviews, which is divided into 3 sections.

#### 3.1. Problem Description

Figure 3 describes the specific process of product ranking based on online reviews. As Figure 3 presents, the consumer plans to purchase a product from his/her acceptable set of alternative products. In order to select the most ideal products from several alternative products, it will be helpful to obtain product rankings by analyzing relevant online reviews. Since the online reviews are mass, scattered, and trivial, consumers cannot quickly determine product attributes based on online reviews. It is necessary to extract product attributes and calculate product attribute weights through online reviews of products. On this basis, alternative products can be ranked through the product ranking method. The obtained ranking results can help consumers to select alternatives.

For the convenience of the following analysis, the following symbols are used to represent the sets or variables that will be covered in the remainder of this paper.

\[
X = \{X_1, X_2, \ldots, X_m\}: \text{the set of } m \text{ alternative products which meet consumer requirements, where } X_i \text{ represents the } i\text{th alternative product, } i = 1, 2, \ldots, m
\]

\[
A = \{A_1, A_2, \ldots, A_n\}: \text{the set of } n \text{ product attributes extracted from online reviews, in which } A_j \text{ represents the } j\text{th product attribute, } j = 1, 2, \ldots, n
\]

\[
w = (w_1, w_2, \ldots, w_n)^T: \text{the product attribute weights, in which } w_j \text{ represents the weight of product attribute } A_j, \text{ and } \sum_{j=1}^{n} w_j = 1, \; w_j \geq 0, \; j = 1, 2, \ldots, n
\]

\[
K = (K_1, K_1, \ldots, K_m)^T: \text{the number of online reviews for each alternative product, in which } K_j \text{ represents the number of online reviews on alternative product } X_i, \; i = 1, 2, \ldots, m
\]

\[
R_{ik} = [R_{ik}^1, R_{ik}^2, \ldots, R_{ik}^n]: \text{the } k\text{th online review on the alternative product } X_i, \text{ in which } R_{ik} \text{ represents the online review related to attribute } A_j \text{ in } k\text{th online review on the alternative product } X_i, \; i = 1, 2, \ldots, m, \quad j = 1, 2, \ldots, n, \quad k = 1, 2, \ldots, K_i
\]

Let \( Ew_{ik} \) be the emotion word related to attribute \( A_j \) in the \( k\)th online review on the alternative product \( X_i \).

#### 3.2. Online Reviews of Product Selection Methods

This section puts forward a product selection model based on online review, which mainly includes 5 parts.

##### 3.2.1. Extract Attributes

(1) Data Preprocessing. Data preprocessing is mainly divided into the two following steps: (1) word segmentation and part-of-speech tagging and (2) deletion of stop word. Firstly, the NLPRI-ICTCLAS Chinese word segmentation system helps us to implement word segmentation and part-of-speech tagging in online reviews. According to the grammar of sentences or the context of online reviews, every sentence in online reviews is divided into words or technical terms, which are marked as nouns, verbs, adjectives, adverbs, and so forth. Secondly, stop words are those words that often appear in online reviews but have no practical significance [48]. To improve the efficiency of sentiment analysis, stop words need to be removed.

(2) Extract Attributes. Attributes are essential when making product selection. For products that consumers are not familiar with, it is difficult for consumers to clearly state every attribute of the products. Therefore, it is necessary to
extract product attributes that most consumers are more concerned about from online reviews of products. The extraction of alternative product attributes is essentially the extraction of feature words, which are then manually classified to obtain the product attributes. There are many methods for feature word extraction [58, 59], and the most common ones are Word2vec, word frequency statistics, Term Frequency-Inverse Document Frequency (TF-IDF), latent Dirichlet allocation (LDA), and so forth. Then, we adopt TF-IDF algorithm to extract the attributes contained in consumers’ online reviews. TF-IDF is on the basis of the bag-of-word mode, where a document can be expressed by a set of words used in the document. The specific algorithm steps are as follows:

\[
TF = \frac{N_w}{N},
\]

where \(N_w\) represents the number of times a word appears in a document and \(N\) represents the total words contained in the document.

\[
IDF = \log \left( \frac{M}{DF + 1} \right).
\]

In the above formula, \(M\) represents the total number of documents in corpus and \(DF\) represents the total number of documents containing the word.

\[
TF - IDF = TF \times IDF.
\]

According to the above formula, if a word is important to this document, it needs a large value of TF and a small value of DF. For instance, prepositions such as “in” and “on” and conjunctions such as “and,” “or,” “also,” and “then” often appear in documents. As they are common to all documents, they are not considered important terms in this document [59].

3.2.2. Utility Values of Online Reviews. In order to maximize the reference value of online reviews, we considered the utility of online reviews. The utility of online reviews is mainly measured by the time the online reviews are published and the degree of support of the online reviews.

Some electronic products continue to be updated and upgraded by merchants after their launch. Hence the reference value of the online reviews for different dates is different. As a result, the reference value of online reviews that are released closer to the time is higher. On the contrary, the reference value of online reviews that are released farther away is lower.

Suppose that \(T_k^i\) indicates the release time of the \(k\)th online reviews for alternative \(A_i\), \(T_{first}^i\) indicates the product publishing time of alternative \(A_i\), and \(T_{now}^i\) denotes the time to get online reviews. We can obtain the time utility values of online reviews:

\[
t_{ik} = \frac{T_k^i - T_{first}^i}{T_{now}^i - T_{first}^i}, \quad i = 1, 2, \ldots, m, k = 1, 2, \ldots, K_i.
\]

The number of “likes” provided by consumers’ online reviews reflects the usefulness of the online reviews to other consumers. When a consumer browses an online review with same views, he or she will hit the “like” button of the online review. Thus, the reference value of an online review is positively correlated with the number of “likes” an online review receives.

Let \(L_k^i\) indicate the number of “likes” received for the \(k\)th online reviews on alternative \(A_i\). Then, the support utility value of online reviews is

\[
l_{ik} = \log(l_k^i + 1) + 1, \quad i = 1, 2, \ldots, m, k = 1, 2, \ldots, K_i.
\]
Let $u_{ik}$ indicate the total utility of the $k$th online reviews on alternative $A_i$. On the basis of obtaining the time utility $t_{ik}$ and the support utility $l_{ik}$ of online reviews, the total utility of online reviews can be calculated by the following formula:

$$ u_{ik} = l_{ik} \times t_{ik}, \quad i = 1, 2, \ldots, m, k = 1, 2, \ldots, K_i. \quad (17) $$

### 3.2.3. Sentiment Analysis

(1) **Constructing Emotion Dictionary.** In this section, based on HowNet dictionary, an emotion dictionary which belongs to alternative products is established. Since the HowNet dictionary contains basic emotional words, there are limitations in performing sentiment analysis of online reviews of specific products. A specific emotion dictionary needs to be constructed based on the product being analyzed. Therefore, based on HowNet dictionary, combined with the emotion words contained in online reviews of products, we construct an exclusive emotion dictionary of alternative products. Let $E_{ik}$ represent the set of emotional words in the HowNet dictionary, let $E_{ik}$ represent the set of emotional words contained in the online reviews, and let $E_{ik}$ represent the set of emotion dictionaries of exclusive alternative products. Then, $E_{ik}$ is calculated as follows:

$$ E_{ik} = E_{ik} \cup E_{ik}, \quad (18) $$

where $E_{ik} = E_{ik} \cup E_{ik} \cup \cdots \cup E_{ik}$ and $E_{ik} = E_{ik} \cup \cdots \cup E_{ik} \cup E_{ik}$. The emotional value of the emotional word $E_{ik}$ needs to be constructed based on the product being analyzed. Therefore, based on HowNet dictionary, the emotional value of the emotional word $E_{ik}$ can be quantified by

$$ E_{ik} = \begin{cases} 1, & \text{if } E_{ik} \in E^+, \\ 0, & \text{if } E_{ik} \in E', \\ -1, & \text{if } E_{ik} \in E^-. \end{cases} \quad (19) $$

(2) **The Treatment of Degree Adverbs and Negative Words.** The treatment of degree adverbs and negative words is a critical step in the process of emotion analysis. Degree adverbs can strengthen or weaken the emotional intensity of emotional words, while negative words can change the emotional tendency. In this study, some common degree adverbs are selected from the HowNet dictionary and they are divided into two levels according to the different degrees expressed. Let $D(E_{ik})$ be the degree of hierarchy, and $D(E_{ik}) = 1$ or 2; the common degree adverbs and their corresponding scores are shown in Table 1.

Let $N(E_{ik})$ be the number of negative words. The deal with negative words is crucial, because negative words can change the emotional tendency. There are two possibilities for the appearance of negative words: (1) When the number of negative words in front of emotional words is even, this indicates the affirmation of emotional words. The emotional tendency of the emotional words does not change. (2) When the number of negative words in front of the emotional words is odd, this indicates the negation of the emotional words. The emotional tendency of the emotional words changes.

### 3.2.4. Construction of Linguistic Distributions

Inspired by Bi et al. [11] and based on mathematical statistics, we use LD to express the results of online reviews sentiment analysis.

We suppose that $a_{ik}^j$ indicates the subscript of linguistic term $S_{ik}$, which is calculated by the following formula:

$$ a_{ik}^j = (-1)^{N(E_{ik})} \times D(E_{ik}) \times E_{ik}. \quad (20) $$

From the above equation, we can see that $a_{ik}^j$ is an integer and $a_{ik}^j \in [-2, 2]$, so the LTS is $S = \{S_{i1}, S_{i2}, S_{i3}, S_{i4}, S_{i5}, S_{i6}\}$.

Let $N_{ik}^j = (\varepsilon_{ik}^j, \beta_{ik}^j, \gamma_{ik}^j, \delta_{ik}^j, \eta_{ik}^j)$ be the vector related to the sentiment values of online reviews $R_{ik}$, where $\varepsilon_{ik}^j, \beta_{ik}^j, \gamma_{ik}^j, \delta_{ik}^j, \eta_{ik}^j = 0$ or 1, $i = 1, 2, \ldots, m$, $j = 1, 2, \ldots, n$, $k = 1, 2, \ldots, K_i$. There are six possible cases for the value of $N_{ik}^j$: $N_{ik}^j = (1, 0, 0, 0, 0), N_{ik}^j = (0, 1, 0, 0, 0), N_{ik}^j = (0, 0, 1, 0, 0), N_{ik}^j = (0, 0, 0, 1, 0), N_{ik}^j = (0, 0, 0, 0, 1)$, and $N_{ik}^j = (0, 0, 0, 0, 0)$, with each value indicating that the emotional value of $R_{ik}$ is $-2, -1, 0, 1, 2$, or missing; $i = 1, 2, \ldots, m$, $j = 1, 2, \ldots, n$, and $k = 1, 2, \ldots, K_i$.

Let $q_{ij}^r, q_{ij}^r, q_{ij}^r$, and $q_{ij}^r$ represent that the number of the emotion values of online review is $-2, -1, 0, 1, 2$, and $2$ in $R_{ik} = \{R_{ik1}, R_{ik2}, \ldots, R_{ikK_i}\}$, respectively; $i = 1, 2, \ldots, m$ and $k = 1, 2, \ldots, K_i$. Based on $N_{ik}^j = (\varepsilon_{ik}^j, \beta_{ik}^j, \gamma_{ik}^j, \delta_{ik}^j, \eta_{ik}^j)$, the values of $q_{ij}^r, q_{ij}^r, q_{ij}^r$, and $q_{ij}^r$ can be calculated by the five following equations, respectively:

$$ q_{ij}^r = \sum_{k=1}^{K_i} u_{ik} e_{ik}^j, \quad i = 1, 2, \ldots, m, j = 1, 2, \ldots, n, \quad (21) $$
3.2.5. Selecting Product Attribute Values by Roulette Wheel Selection Algorithm. In real life, some electronic products continue to be updated and upgraded by merchants after their launch. In addition, different consumers have different evaluation criteria for the products. Moreover, since the evaluation value cannot reflect these differences, we need to choose the attribute value randomly. On the basis of the LD of different alternative products on different attributes, we can get the distribution function $F$ of attribute value of the alternative product on different attributes.

The distribution function of attribute value of alternative $X_i$ on attribute $A_j$ can be expressed as

$$
F_{ij} = \begin{cases}
0, & \alpha < -2, \\
p^{-2}_{ij}, & -2 \leq \alpha < -1, \\
p^{-2}_{ij} + p^{-1}_{ij}, & -1 \leq \alpha < 0, \\
p^{-2}_{ij} + p^{-1}_{ij} + p^{0}_{ij}, & 0 \leq \alpha < 1, \\
1 - p^{2}_{ij}, & 1 \leq \alpha < 2, \\
1, & 2 \leq \alpha,
\end{cases}
$$

$i = 1, 2, \ldots, m, j = 1, 2, \ldots, n.$

Suppose that $P = (P^{2}_{ij}, P^{-1}_{ij}, P^{0}_{ij}, P^{1}_{ij}, P_{ij})$ denotes the probability distribution with different emotion values, where $P^{2}_{ij}, P^{-1}_{ij}, P^{0}_{ij}, P^{1}_{ij},$ and $P_{ij}$, respectively, denote the probabilities with the emotion values of online review in $R_{ik} = \{R_{ik}^1, R_{ik}^2, \ldots, R_{ik}^n\}, \ i = 1, 2, \ldots, m, \ k = 1, 2, \ldots, K_i$. Based on the obtained values of $q^{\beta}_{ij}, q^{\delta}_{ij}, q^{\gamma}_{ij}, q^{\phi}_{ij},$ and $q^{\psi}_{ij}$, the elements in the probability distribution $P = (P^{2}_{ij}, P^{-1}_{ij}, P^{0}_{ij}, P^{1}_{ij}, P_{ij})$ can be calculated by the following equations, respectively:

$$
P^{2}_{ij} = \frac{q^{\beta}_{ij}}{q^{\beta}_{ij} + q^{\delta}_{ij} + q^{\gamma}_{ij} + q^{\phi}_{ij} + q^{\psi}_{ij}}, \ i = 1, 2, \ldots, m, j = 1, 2, \ldots, n,
$$

$$
P^{-1}_{ij} = \frac{q^{\gamma}_{ij}}{q^{\beta}_{ij} + q^{\delta}_{ij} + q^{\gamma}_{ij} + q^{\phi}_{ij} + q^{\psi}_{ij}}, \ i = 1, 2, \ldots, m, j = 1, 2, \ldots, n,
$$

$$
P^{0}_{ij} = \frac{q^{\phi}_{ij}}{q^{\beta}_{ij} + q^{\delta}_{ij} + q^{\gamma}_{ij} + q^{\phi}_{ij} + q^{\psi}_{ij}}, \ i = 1, 2, \ldots, m, j = 1, 2, \ldots, n,
$$

$$
P^{1}_{ij} = \frac{q^{\psi}_{ij}}{q^{\beta}_{ij} + q^{\delta}_{ij} + q^{\gamma}_{ij} + q^{\phi}_{ij} + q^{\psi}_{ij}}, \ i = 1, 2, \ldots, m, j = 1, 2, \ldots, n.
$$

Therefore, the evaluation matrix $B = (b_{ij})_{m \times n}$ can be obtained, where element $b_{ij}$ indicates that the LD of alternative $X_i$ relates to attribute $A_j$. It can be calculated by the following equation:

$$
b_{ij} = \left\{ \left(S_{\alpha}, P_{ij}^{\alpha} \right) \mid S_{\alpha} \in S, P_{ij}^{\alpha} \geq 0, \alpha = -\tau, 0, \ldots, \tau, \sum_{\alpha=-\tau}^{\tau} P_{ij}^{\alpha} = 1 \right\}, \ i = 1, 2, \ldots, m, j = 1, 2, \ldots, n.
$$

Table 1: Some adverbs of degree and their scores.

| Score | Degree adverb                                      |
|-------|---------------------------------------------------|
| 1     | A little, slightly, quite, more, briefly, pretty,  |
|       | rather, fairly, considerably, etc.                |
| 2     | Very, super, most, especially, absolutely, special, |
|       | extra, incomparable, etc.                         |
Therefore, according to the distribution function $F$ of attribute values, random selection of product attributes will be more realistic. We can get the decision matrix $D = (d_{ij})_{m \times n}$ where element $d_{ij}$ is obeying distribution function $F$, $i = 1, 2, \ldots, m$, $j = 1, 2, \ldots, n$.

### 3.3. Product Ranking

#### 3.3.1. Calculating Attribute Weights

In the product ranking process, the determination of product attribute weight plays a vital role. Diakoulaki [60] put forward the Criteria Importance through Intercriteria Correlation (CRITIC) method, which is an objective weighting method. CRITIC is that the objective weight of the attribute is determined according to two basic concepts. One is contrast intensity, which indicates the magnitude of attribute value difference of each alternative on the same attribute in the form of standard deviation. That is, the magnitude of the standard indicates the magnitude of the difference between the attribute values of each alternative on the same attribute. When the standard deviation is larger, the difference between the attribute values of each alternative is greater. The second is the conflict between attributes, which is measured by correlation between attributes. That is to say, if the positive correlation between two attributes is stronger, the conflict between these two attributes will be weaker.

The expectation matrix $E$ obtained from that evaluation matrix $B$ is

$$E = \begin{bmatrix}
e_{11} & e_{12} & \cdots & e_{1n} \\
e_{21} & e_{22} & \cdots & e_{2n} \\
\vdots & \vdots & \ddots & \vdots \\
e_{m1} & e_{m2} & \cdots & e_{mn}
\end{bmatrix}.$$  \hspace{1cm} (33)

Element $e_{ij} = E(b_{ij})$ is the expected value of the LD of alternative product $X_i$ on attribute $A_j$, $i = 1, 2, \ldots, m$, $j = 1, 2, \ldots, n$.

Step 1: matrix $E$ standardization.

$$N = \begin{bmatrix}
n_{11} & n_{12} & \cdots & n_{1n} \\
n_{21} & n_{22} & \cdots & n_{2n} \\
\vdots & \vdots & \ddots & \vdots \\
n_{m1} & n_{m2} & \cdots & n_{mn}
\end{bmatrix}.$$  \hspace{1cm} (34)

The element standardization method of expectation matrix $E$ is the same as the classical MABAC method.

Step 2: calculating the correlation coefficient matrix $R = (r_{jk})_{m \times n}$ based on the normalized matrix $N$. Here, $r_{jk}$ is correlation coefficient between attributes $A_j$ and $A_k$ and it is calculated in the following equation:

$$r_{jk} = \frac{\sum_{i=1}^{m} (x_{ij} - \bar{x}_j) (x_{ik} - \bar{x}_k)}{\sqrt{\sum_{i=1}^{m} (x_{ij} - \bar{x}_j)^2 \sum_{i=1}^{m} (x_{ik} - \bar{x}_k)^2}}$$  \hspace{1cm} (35)

where $\bar{x}_j$ and $\bar{x}_k$ denote the means of the $j$th and $k$th attributes.

Step 3: on the basis of the correlation coefficient, the value of index $C$ can be calculated by

$$C_j = \sigma_j \sum_{i=1}^{n} (1 - r_{ij}), \quad j = 1, 2, \ldots, n.$$  \hspace{1cm} (36)

where $\sigma_j$ is the standard deviation of attribute $A_j$. The standard deviation is used to quantify the contrast intensity of the corresponding attribute and it can be computed by the following equation:

$$\sigma_j = \sqrt{\frac{1}{n-1} \sum_{j=1}^{n} (n_{ij} - \bar{n}_j)^2}, \quad j = 1, 2, \ldots, n.$$  \hspace{1cm} (37)

Step 4: computing the weight of each attribute. According to the value of index $C$, the weight of each attribute can be obtained by the following formula:

$$w_j = \frac{C_j}{\sum_{j=1}^{n} C_j}, \quad j = 1, 2, \ldots, n$$  \hspace{1cm} (38)

#### 3.3.2. The Extended Multiattributive Border Approximation Area Comparison Method

Compared with the classical MABAC method, the extended MABAC method with mixed information has differences in determining the border approximation area and other steps. The specific steps of the extended MABAC method are as follows:

Step 1: forming an initial decision matrix $D = (d_{ij})_{m \times n}$.

The initial decision matrix element $d_{ij}$ is the attribute value of the alternative product $X_i$ on attribute $A_j$, and the attribute value $d_{ij}$ can be in the form of linguistic term or real number ($i = 1, 2, \ldots, m$, $j = 1, 2, \ldots, n$).

Step 2: normalizing the initial decision matrix. Only normalize real number elements; the normalization calculation method is the same as the classical MABAC method.

Step 3: determining the border approximation area matrix $G$. The formula for computing the elements in the border approximation area matrix $G$ is

$$g_{ij} = \frac{1}{m} \sum_{i=1}^{m} n_{ij}, \quad i = 1, 2, \ldots, m; \quad j = 1, 2, \ldots, n.$$  \hspace{1cm} (39)
Step 4: computing the score of each alternative on each attribute. The score of the alternative product $X_i$ on attribute $A_j$ is determined by the following equation:

$$q_{ij} = \begin{cases} 
  d(n_{ij}, g_j), & \text{if } n_{ij} > g_j, \\
  0, & \text{if } n_{ij} = g_j, \\
 -d(n_{ij}, g_j), & \text{if } n_{ij} < g_j,
\end{cases} \quad i = 1, 2, \ldots, m, j = 1, 2, \ldots, n. \quad (40)$$

where $d(n_{ij}, g_j)$ denotes the distance between the value of $n_{ij}$ and the value of $g_j$.

If $q_{ij} = 0$, the alternative products $X_i$ may belong to the border approximation area ($G$). If $q_{ij} < 0$, the alternative products $X_i$ could belong to the lower approximation area ($G^-$). If $q_{ij} > 0$, the alternative products $X_i$ might belong to the upper approximation area ($G^+$). Thus, same as the classical MABAC method, in order for alternative product $X_i$ to be selected as the best one in alternative products set, it is necessary for alternative product $X_i$ to belong to the upper approximation area ($G^+$) under as many attributes as possible.

Step 5: alternative product ranking.

$$S_i = \sum_{j=1}^{n} w_j q_{ij}, \quad i = 1, 2, \ldots, m. \quad (41)$$

Based on the comprehensive score $S_i$ of the alternative product $X_i$, we can get the ranking result of alternative products. The bigger the comprehensive score $S_i$ is, the better the alternative product $X_i$ is.

4. Case Study

With the continuous updating and iteration of communication technology and information technology, we are gradually entering the 5G communication era. The advent of 5G is not only the improvement of network speed and capacity but also a disruptive industrial change and another explosion of consumer digital life. With the advent of the 5G era, 5G products have mushroomed with rapid development. People have given 5G unprecedented expectations because this is a leap forward in the new era. However, while 5G has brought about changes in consumer-oriented digital life, new issues have emerged regarding consumers’ choice of 5G products. Due to the fact that the 5G era is not yet fully popularized, the value of 5G products is relatively high for ordinary consumers. Therefore, selecting a 5G product which meets consumers’ expectations is worth studying. Suppose that a consumer plans to buy a 5G mobile phone in acceptable price range of CNY4000-5000. By searching for products that meet the requirements of consumers in JD.COM, five alternative products with 8 GB of operating memory and 256 G of fuselage storage capacity were selected as shown in Figure 4.

First, online reviews of five alternatives are collected by crawler software. Based on the preprocessing of online reviews data, the TF-IDF algorithm is used to extract product attributes. The attributes are $A_1$: Exterior, $A_2$: Screen, $A_3$: Performance, $A_4$: Loudspeaker, and $A_5$: Camera. Through the sentiment analysis results of online reviews, we can obtain the LD of different alternative products on different attributes. Since every consumer wants to obtain a cost-effective product, the product’s price should be added to the attributes. Thus, attribute $A_6$ is Price. Then, the values of the attributes of each alternative are shown in Table 2.

According to the LD of each product on each attribute, the expectation matrix $E$ can be obtained as follows:

$$E = \begin{bmatrix} 
1.31 & 0.83 & 1.30 & 0.88 & 1.00 & 4998 \\
1.26 & 1.02 & 1.11 & 0.73 & 1.02 & 4999 \\
1.24 & 1.03 & 1.42 & 0.93 & 1.05 & 4299 \\
1.28 & 0.91 & 1.12 & 1.00 & 1.13 & 4988 \\
1.03 & 0.70 & 1.21 & 0.69 & 1.07 & 4299 
\end{bmatrix}. \quad (42)$$

On the basis of the expectation matrix $E$, we can calculate the attribute weight vector by the CRITIC method as $w = (0.1121, 0.1533, 0.2177, 0.1211, 0.2003, 0.1955)^T$.

4.1. The Classical MABAC Method. With the expectation matrix $E$ as the initial decision matrix, the alternative ranking results can be obtained by the classical MABAC method. From equations (5)–(11), we can have the following comprehensive score values for each alternative: $S_1 = 0.0167$, $S_2 = 0.1124$, $S_3 = 0.3088$, $S_4 = 0.1264$, and $S_5 = 0.0286$. Based on the comprehensive score values, the ranking results of the alternative can be obtained as $X_3>X_4>X_1>X_5>X_2$. It can be seen that MI11 is the optimal alternative.

4.2. The Extended MABAC Method with Mixed Information. The extended MABAC method was used to randomly simulate the ranking of alternatives 500 times, and the simulation results are listed in Table 3.

According to the average value of simulation results, the ranking results of the alternative products are $X_3>X_4>X_1>X_5>X_2$. It is not difficult to see that the ranking results are roughly the same as the ranking result using the classical MABAC method based on the expectation matrix $E$. It can be seen that MI11 is still the most preferable.
acceptable set of alternative products.

Table 2: Attribute values for each attribute for each alternative product.

|   | A1      | A2      | A3      | A4      | A5      | A6  |
|---|---------|---------|---------|---------|---------|-----|
| X1 | (S1, 0.11) | (S1, 0.01) | (S0, 0.04) | (S1, 0.02) | (S0, 0.17) | 4998 |
|   | (S1, 0.47) | (S0, 0.31) | (S1, 0.61) | (S0, 0.27) | (S0, 0.65) |
|   | (S2, 0.42) | (S2, 0.52) | (S2, 0.34) | (S1, 0.52) | (S2, 0.18) |
| X2 | (S1, 0.02) | (S0, 0.16) | (S0, 0.11) | (S0, 0.33) | (S0, 0.11) | 4999 |
|   | (S0, 0.03) | (S1, 0.65) | (S1, 0.67) | (S0, 0.61) | (S0, 0.75) |
|   | (S2, 0.32) | (S2, 0.19) | (S2, 0.22) | (S2, 0.06) | (S2, 0.13) |
| X3 | (S1, 0.76) | (S0, 0.08) | (S0, 0.03) | (S0, 0.20) | (S0, 0.12) | 4299 |
|   | (S2, 0.24) | (S1, 0.82) | (S1, 0.52) | (S1, 0.66) | (S1, 0.71) |
|   |             | (S2, 0.10) | (S1, 0.45) | (S2, 0.14) | (S2, 0.17) |
| X4 | (S1, 0.01) | (S0, 0.14) | (S0, 0.06) | (S0, 0.12) | (S0, 0.07) | 4988 |
|   | (S0, 0.01) | (S1, 0.82) | (S1, 0.76) | (S1, 0.77) | (S0, 0.72) |
|   | (S1, 0.67) | (S2, 0.04) | (S2, 0.18) | (S2, 0.11) | (S2, 0.21) |
| X5 | (S1, 0.04) | (S1, 0.09) | (S1, 0.06) | (S1, 0.04) | (S0, 0.12) | 4299 |
|   | (S0, 0.05) | (S0, 0.14) | (S0, 0.06) | (S0, 0.36) | (S0, 0.70) |
|   | (S1, 0.74) | (S1, 0.51) | (S1, 0.46) | (S0, 14)   | (S2, 0.18) |
|   | (S2, 0.17) | (S2, 0.38) | (S2, 0.14) |            |            |

Table 3: Random simulated ranking values for each alternative product.

|   | X1      | X2      | X3      | X4      | X5      |
|---|---------|---------|---------|---------|---------|
| 1 | -0.0827 | -0.0688 | 0.1267  | -0.1041 | 0.0563  |
| 2 | 0.0730  | -0.1255 | -0.0980 | -0.0702 | 0.0980  |
| 3 | -0.0762 | -0.1709 | 0.1470  | -0.0152 | 0.1273  |
| 4 | 0.1195  | -0.0308 | 0.1647  | -0.1064 | -0.1042 |
| 5 | -0.1114 | 0.0771  | -0.0839 | -0.0642 | 0.1567  |
| 6 | -0.0517 | -0.0800 | 0.1699  | -0.0770 | -0.0096 |
| 7 | -0.0594 | -0.1260 | 0.1138  | -0.0725 | 0.0378  |
| 496| -0.1346 | -0.0141 | 0.0665  | -0.1038 | 0.1254  |
| 497| -0.0931 | -0.1337 | 0.1546  | -0.0923 | -0.0128 |
| 498| -0.1479 | -0.0818 | -0.1137 | 0.0243  | 0.1402  |
| 499| 0.0396  | -0.1688 | 0.1437  | -0.0669 | -0.1800 |
| 500| -0.1015 | -0.1079 | 0.2087  | -0.0707 | 0.0715  |
| Average| -0.0722 | -0.0791 | 0.0448  | 0.0108  | -0.0135 |

alternative product. There is a slight difference in the ranking of alternative products X1 and X5. Because the expectation value is susceptible to extreme values in the data, when a set of data is obviously skewed, the expectation value is less representative. Since the LDs of alternatives X1 and X5 are significantly skewed distribution on multiple attributes, the ranking result obtained with the expectation value is inaccurate. The differences between the two methods will be explained in detail in the simulation experiments.

For the sake of further proving the rationality of the proposed method, we compare the ranking results obtained by this method with those obtained by other methods. The method proposed by Jiang et al. [61] can be considered as the most relevant method in this study. In their method, a measurement method for measuring the distance between two discrete stochastic variables is proposed. Then, according to the basic principle of TOPSIS method, the relative closeness between each alternative product and the ideal and nadir solutions is calculated. Finally, the alternative ranking results are obtained based on the relative closeness. On the basis of the data of this case, the ranking result by the method proposed by Jiang et al. [61] was obtained as X3 > X1 > X2 = X3 > X4. It can be seen that the ranking results obtained by the method proposed in this paper are the same as those obtained by the method proposed by Jiang et al. Thus, the rationality of the proposed method in this paper is demonstrated.
4.3. Simulation Experiment. The specificity of the above case data is such that it is difficult to demonstrate the differences between the method proposed in this paper and the classical MABAC method based on expectation value. Therefore, this section will illustrate the differences between the two methods through a simulation experiment. We assume that there are five alternatives: $x_1^*$, $x_2^*$, $x_3^*$, $x_4^*$, and $x_5^*$. Then randomly generate online review sentiment analysis results for each alternative on the five attributes. The sentiment analysis results are still represented by LD. To reduce the influence of the weights on the ranking results, we assume that the weight of each attribute is $\omega = (0.15, 0.2, 0.25, 0.3, 0.1)^T$ and it remains constant. We use MATLAB to implement the simulation. The ranking values for each alternative obtained during the simulation are shown in Figures 5–9. The ranking values for each alternative during the simulation were counted and the results are shown in Table 4. As shown in Figures 5–9 and Table 4, the ranking values for alternative $x_1^*$ are more concentrated in ranking 2. The remaining alternatives have more dispersed ranking values. It may be that the LD values of the remaining alternatives are discrete on each attribute. In the actual product selection process, this occurs when consumers give these alternatives mixed reviews. The final ranking of these alternatives will depend on how they are perceived by consumers as a whole.

The ranking results of the alternative by simulation using the method proposed in this paper are $x_1^* \succ x_2^* \succ x_3^* \succ x_4^* \succ x_5^*$. The ranking results of the traditional MABAC method based on expectation are $x_1^* \succ x_3^* \succ x_4^* \succ x_5^* \succ x_2^*$. It can be seen that although the optimal alternative is the same, there are differences in the ranking results of the remaining alternatives. This is because although the two methods use the same set of data, the two methods lead to different levels of information loss. Representing sentiment analysis results through the expectation value of LD exacerbates the loss of information. Suppose that the sentiment analysis results of two alternatives on an attribute are represented by the LDs as $L_1 = \{(S_1, 0.25), (S_0, 0.32), (S_1, 0.2), (S_2, 0.23)\}$ and $L_2 = \{(S_0, 0.69), (S_1, 0.21), (S_1, 0.1)\}$. It can be calculated that the expectation values of these two LDs are the same; $E(L_1) = E(L_2) = 0.41$. However, observing the two LDs shows that $L_2$ is better than $L_1$. Thus, representing sentiment analysis results through the expectation value of LD would ignore the differences between alternatives. However, it is possible that these differences are the key to the ranking of the alternatives.

Although the loss of information is inevitable when ranking products, we should try to reduce it as much as possible. In this paper, the method of randomly selecting product attribute values using the roulette wheel selection algorithm reduces the loss of information as much as possible. This method not only preserves the differences between product attributes but also gets more convincing ranking results.
5. Conclusion

In this study, we propose a new method to rank alternative products on the basis of online reviews. The specific contributions of this paper are summarized as follows.

First of all, this study overcomes the problem that the reference value of online reviews is not fully utilized in the existing research. In this paper, the reference value of online reviews is maximized by calculating the time utility and support utility of online reviews. Online reviews with different publication times and support numbers have different reference values.

Secondly, we propose a method to represent the sentiment analysis results of online reviews by LD. Compared with other transformation methods, the method proposed in this paper is simpler. The linguistic term set is constructed according to the emotional tendency of the online reviews. Next, the percentage of each emotional tendency of each attribute in the online reviews is calculated. Then, the LD is constructed based on the percentage of emotional tendency.

In addition, this paper proposes a method for selecting product attributes by means of the roulette wheel selection algorithm. Representing sentiment analysis results through the expectation value of linguistic distribution will exacerbate the loss of information. This paper provides a method to solve this problem. The attributes of products are randomly selected by roulette wheel selection algorithm. Then, the ranking results are achieved by extended MABAC method with mixed information. Consequently, not only does the proposed method in this paper reduce the information loss but also the ranking results are more relevant to reality.

Finally, this paper provides a method to solve the problem of unknown product attributes and their weights. This method can extract the attributes of alternative products from online reviews. Then, the attribute objective weights are determined through the sentiment analysis results of online reviews.

There are still a number of limitations that need to be further addressed:

1. This paper did not consider the utility of follow-up reviews of online reviews. The follow-up reviews of online reviews include the consumer experience of using products for a period of time, which has high reference value. Therefore, how to make full use of online reviews needs further study.

2. In this paper, only the objective weight of each attribute is considered in the ranking process of alternatives. However, the subjective weight provided by consumers is also crucial in the ranking process. Consequently, further research is needed to consider both subjective and objective weights.

In terms of future research, the proposed method for representing the sentiment analysis results of online reviews in LDs can be used in other management decision analysis issues. Furthermore, it would be interesting to extend the method proposed in this paper to Pythagorean fuzzy uncertain environment [62]. In addition, as merchants become aware of the impact online reviews have on consumers, they are beginning to resort to online review manipulation [63, 64]. Consequently, how to identify genuine and valid online reviews in the case of online review manipulation is something that needs further research. Additionally, how to make product selection based on online reviews in the context of merchant online review manipulation is also an issue that deserves further research.

Data Availability

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

Table 4: Frequency statistics of the simulation ranking results for each alternative.

| Value | Alternative 1, $x_1^*$ | Alternative 2, $x_2^*$ | Alternative 3, $x_3^*$ | Alternative 4, $x_4^*$ | Alternative 5, $x_5^*$ |
|-------|------------------------|------------------------|------------------------|------------------------|------------------------|
|       | Count | Percent | Count | Percent | Count | Percent | Count | Percent | Count | Percent | Count | Percent |
| 1     | 85    | 17.00   | 115   | 23.00   | 117   | 23.40   | 88    | 17.60   | 95    | 19.00   |
| 2     | 253   | 50.60   | 119   | 23.80   | 62    | 12.40   | 51    | 10.20   | 15    | 3.00    |
| 3     | 48    | 9.60    | 83    | 16.60   | 98    | 19.60   | 114   | 22.80   | 157   | 31.40   |
| 4     | 37    | 7.40    | 86    | 17.20   | 107   | 21.40   | 122   | 24.40   | 148   | 29.60   |
| 5     | 77    | 15.40   | 97    | 19.40   | 116   | 23.20   | 125   | 25.00   | 85    | 17.00   |

Figure 9: Simulated ranking values for alternative $x_5^*$. 
Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this study.

Acknowledgments

This work was jointly supported by the National Natural Science Foundation of China (NSFC, Project nos. 71801142 and 71771140), the Shandong Province Social Science Planning Project (no. 21CGLJ12), and Natural Science Planning Project of Shandong Province (no. ZR2019BG006).

References

[1] Z. Zhang, Q. Ye, R. Law, and Y. Li, “The impact of e-word-of-mouth on the online popularity of restaurants: a comparison of consumer reviews and editor reviews,” International Journal of Hospitality Management, vol. 29, no. 4, pp. 694–700, 2010.

[2] B. Fang, Q. Ye, D. Kucukusta, and R. Law, “Analysis of the perceived value of online tourism reviews: influence of readability and reviewer characteristics,” Tourism Management, vol. 52, pp. 498–506, 2016.

[3] Z. Zhang, Z. Zhang, and Y. Yang, “The power of expert identity: how website-recognized expert reviews influence travelers’ online rating behavior,” Tourism Management, vol. 55, pp. 15–24, 2016.

[4] R. Gang and T. Hong, “Examining the relationship between specific negative emotions and the perceived helpfulness of online reviews,” Information Processing and Management, vol. 56, no. 4, pp. 1425–1438, 2019.

[5] M. Malik and A. Hussain, “An analysis of review content and reviewer variables that contribute to review helpfulness,” Information Processing and Management, vol. 54, no. 1, pp. 88–104, 2018.

[6] M. Farhadloo, R. A. Patterson, and E. Rolland, “Modeling customer satisfaction from unstructured data using a Bayesian approach,” Decision Support Systems, vol. 90, pp. 1–11, 2016.

[7] D. E. Pournarakis, D. N. Sotiropoulos, and G. M. Giaglis, “A computational model for mining consumer perceptions in social media,” Decision Support Systems, vol. 93, pp. 98–110, 2017.

[8] J. Qi, Z. Zhang, S. Jeon, and Y. Zhou, “Mining customer requirements from online reviews: a product improvement perspective,” Information and Management, vol. 53, no. 8, pp. 951–963, 2016.

[9] S. Xiao, C.-P. Wei, and M. Dong, “Crowd intelligence: analyzing online product reviews for preference measurement,” Information and Management, vol. 53, no. 2, pp. 169–182, 2016.

[10] J.-W. Bi, Y. Liu, and Z.-P. Fan, “Representing sentiment analysis results of online reviews using interval type-2 fuzzy numbers and its application to product ranking,” Information Sciences, vol. 504, pp. 293–307, 2019.

[11] Y. Liu, J. W. Bi, and Z. P. Fan, “Ranking products through online reviews: a method based on sentiment analysis technique and intuitionistic fuzzy set theory,” Information Fusion, vol. 36, pp. 149–161, 2016.

[12] W. Wang, H. Wang, and Y. Song, “Ranking product aspects through sentiment analysis of online reviews,” Journal of Experimental and Theoretical Artificial Intelligence, vol. 29, no. 2, pp. 227–246, 2017.

[13] C. Guo, Z. Du, and X. Kou, “Products ranking through aspect-based sentiment analysis of online heterogeneous reviews,” Journal of Systems Science and Systems Engineering, vol. 27, no. 5, pp. 542–558, 2018.

[14] D. Zhang, Y. Li, and C. Wu, “An extended TODIM method to rank products with online reviews under intuitionistic fuzzy environment,” Journal of the Operational Research Society, vol. 71, no. 2, pp. 322–334, 2020.

[15] L. Ying, J. Jian, J. Ping, J. A. Harding, and R. Fung, “Identifying helpful online reviews: a product designer’s perspective,” Computer-Aided Design, vol. 45, no. 2, pp. 180–194, 2013.

[16] M. Zhang, B. Fan, N. Zhang, W. Wang, and W. Fan, “Mining product innovation ideas from online reviews,” Information Processing and Management, vol. 58, no. 1, Article ID 102389, 2021.

[17] S. Tirunillai and G. J. Tellis, “Mining marketing meaning from online chatter: strategic brand analysis of big data using latent Dirichlet allocation,” Journal of Marketing Research, vol. 51, no. 4, pp. 463–479, 2014.

[18] A. Culotta and J. Cutler, “Mining brand perceptions from twitter social networks,” Marketing Science, vol. 35, no. 3, pp. 343–362, 2016.

[19] O. Netzer, R. Feldman, J. Goldenberg, and M. Fresko, “Mine your own business: market-structure surveillance through text mining,” Marketing Science, vol. 31, no. 3, pp. 521–543, 2012.

[20] J.-W. Bi, Y. Liu, Z.-P. Fan, and E. Cambria, “Modelling customer satisfaction from online reviews using ensemble neural network and effect-based Kano model,” International Journal of Production Research, vol. 57, no. 22, pp. 7068–7088, 2019.

[21] X. Xu, X. Wang, Y. Li, and M. Haghhighi, “Business intelligence in online customer textual reviews: understanding consumer perceptions and influential factors,” International Journal of Information Management, vol. 37, no. 6, pp. 673–683, 2017.

[22] Y. Meng, N. Yang, Z. Qian, and G. Zhang, “What makes an online review more helpful: an interpretation framework using XGBoost and SHAP values,” Journal of Theoretical and Applied Electronic Commerce Research, vol. 16, no. 3, pp. 466–490, 2021.

[23] Y. Guo, S. J. Barnes, and Q. Jia, “Mining meaning from online ratings and reviews: tourist satisfaction analysis using latent Dirichlet allocation,” Tourism Management, vol. 59, pp. 467–483, 2017.

[24] H. Kang, S. J. Yoo, and D. Han, “Senti-lexicon and improved Naïve Bayes algorithms for sentiment analysis of restaurant reviews,” Expert Systems with Applications, vol. 39, no. 5, pp. 6000–6010, 2012.

[25] A. Onan, S. Koruko˘glu, and H. Bulut, “A multiobjective weighted voting ensemble classifier based on differential evolution algorithm,” Expert Systems with Applications, vol. 42, no. 4, pp. 1857–1863, 2015.

[26] Y. Ganin, E. Ustinova, H. Ajakan et al., “Domain-adversarial training of neural networks,” The Journal of Machine Learning Research, vol. 17, no. 1, pp. 2096–2108, 2016.

[27] D. Zhang, H. Xu, Z. Su, and Y. Xu, “Chinese reviews sentiment classification based on word2vec and SVMPerf,” Expert Systems with Applications, vol. 42, no. 4, pp. 1857–1863, 2015.

[28] Y. Liu, J.-W. Bi, and Z.-P. Fan, “A method for multi-class sentiment classification based on an improved one-vs-one (OVO) strategy and the support vector machine (SVM) algorithm,” Information Sciences, vol. 394, pp. 38–52, 2017.
[29] A. Moreo, M. Romero, J. L. Castro, and J. M. Zurita, “Lexicon-based comments-oriented news sentiment analyzer system,” *Expert Systems with Applications*, vol. 39, no. 10, pp. 9166–9180, 2012.

[30] W. Zhang, H. Xu, and W. Wan, “Weakness Finder: find product weakness from Chinese reviews by using aspects based sentiment analysis,” *Expert Systems with Applications*, vol. 39, no. 11, pp. 10283–10291, 2012.

[31] J. Wu, K. Lu, S. Su, and S. Wang, “Chinese micro-blog sentiment analysis based on multiple sentiment dictionaries and semantic rule sets,” *IEEE Access*, vol. 7, pp. 183924–183939, 2019.

[32] Z. Cui, Q. Qiu, C. Yin, J. Yu, Z. Wu, and A. Deng, “A barrage sentiment analysis scheme based on expression and tone,” *IEEE Access*, vol. 7, pp. 180324–180335, 2019.

[33] M. Ahmed, Q. Chen, and Z. Li, “Constructing domain-dependent sentiment dictionary for sentiment analysis,” *Neural Computing and Applications*, vol. 32, no. 18, pp. 14719–14732, 2020.

[34] Y. Liu, J.-W. Bi, and Z.-P. Fan, “Multi-class sentiment classification: the experimental comparisons of feature selection and machine learning algorithms,” *Expert Systems with Applications*, vol. 80, pp. 323–339, 2017.

[35] Y. Liu, J. W. Bi, and Z. P. Fan, “A method for ranking products through online reviews based on sentiment classification and interval-valued intuitionistic fuzzy TOPSIS,” *International Journal of Information Technology and Decision Making*, vol. 16, no. 6, pp. 1497–1522, 2017.

[36] R.-x. Nie, Z.-p. Tian, J.-q. Wang, and K. S. Chin, “Hotel selection driven by online textual reviews: applying a semantic partitioned sentiment dictionary and evidence theory,” *International Journal of Hospitality Management*, vol. 88, Article ID 102495, 2020.

[37] D. Zhang, C. Wu, and J. Liu, “Ranking products with online reviews: a novel method based on hesitant fuzzy set and sentiment word framework,” *Journal of the Operational Research Society*, vol. 71, no. 3, pp. 528–542, 2020.

[38] X. Liang, P. Liu, and Z. Wang, “Hotel selection utilizing online reviews: a novel decision support model based on sentiment analysis and DL-VIKOR method,” *Technological and Economic Development of Economy*, vol. 25, no. 6, pp. 1139–1161, 2019.

[39] P. Liu and F. Teng, “Probabilistic linguistic TODIM method for selecting products through online product reviews,” *Information Sciences*, vol. 485, pp. 441–455, 2019.

[40] E. Najmi, K. Hashmi, Z. Malik, A. Rezgui, and H. U. Khan, “CAPRA: a comprehensive approach to product ranking using customer reviews,” *Computing*, vol. 97, no. 8, pp. 843–867, 2015.

[41] X. Fu, T. Ouyang, Z. Yang, and S. Liu, “A product ranking method combining the features-opinion pairs mining and interval-valued Pythagorean fuzzy sets,” *Applied Soft Computing*, vol. 97, Article ID 106803, 2020.

[42] M. Y. Li, X. J. Zhao, L. Zhang, X. Ye, and B. Li, “Method for product selection considering consumer’s expectations and online reviews,” *Cybernetes*, vol. 50, no. 9, 2020.

[43] R. M. Rodriguez, L. Martinez, and F. Herrera, “Hesitant fuzzy linguistic term sets for decision making,” *IEEE Transactions on Fuzzy Systems*, vol. 20, no. 1, pp. 109–119, 2011.

[44] G. Zhang, Y. Dong, and Y. Xu, “Consistency and consensus measures for linguistic preference relations based on distribution assessments,” *Information Fusion*, vol. 17, pp. 46–55, 2014.

[45] Y. Wu, Z. Zhang, G. Kou et al., “Distributed linguistic representations in decision making: taxonomy, key elements and applications, and challenges in data science and explainable artificial intelligence,” *Information Fusion*, vol. 65, pp. 165–178, 2020.

[46] Z. Xu, “A method based on linguistic aggregation operators for group decision making with linguistic preference relations,” *Information Sciences*, vol. 166, no. 1–4, pp. 19–30, 2004.

[47] Y. Dong, Y. Xu, H. Li, and B. Feng, “The OWA-based consensus operator under linguistic representation models using position indexes,” *European Journal of Operational Research*, vol. 203, no. 2, pp. 455–463, 2010.

[48] X. Liang, F. Teng, and Y. Sun, “Multiple group decision making for selecting emergency alternatives: a novel method based on the LDWPA operator and LD-MABAC,” *International Journal of Environmental Research and Public Health*, vol. 17, no. 8, p. 2945, 2020.

[49] W. Qian, J. Chai, Z. Xu, and Z. Zhang, “Differential evolution algorithm with multiple mutation strategies based on roulette wheel selection,” *Applied Intelligence*, vol. 48, no. 10, pp. 3612–3629, 2018.

[50] R. Bakhat and M. Rajaa, “Risk assessment of a wind turbine using an AHP-MABAC approach with grey system theory: a case study of Morocco,” *Mathematical Problems in Engineering*, vol. 2020, Article ID 2496914, 2020.

[51] D. Pamucar and G. Ćirović, “The selection of transport and handling resources in logistics centers using Multi-Attributive Border Approximation area Comparison (MABAC),” *Expert Systems with Applications*, vol. 42, no. 6, pp. 3016–3028, 2015.

[52] X. Peng and J. Dai, “Approaches to single-valued neutrosophic MADM based on MABAC, TOPSIS and new similarity measure with score function,” *Neural Computing and Applications*, vol. 29, no. 10, pp. 939–954, 2018.

[53] D. Pamucar, I. Petrović, and G. Ćirović, “Modification of the Best–Worst and MABAC methods: a novel approach based on interval-valued fuzzy-rough numbers,” *Expert Systems with Applications*, vol. 91, pp. 89–106, 2018.

[54] Y. Dorfeshan and S. M. Mousavi, “A novel interval type-2 fuzzy decision model based on two new versions of relative preference relation-based MABAC and WASPAS methods (with an application in aircraft maintenance planning),” *Neural Computing and Applications*, vol. 32, no. 8, pp. 3367–3385, 2020.

[55] S.-m. Yu, J. Wang, and J.-q. Wang, “An interval type-2 fuzzy likelihood-based MABAC approach and its application in selecting hotels on a tourism website,” *International Journal of Fuzzy Systems*, vol. 19, no. 1, pp. 47–61, 2017.

[56] X. Peng and Y. Yang, “Pythagorean fuzzy Choquet integral based MABAC method for multiple attribute group decision making,” *International Journal of Intelligent Systems*, vol. 31, no. 10, pp. 989–1020, 2016.

[57] A. Onan, S. Korukoğlu, and H. Bulut, “Ensemble of keyword extraction methods and classifiers in text classification,” *Expert Systems with Applications*, vol. 57, pp. 232–247, 2016.

[58] D. Kim, D. Seo, S. Cho, and P. Kang, “Multi-co-training for document classification using various document representations: TF-IDF, LDA, and Doc2Vec,” *Information Sciences*, vol. 477, pp. 15–29, 2019.

[59] D. Diakoulaki, G. Magvotas, and L. Papayannakis, “Determining objective weights in multiple criteria problems: the critic method,” *Computers and Operations Research*, vol. 22, no. 7, pp. 763–770, 1995.
[61] Y.-P. Jiang, H.-M. Liang, and M. Sun, “A method for discrete stochastic MADM problems based on the ideal and nadir solutions,” *Computers and Industrial Engineering*, vol. 87, pp. 114–125, 2015.

[62] L. Wang, H. Garg, and N. Li, “Pythagorean fuzzy interactive Hamacher power aggregation operators for assessment of express service quality with entropy weight,” *Soft Computing*, vol. 25, no. 2, pp. 973–993, 2021.

[63] H. Cao, “Online review manipulation by asymmetrical firms: is a firm’s manipulation of online reviews always detrimental to its competitor?” *Information and Management*, vol. 57, no. 6, Article ID 103244, 2020.

[64] H. Li, F. Meng, and B. Pan, “How does review disconfirmation influence customer online review behavior? a mixed-method investigation,” *International Journal of Contemporary Hospitality Management*, vol. 32, no. 11, pp. 3685–3703, 2020.