RESEARCH ARTICLE

A Novel Hotel Selection Decision Support Model Based on the Online Reviews from Opinion Leaders by Best Worst Method

Jian Wu[1] · Chenhang Liu[1] · Yang Wu[1] · Mingshuo Cao[1] · Yujia Liu[1]

Abstract

Hotel selection is an important decision in making travel plans. Since hotel selection is a typical non-expert decision, online reviews provide people with information about the hotels and travel destinations they never went to. Several studies construct the decision model based on online reviews with the subjective weights of criteria but ignore the objective weight of criteria derived by opinion leaders, which contributes to the review helpfulness. This study proposes a decision-making model based on online reviews for satisfactory hotel selection. Firstly, an RFMP model is proposed to extract the online reviews of opinion leaders, and the Word2vec method is used to extract the criteria from the online review of opinion leaders. Secondly, obtain the objective weight of criteria from the online reviews of opinion leaders by Word2Vec. Meanwhile, obtain the subjective weight of criteria by the best worst method(BWM) method. Thus, the weight of the hotel selection criteria can be obtained by a linear weighting of objective and subjective weight with a parameter. Thirdly, the hotel selection process based on TOPSIS is employed. Finally, a case study of 8 alternative hotels on Mafengwo.com is applied to verify the proposed model. Comparison experiments and sensitivity analysis are given to illustrate the reasonableness and advantage of the proposed model.

Keywords Online reviews · Opinion leader identification · Best-worst method (BWM) · Multi-criteria decision making · Hotel selection

1 Introduction

Hotel selection is an important decision of making a travel plan, which is directly related to the satisfaction of the travel experience. Since most tourists know nothing about the destination, they may find it very difficult to select a satisfactory hotel. With the development of online hotel booking websites and social media such as booking.com, Airbnb.com, facebook.com, online reviews of hotels, which reflect the real experiences and opinions of consumers [2], have exploded [1]. Benefit from the online reviews, tourists can learn information about the alternative hotels in their travel destinations without being there [3]. About 88% of customers tend to check online reviews before purchasing a product, and the percentage is growing [4]. Meanwhile, customers tend to share the purchase experience after buying products by leaving online reviews on websites and social media. Due to a large number of online reviews on websites, potential customers need to spend a lot of time and effort on identifying the helpful information of various products and services in the hospitality industry [5], such as travel products and restaurant services [6]. The method of review helpfulness evaluation attracted a lot of attention by researchers [7, 8]. Fake online review is also a kind of useless information that needs to be considered to delete from data sets [9].

Multi-criteria decision-making is a wildly used method in hotel selection with online reviews [10]. To develop reasonable and effective hotel selection methods, the existing studies focused on three main problems: Firstly, some researchers proposed sentiment analysis methods of online reviews for hotel selection [11]. Secondly, the criteria extraction and

✉ Yujia Liu
lyj71@126.com

Jian Wu
jyajian@163.com

Chenhang Liu
liuchenhangup@126.com

Yang Wu
yangwu@shmtu.edu.cn

Mingshuo Cao
caomingshuo0606@163.com

1 School of Economics and Management, Shanghai Maritime University, Shanghai 201306, China

prioritization methods of hotel selection are the main issues that have been studied by researchers [21]. Zhang et al. [12] extracted 8 criteria and 149 keywords related to them from online. Bi et al. [13] used 6 criteria and found asymmetric effects of attribute performance. Thirdly, some researchers are interested in group tourism problem [3]. However, the existing studies have the following issues to be addressed:

The first issue is to obtain the helpful reviews and criteria of hotel selection [15]. the helpfulness of online reviews is ignored by the proposed hotel selection methods. The online reviews of opinion leaders provided a solution for the above problem. RFM is an effective method for identifying key customers in marketing [25, 26]. For identifying the opinion leaders, the social network is a vital factor which can not be ignored. Furthermore, criteria of hotel selection obtained from the opinion leaders will be more helpful.

The second issue is to obtain the weight vectors of criteria. In real life, as hotel selection is a non-expert decision-making problem, even if tourists know the criteria for selecting a hotel, the weights of these criteria will vary depending on the travel destination. For example, public transportation in specific travel destinations is underdeveloped, and the "location" should be given a higher weight. Conventional researches only considered either objective weight vectors of criteria (extract from the online reviews) or subjective weight vectors (given by travelers), which is not comprehensive. Besides, the subjective weighting method is difficult to ensure the consistency of the weight vector. The BWM has an advantage in ensuring the consistency of criteria weights by reducing comparisons time [14].

To address the above research gaps, this paper proposes a novel hotel selection decision support model based on the online reviews from opinion leaders. The main purposes of this study are as follows:

(1) To obtain the helpful online reviews, a RFMP model is proposed to identify the opinion leaders among the customers who leave reviews on the hotel booking websites. Then, hotel selection criteria can be extract by Word2Vec method from the online reviews of opinion leaders.
(2) To obtain the weight vectors of criteria, a subjective and objective weighting method is proposed. The objective weight vectors are obtained from the opinion leaders. The subjective weight vectors are obtained by the BWM.

Thus, this study is organized as follows: Section 2 investigates an overview of the related work. Section 3 illustrate the proposed methodology for hotel selection with online reviews. Section 4 presents a case study of the Mafengwo.com to illustrate the application of the proposed method. Comparative analyses is also presented in this section.

Section 5 presents the Result and sensitivity analysis. Finally, conclusion and future work are presented in Sect. 6.

2 Literature Review

Previous research on hotel choice has focused on two categories [27, 28]. The first type of research aspect emphasizes the criteria that tourists may consider when booking and selecting a hotel, and the model based on hotel choice can help tourists is an important research aspect that has witnessed a growing number of cases supported by the fuzzy MCDM approach. Xu and Li [29] state that the most valuable criteria that may significantly influence visitors' decisions are room quality, staff attitudes and behavior, location, transportation, value, and food. Yen and Tang [30] specifically analyze the impact of hotel criteria on eWOM behavior and provide evidence to support the relevance of these criteria to hotel performance. Kwok and Lau [31] explored an improved TOP-SIS-based decision support algorithm for tourists to choose Adal [32] introduced a new integrated MCDM approach in this context, in which a stepwise weight assessment ratio analysis method was used to determine the criteria weights. An operational competitive rating analysis was also used to find the optimal hotel. Yu et al. [33] developed an MCDM model based on Vlsekriterijumska optimizacija I KOmpromisno Resenje (Serbian) (VIKOR) using online rating-based LDAs to solve the hotel selection problem. Peng et al. [27] used probabilistic linguistic term sets (PLTSs) converted from customer ratings to represent evaluation information and investigated a cloud decision support model to search for the best choice for tourists. They confirmed the advantages of applying PLTSs and LDAs in a hotel setting. However, because users have not been to the local area, users are not as confident in their evaluation criteria weights, they need to refer to the evaluation criteria of opinion leaders, and the evaluation criteria weights are too complex, and some current traditional methods have many shortcomings [38].

As hotel selection is a non-expert decision- making problem, online review provide decision-making opinions and weight information for tourists who have never book the alternative Hotels. The main issue is how to obtain the helpfulness online reviews to better leverage the impact of online evaluation on decision-makers. In general, the validity of online reviews is analyzed predominantly with two approaches. Firstly, Handcrafted features like structural statistics, the sentiment, features of online reviews are used to predict or identify the helpfulness of online reviews [15–18]. In some studies, comments are classified according to the quality of the information, readability, and subjectivity of the text content [19, 20]. Secondly, the automatic feature extraction process base on deep learning [21]. Chen [22] conducted an sentiment analysis based method and concluded

that negative emotions attract more helpful votes, while male readers prefer comments with positive emotions. Although both statistical and NLP-based approaches were incorporated by Zhou [24], the semantic understanding of the comment content remained unattended.

The existing methods of obtaining weights for hotel selection are mainly divided into two types: given subjectively by decision-makers and objective weights calculated using machine learning algorithms. There are two main problems in the research in online reviews regarding the acquisition of weights, namely, the method of weight acquisition is complicated and does not take into account both subjective and objective weights [36]. Rezaei [14] developed a Multi-criteria Decision Making (MCDM) method named the Best Worst Method (BWM). Compared to the Analytic Hierarchy Process (AHP), one of the MCDM methods based on pairwise comparisons, BWM requires less comparative data while producing more consistent comparisons, allowing it to produce more reliable results based on previous analyses [14]. Due to its simplicity and reliability, BWM has been widely applied to solve a range of different problems [34, 35]. Besides, in the design of weights few past studies of online hotel evaluations have considered both subjective and objective weights. Nowadays, decision-makers want to seek the decision-making opinions of opinion leaders on the one hand, and ensure that their preferences are reflected on the other [37].

3 Methodology

3.1 Problem Description

To help travelers select the most appropriate hotel in the travel destination they have never been, a framework of hotel selection decision support model based on the online reviews

from opinion leaders is constructed. The proposed framework are consisted of the following four parts: (1) Obtain the opinion leaders and their online reviews by the proposed RFMP method; (2) Extract the hotel selection criteria from online reviews of the opinion leaders by Word2Vec method; (3) Calculate the weight vectors of criteria which is a linear weighting of objective and subjective weight with a parameter.(4) Select the ideal hotel by TOPSIS method. The proposed framework of hotel selection decision support model based on the online reviews from opinion leaders can be shown in Fig. 1,

The following notations are used to denote the sets and variables in the problem, which will be used throughout this paper:

- $A = \{A_1, A_2, ..., A_m\}$: the set of m hotels, where A_i denotes the i-th alternative, $i = 1, ..., m$.
- $C = \{C_1, C_2..., C_n\}$: the set of n hotel criteria, where C_j denotes the j-th criterion, $j = 1, ..., n$.
- $W_j = \{W_1, W_2, ..., W_n\}$: the vector of criteria weights, where W_j denotes the weights of criteria C_j, $W_j \geqslant 0$ and $\sum_{j=1}^{n} W_j \geqslant 0$, $j = 1, 2, ..., n$
- $U\_ID = \left\{ User_h^{A_j} \right\}_{k \times n}$: the users set, where $User_h^{A_j}$ denotes the h-th user who gave online reviews of A_j hotels, $h = 1, ..., k$, $j = 1, ..., n$.

3.2 Data Preparation

Since the identification of opinion leaders improves the helpfulness of online reviews on the online travel platforms,This study crawled the online hotel reviews from the online travel platforms, as well as the features and social network centrality of users, including online evaluation time, evaluation score, evaluation text, numbers of evaluations for a moment, impact value of online reviews for a moment and user levels, user fans, number of user followings, which are
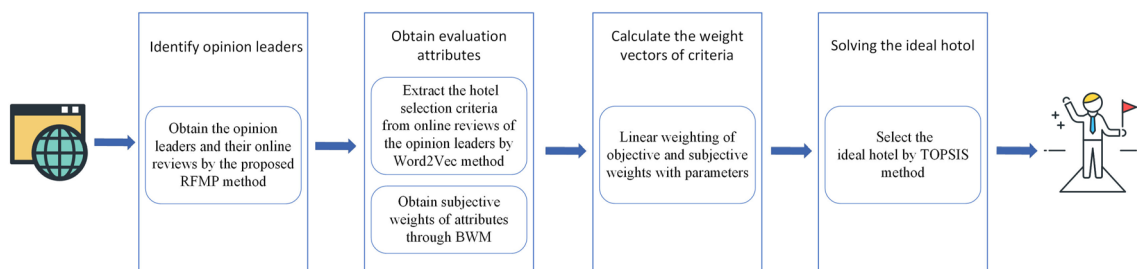

Fig. 1 The structure of a novel hotel selection decision support model

Table 1 List of Abbreviations

Proprietary nouns	Abbreviated nouns
Online travel platforms	*OTPs*
Online evaluation time	*OET*
Evaluation score	*ES*
Evaluation text	*ET*
Numbers of evaluations for a moment	*NEMs*
Impact value of online reviews for a moment	*IVORs*
User levels	*ULs*
User fans	*UFs*
Number of user followings	*NUFs*

pre-processed to filter opinion leaders for subsequent analysis using the RFMP model. The abbreviations of each proper noun are shown in the following Table 1.

In this study, as Mafengwo.com is one of the leading OTPs in China, it is selected to crawl the user online reviews and the related users' data of multiple hotels in a specified destination using python crawler and store the crawled data by applying certain rules. The crawled data need to be pre-processed by the following steps:

Step 1: Crawling for OHRs

The crawled dataset includes hotel name, user ID, ET, ES, OET, time of the last review, number of reviews in three years, IVORs, ULs, UFs, NUFs.

Step 2: Deleting incomplete online reviews

Each valid data crawled is made up of 11 dimensions, and we do a cleanup of all the data crawled in Step 1, eliminating data with missing dimensions.

Step 3: Standardizing data

Definition 1 Date of last online reviews (DLORs), NT means now time.

$$T^{user_h} = \{OET_1, OET_2, OET_3, OET_4, ..., OET_n\}, h = 1, 2, 3, ..., k \quad (1)$$

$$T^{user_h}_{DLORs} = \min\{NT - OET_h\}, h = 1, 2, 3, ..., k \quad (2)$$

where OET_j denotes the jth comment of the h-th user, and T^{user_h} denotes the set of online reviews of the ith user, $T^{user_h}_{DLORs}$ denotes the time interval between the last evaluation of the h-th user and the current one.

After getting the DLORs for each user, this paper standardizes on different dimensions of each data item, eliminating the effect of different measures on the metrics.

$$S\_Data = \begin{cases} \frac{Data_h^{p_j} - Data_{\min}^{p_j}}{Data_{\max}^{p_j} - Data_{\min}^{p_j}} & p_j = \{NEMs, IVORs, ULs, UFs, NUFs\} \\ -\frac{Data_h^{p_j} - Data_{\min}^{p_j}}{Data_{\max}^{p_j} - Data_{\min}^{p_j}} & p_j = T^{user_h}_{DLORs} \quad h = 1, 2, ..., k \end{cases}$$

$$(3)$$

where $Data_h^{p_j}$ denotes the p-th value of the p_j-th indicator, $Data_{\max}^{p_j}$ and $Data_{\min}^{p_j}$ denote the h-th max-value and min-value of the p_j-th indicator respectively.

3.3 Best-Worst Method (Algorithm 1)

This paper adopts the Best Worst Method (BWM) [27], which was proved to be superior to the AHP method with respect to the computational complexity. The computational procedure of the method contains the detailed steps of the six processes involved in the proposed method. They are as follows:

Step 1. Defining a set of evaluation criteria: $C = \{C_1, C_2, ..., C_n\}$.

Step 2. Selecting the best and the worst criteria, respectively.

Step 3. Determine the priority of the best criterion over each of the other criteria

Using a number between 1 and 9 to determine the preference of the best criterion over all the other criteria . The resulting Best-to-Others vector (BOV) would be: $A_B = \{a_{B1}, a_{B2}, ..., a_{Bn}\}$, where a_{Bj} indicates the preference of the best criterion over criterion C_j. It is clearly indicated that $a_{BB} = 1$.

Step 4. Determine the priority of each criterion over the worst criteria

Using a number between 1 and 9 to determine the preference of all the criteria over the worst criterion. The resulting Others-to-Worst vector (OWV) would be: $A_W = \{a_{1W}, a_{2W}, ..., a_{nW}\}^T$, where a_{jW} indicates the preference of the criterion j over the worst criterion W. It is clearly indicated that $a_{WW} = 1$.

The values of BOV and OWV are shown in Table 2.

Step 5. Solve the optimal weights solution $W^* = \{W_1^*, W_2^*, ..., W_n^*\}$. The optimal weight for the criteria is the one where, for each pair of W_B/W_j and W_j/W_w, we have $W_B/W_j = a_{Bj}$ and $W_j/W_W = a_{jW}$. To satisfy these conditions for all j, we should find a solution where the maximum absolute differences $\left|\frac{W_B}{W_j} - a_{Bj}\right|$ and $\left|\frac{W_j}{W_W} - a_{jW}\right|$ for all j is minimized. Considering the non-negativity and sum condition for the weights, the following problem is resulted:

Table 2 Pairwise comparison vector for BOV and OWV

criteria	C_1	C_2	C_3	⋯	C_n
Best criterion	a_{B1}	a_{B2}	a_{B3}	⋯	a_{Bn}
Worst criterion	a_{1W}	a_{2W}	a_{W3}	⋯	a_{Wn}

$$\min_{j} \max \left\{ \left| \frac{W_B}{W_j} - a_{Bj} \right| \text{ and } \left| \frac{W_j}{W_W} - a_{jW} \right| \right\}$$

$$s.t. \begin{cases} \sum_j W_j = 1 \\ W_j \geq 0, \text{for all } j \end{cases}$$

(4)

Where W_j denotes the criteria weight. W_B and W_w denotes the best weight, the worst weight, respectively. a_{Bj} denotes the preference score of the best criteria for the j-th criteria, similarly, a_{jW} denotes the preference score of the j-th criteria for the worst criteria. The problem can be transferred to the following problem:

$$\min \delta$$

$$s.t. \begin{cases} \left| \frac{W_B}{W_j} - a_{Bj} \right| \leq \delta \\ \left| \frac{W_j}{W_W} - a_{jW} \right| \leq \delta \\ \sum_j W_j = 1 \\ W_j \geq 0, \text{for all } j \end{cases}$$

(5)

Where W_j denotes the criteria weight. W_B and W_w denotes the best weight, the worst weight, respectively. a_{Bj} denotes the preference score of the best criteria for the j-th criteria, similarly, a_{jW} denotes the preference score of the j-th criteria for the worst criteria, and the δ is an infinitesimal value.

Solving the problem, the optimal weights $W^* = \{W_1^*, W_2^*, ..., W_n^*\}$ and δ^* are obtained

Step 6. Calculate the consistency ratio and improve the consistency of criteria weight which have been obtained. Next, the consistency of the obtained results is verified. For the consistency ratio, determining firstly as, then the consistency index(CI) can be found through the established data relationship.

$$CI = \frac{\lambda_{\max} - n}{n - 1}$$

(6)

Where $\lambda_{\max}$ is the maximum eigenvalue of the matrix.

Then, by calculating the ratio of a and the consistency index, it is the required consistency ratio (CR).

$$CR = \frac{\delta^*}{CI^{a_{BW}}}$$

(7)

When $CR \leqslant 0.1$, the expected level of consistency is achieved. Otherwise, the consistency of preference relations (PRs) can be improved by modifying some values in the PRs. The smaller the CR value obtained, the better the consistency and the more scientific the solution [27]

Thus, the pseudo code of the Best Worst Method is shown in **Algorithm 1** for obtaining the criteria weight of hotel selection.

Algorithm 1: Using BWM to Solve Fuzzy Multi-criteria Decision Problems

---

begin

    Input: A set of criteria $C = \{C_1, C_2, ..., C_n\}$;The best and worst criterion were selected respectively; $BOV = A_B = \{a_{B1}, a_{B2}, ..., a_{Bn}\}$ and $OWV = A_W = \{a_{1W}, a_{2W}, ..., a_{nW}\}^T$ based on the decision maker's preference is gotten;

    Output: Ranking of alternatives;

    1: Select a set of evaluation criteria

    2: Identify the best criteria, determining the preference of the best criterion over all the other criteria $BOV = A_B = \{a_{B1}, a_{B2}, ..., a_{Bn}\}$

    3: Identify the worst criteria, determining the preference of all the criteria over the worst criterion $OWV = A_W = \{a_{1W}, a_{2W}, ..., a_{nW}\}$

    4: Mathematical Modeling of Multi-Objective Nonlinear Planning from BOV and OWV

    5: Solve the model

        If this model can be addressed:

            Obtain $W* = \{W_1^*, W_2^*, ..., W_n^*\}$ and δ^*

        else:

            revise BOV and OWV, reconstructing the solution model

    6: Judging Consistency $CR = \frac{\delta^*}{CI^{a_{BW}}}$

        If $CR \geq$ Threshold:

            The weight of the solution is not reasonable, PRs can be improved by modifying some values

        else:

            The weight of the solution is reasonable

end

3.4 Extract the Online Text Review of Opinion Leaders

3.4.1 Basic Marketing Model (RFM)

According to Arthur Hughes of the Database Marketing Institute, three magic elements in the customer database make up the best metrics for data analysis. the RFM model based on three important indicators of customer behavior,is utilized to analyze customer value, in which, R represents recency, F represents frequency, and M represents monetary.

The RFM analysis is based on the following assumptions:(1) Customers who have recently made purchases are more likely to purchase again than customers who have not made purchases recently;(2) Customers who purchase more frequently are more likely than those who purchase less frequently. May buy the company's products (services) again; (3) Customers with a higher total purchase amount are more likely to buy again and are customers with a higher value.

Definition 2 (Recency) For a $Reviewer_h$, the recency of his consumption on OTPs can be expressed as:

$$R^{user_h} = T_{DLORs}^{user_h}, h = 1, 2, 3, ..., k \tag{8}$$

where R^{user_h} denotes the recency of the h-th user to posting online reviews in OTPs.

Recency represents the time of last consumption, which indicates the time since the user last consumed lasting to the present. The more recent the consumption time, the greater the customer value. Here, time is measured in days. And there is a problem here, as R indicates the length of time since the users latest consume, the closer the consumption time, the greater the customer value.

Definition 3 (Frequency) For a *Reviewer*$_h$, the frequency of his consumption on OTPs can be expressed as:

$$F^{user_h} = \sum_{l=1}^{e} Reviewer_l^{user_h}, \, l = 1, 2, ..., e; h = 1, 2, ..., k \quad (9)$$

where F^{user_h} denotes the frequency of the h-th user to posting online reviews in OTPs, and Re*viewer*$_l^{user_h}$ denotes the l-th online rating posted by the h-th user in OTPs. Frequency indicates how many times the user has spent within a certain period, which can reflect user activity.

Definition 4 (Monetary) For a *reviewer*$_h$, the monetary of his consumption on OTPs can be expressed as:

$$M^{user_h} = \sum_{l=1}^{e} Mone_l^{user_h}, \, l = 1, 2, ..., e; h = 1, 2, ..., k \quad (10)$$

where M^{user_h} denotes the the power of impact of the h-th user to posting online reviews in OTPs, and $Mone_l^{user_h}$ denotes the l-th online rating impact posted by the h-th user in OTPs. The amount of money a user spends over some time. A score based on the transaction amount, the higher the transaction amount, the higher the score. Reflects customer value.

3.4.2 RFMP Medel

Based on the RFM marketing model, this paper considers the popularity (P) as a new indicator, and proposes the RFMP model to aggregate criteria for each online review, applies the BWM to address multi-criteria decision problems to measure weights and the influence value of online review. To measure the publisher of online review credibility, the overall evaluation helpfulness of the publisher of online review under the RFMP model can be expressed as Eq. 11:

$$S_{user_h} = w_R \times \overline{S_{R_h}} + w_F \times \overline{S_{F_h}} + w_M \times \overline{S_{M_h}} + w_P \times S_{P^{user_h}},$$
$$h = \{1, 2, ..., k\}$$
$$(11)$$

where the $\overline{S_{R_h}}, \overline{S_{F_h}}, \overline{S_{M_h}}$ were the data after $S_{R_h}, S_{F_h}, S_{M_h}$ standardization respectively, $S_{P^{user_h}}$ was the solution from the above model.

The parameter P is a function of three indicators: ULs, UFs, and NUFs. Since the contribution of these three indicators is not the same, which weight allocation to assemble them is one of the first issues that the model must address. According to **Algorithm 1**, this paper selected an criteria set $BOV = A_B = \{a_{NF_{s,1}}, a_{NF_{s,2}}, a_{NF_{s,3}}\}$ and $OWV = A_W = \{a_{1,ULs}, a_{2,ULs}, a_{3,ULs}\}^T$ also are given by experts to assess the degree of significant preference between criteria. The centrality score of the publisher of online review can be expressed as:

$$S_{P^{user_h}} = w_{ULs} \times \overline{S_{ULs}} + w_{UFs} \times \overline{S_{UFs}} + w_{NUFs} \times \overline{S_{NUFs}},$$
$$h = \{1, 2, ..., k\}$$
$$(12)$$

where the $\overline{S_{ULs}}, \overline{S_{UFs}}, \overline{S_{NUFs}}$ were the data $S_{ULs}, S_{UFs}, S_{NUFs}$ after standardization respectively.

Similar to previous studies that applied RFM analysis to calculate customer value, the value of the publisher of online review is measured here.

The best and worst criteria are selected from the set of $C = \{R, F, M, P\}$ under the expert's preference, and the expert is given the BOV and OMV to compute the respective weights of the criteria based on **Algorithm 1**. the process of data standardized is implemented in the data pre-processing to eliminating the influence of different data frames.

We cluster the four criterion, calculate the weights of each indicator, and multiply each indicator score. The publisher of online review with high overall scores is selected by setting thresholds and is considered opinion leaders.

3.5 Obtaining the Weight Vector of the Hotel Selection Criteria

This study utilized the following six first-level criteria of hotel selection: Position, Service, Cleanliness, Comfort, Facility, and Food, which are crawled from the mafengwo.com. Based on the crawled criteria, the word2vec algorithm is utilized to aggregate the preferences of the opinion leaders with them to calculate the objective evaluation criteria, which are the objective weight vectors proposed in this paper. Besides, the subjective weight vectors are obtained by calculating the preferences of each decision-maker based on the BWM.

Hotel selection is a multi-criteria decision making problem based on online reviews. Tourists who have already consumed a certain hotel used to leave an online review of several criteria on the website based on their living experience. However, it is unrealistic to expect all users to comment on a particular criteria in the same words. Some examples of hotel reviews show that both "It's so convenient to go to the subway station" and "The hotel is in the center of the city,

very convenient transportation" is about the position, but use different words. Therefore, words describing the same criteria need to be organized into a first-level criteria word, which includes a number of feature words.

In this subsection, the Word2Vec algorithm is used to construct a first-level criteria set of hotel selection and its related feature words. Since the Word2Vec algorithm is used to determine the similarity between two text data sets, in this study, it is used to calculate the similarity in semantic space between a feature word extracted from an online review and the first-level criteria words and the semantic space similarity between the first-level criteria words. Words with higher similarity are constructed as the corresponding

criteria dictionaries. The specific process is as follows. First, each review in review sets $S = \{s_1, s_2, ..., s_e\}$ needs to be segmented, using the word segmentation technology, into several words according to dependency parsing. Next, the words are converted into corresponding space vectors and then the top N words with the closest similarity to each first-level criteria word are obtained using the the Word2Vec algorithm. This step is implemented based on the word2vec. Word2Vec() method of the gensim.model in Python 3.7. After manually removing some unqualified words from it, the criteria dictionary sets $S = \{s_1, s_2, ..., s_e\}$ is obtained. The pseudo-code of the algorithm for extracting criteria set is shown in **Algorithm 2**.

The following Sects. 3.5.1 and 3.5.2 focus on the acquisition and use of subjective and objective weights

---

Algorithm 2: Extracting and clustering textual information based on Word2Vec algorithms

---

begin

    Input: A set of online review text $S = \{s_1, s_2, ..., s_e\}$; $l = 1, 2, ..., e$ and criteria set

    $D = \{D_1, D_2, ..., D_6\}$

    Output: A set of online review word

    $D_j = \{word_1^{s_i}, word_2^{s_i}, ..., word_j^{s_i}, word_n^{s_i}\},$ *where* $word_j^{s_i}$ *means*

    the $word_j$ in the s_i online review based on OTPs

1: **for** online review in S:

2:    separate sentence to Words in online review based on jieba algorithms

3:    **return** Words

4:    **for** word in Words:

5:      load the word2vec.Word2Vec() to Similarity Values through Dj

6:      **return** top N words with high similar on Dj

7:      **if** word Dj :

8:        classify word to Dj

9:      **else**:

10:        pass

end

3.5.1 Obtaining Objective Weights of the Criteria

Since hotel selection is a typical non-expert decision-making problem, the identification of opinion leader helps decision maker filter out the helpfulness online review and provide objective evidences for determining the weight vector of criteria. Based on the word2vec, opinion leaders (OLs) in the OTPs can be identified, and the objective weight vectors in the evaluation model rely mainly on analyzing the online review texts of the OLs. The online comments of OLs

are analyzed in the following. This subsection contains the detailed steps of the four processes involved in the proposed algorithm. This subsection contains the detailed steps of the four processes involved in the proposed algorithm. They are as follows.

Step 1. Extract OLs' online reviews.

Each evaluation online review is addressed separately. Implemented using the third-party library pandas for Python data analysis.

Step 2. Text pre-processing.

Three basic procedures (tokenization, removal of capitalization, and removal of stop words and non-characters) were used to process the crawled text comments. The aim of Tokenization is to break up text into words, phrases, or other tokens using the third-party library jieba for Python. In this study, each textual comment was broken down into words. Stop words (e.g., a, an, the, of, for he, and she), which do not contribute significantly in text mining, should be filtered before text processing. The applied stop words include the existing list of stop words (see https://github.com/goto456/stopwords). Non-Chinese characters were also removed, such as $[, .?!() ： ;']$ and emoji.

Step 3. Extracting and classify textual online reviews to $D_j = \{word_1^{s_l}, word_2^{s_l}, ..., word_j^{s_l}, word_n^{s_l}\}$ based on **Algorithm 2**

Step 4. Obtain the objective weight vectors based on Eq. 13.

$$P(X = D_j) = \frac{\sum_{j=1}^{6} \sum_{l=1}^{e} word_j^{s_l}}{\sum_{l=1}^{e} word_j^{s_l}}, \tag{13}$$

where $j = \{1, 2, 3....6\}$ are the number of criteria

In this subsection, we need to calculate the weights of the objective criterion and used the percentage to approximate the objective weights of each criterion in the paper.

3.5.2 Obtaining Subjective Weights of the Criteria

Every traveler has his/her own preferences in hotel selection, and it is incomplete to aggregate the decision preference using only objective weight vectors obtained from the online reviews of opinion leaders. How do decision makers reflect their preferences, which is another issue considered in this paper? In this paper, the decision support model we provide is that decision makers can obtain expert's preference opinions through online reviews of OLs, and this information is used as objective weights for the criteria. Meanwhile, the decision maker gives his/her own preferences, and the subjective weights of criteria can be calculated based on the BWM method using personal preferences. The steps to solve the subjective weights of criteria are as follows:

Step 1: Selecting the most desirable and the least important criteria, respectively.

Step 2:   Determining $BOV = A_B = \{a_1, a_2, ..., a_n\}$ and $OWV = A_W = \{b_1, b_2, ..., b_n\}$ to assess the degree of PRs between criteria.

Step 3: Calling **Algorithm 1** to solve for the subjective weights of the criteria.

Step 4: Calculate the consistency of the subjective weight vectors. calculate the consistency ratio (CR) by Eq. 7

The weights derived from the steps above are considered as an important component of making the final decision weights, which is a collection of OLs' recommendations and personal preferences through the weight aggregation method.

3.5.3 Obtaining Combined Weight Vector

We calculated the objective weight vector and subjective weight vector respectively, and the common combined weight vector(CWAs) formula can be expressed as:

Through the above, the CWAs of each criteria are solved as shown in Eq. 14.

$$W_j = W_j^{OWAs} \times \beta + W_j^{SWAs} \times (1-\beta) \tag{14}$$

Where OWAs means objective weight of criteria, SWAs means subjective weight of criteria, $\beta \in [0, 1]$ is a parameter to control the degree of objective weights and subjective weights.

3.6 The Hotel Selection Process

Considering the stability of the decision solution in the decision-making process, the TOPSIS method is proposed to help travelers to select hotels and analyze their decision results. TOPSIS method was first proposed by C.L. Hwang and K. Yoon in 1981 and is a sequential method that approximates the ideal solution [45]. It consists of a 6-step solution process. The step-by-step description of the TOPSIS method hotel selection process is shown below:

Step 1: Constructing a preference matrix

The decision-maker created Matrix A, and is shown as formula 15:

$$A_{ij} = \begin{bmatrix} a_{11} & a_{12} & ... & a_{1n} \\ a_{21} & a_{22} & ... & a_{2n} \\ ... & ... & ... & ... \\ a_{m1} & a_{m2} & ... & a_{mn} \end{bmatrix} \tag{15}$$

In the matrix, *m* represents the number of decision points and *n* represents the number of evaluation criteria.

Step 2: Constructing a Standard Decision Matrix (R)

Using the elements of matrix A and the following formula 16 to determine the elements of The Standard Decision Matrix :

$$r_{ij} = \frac{a_{ij}}{\sqrt{\sum_{i=1}^{m} a_{ij}^2}} \qquad i = 1, 2,..., m; \quad j = 1, 2,..., n; \tag{16}$$

The matrix R is defined by the matrix shown below:

$$R_{ij} = \begin{bmatrix} r_{11} & r_{12} & ... & r_{1n} \\ r_{21} & r_{22} & ... & r_{2n} \\ ... & ... & ... & ... \\ r_{m1} & r_{m2} & ... & r_{mn} \end{bmatrix} \tag{17}$$

Step 3: Creating ideal (R^+) and negative ideal ($R-$) solutions:

Find the ideal solution and the set of negative solutions, as shown in Eq. 18 below:

$$R^+ = \{(maxr_{ij}|j \in J_1), (minr_{ij}|j \in J_2)\} = \{r_1^+, r_2^+, ..., r_n^+\}$$
$$R- = \{(minr_{ij}|j \in J_1), (maxr_{ij}|j \in J_2)\} = \{r_1^-, r_2^-, ..., r_n^-\} \tag{18}$$

Where J_1 is a benefit indicator and J_2 is a cost indicator

Step 4: In the TOPSIS method, the Euclidean distance method is used to find the deviation of the evaluation factor values at each decision point from the ideal solution and the set of negative ideal solutions. The calculation of the ideal discrimination (S^+) measure is shown in the formula 19 and the calculation of the negative ideal discrimination ($S-$) measure is shown in the formula 19:

$$S_i^+ = \sqrt{\sum_{j=1}^{n} [r_{ij}w_j - r_j^+ w_j]^2} = \sqrt{\sum_{j=1}^{n} w_j^2 [r_{ij} - r_j^+]^2}$$
$$i = 1, 2, ..., 8; \quad j = 1, 2, ..., 6;$$
$$S_i^- = \sqrt{\sum_{j=1}^{n} [r_{ij}w_j - r_j^- w_j]^2} = \sqrt{\sum_{j=1}^{n} w_j^2 [r_{ij} - r_j^-]^2}$$
$$i = 1, 2, ..., 8; \quad j = 1, 2, ..., 6; \tag{19}$$

Step 5: The proximity (Ci*) of each decision point relative to the ideal solution are calculated. The calculation of the ideal solution is shown in the following formula 20:

$$C_i^* = \frac{S_i^-}{S_i^- + S_i^+} \tag{20}$$

The value C_i^* is in the range $0 \leq C_i^* \leq 1$ and $C_i^* = 1$ indicates the absolute proximity of the corresponding decision point to the ideal solution.

In the above TOPSIS **Algorithm** (Step 3), this paper adopts a subjective-objective weight set approach to creating the Weighted Standard Decision Matrix, the recommendation system had been obtained the preferences of OLs from a large number of online reviews, calculate the objective weight of criteria based on the preferences of the OLs using **Algorithm 2**, then take into account the preferences of the decision-makers, find the subjective weight of criteria based on **Algorithm 1**, and finally add the β to calculate the combined weights of the criteria and rank the alternatives by TOPSIS. Finally, the stability of the alternative ordering is analyzed by adjusting the parameters β.

3.7 The Proposed Hotel Selection Decision Support Model Based on the Online Reviews by Opinion Leaders

In this part, considering the current demands of decision-makers, the proposed model were introduce in detail, which consists of four stages as illustrated in Fig. 2, namely, data preparation, opinion leaders identification, weighing solution, and decision recommendation, to meeting the needs of decision support. Details of the method are described as follows.

Step 1: Data preparation

Determine the criteria for hotel decisions and crawl user information on OTPs, which includes ULs, UFs, NUFs, Review Recency, Review Frequency, Review Momentary, and Review Content.

Step 2: Opinion leaders identification

The three criteria ULs UFs and NUFs are assembled using the BWM multi-criteria decision method, and the results of the assembly are expressed in terms of Popular user centrality. Based on the RFM model, the parameter P is expanded to become the RFMP model, which is the opinion leader identification model proposed in this paper. Through the scoring value and weight of each indicator, the comprehensive scoring value of each publisher of online review is obtained and the opinion leaders are selected.

Step 3: Obtain the weight vector of criteria

Based on the identification of opinion leaders, the evaluation texts are clustered and learned, and the mathematical frequencies of each indicator are counted, which are used as objective weights of criteria to provide a reference for decision experts. Meanwhile, the individual decision preferences for each criteria are satisfied again by the BWM, which can be used to solve for the weights of each criteria, which are considered here as subjective weights. Finally, the subjective and objective weights are combined, which is used as the basis for solution selection.

Step 4: Selection processes

Considering the rationality and feasibility of the proposed method, the TOPSIS-based scheme ranking method is proposed in this paper. The whole ranking is observed by the change of confidence coefficient.

4 A Case Study and Calculation

In this section, to illustrate the proposed hotel selection decision support model, a case study of Mafengwo.com is conducted based on online reviews by opinion leaders. The data of the site is shown in Fig. 3. The corpus data used in this article comes from Mafengwo.com, which is a typical OTPs, which offers a variety of hotel ranking services, with members who are in the top 100 local rankings in terms of
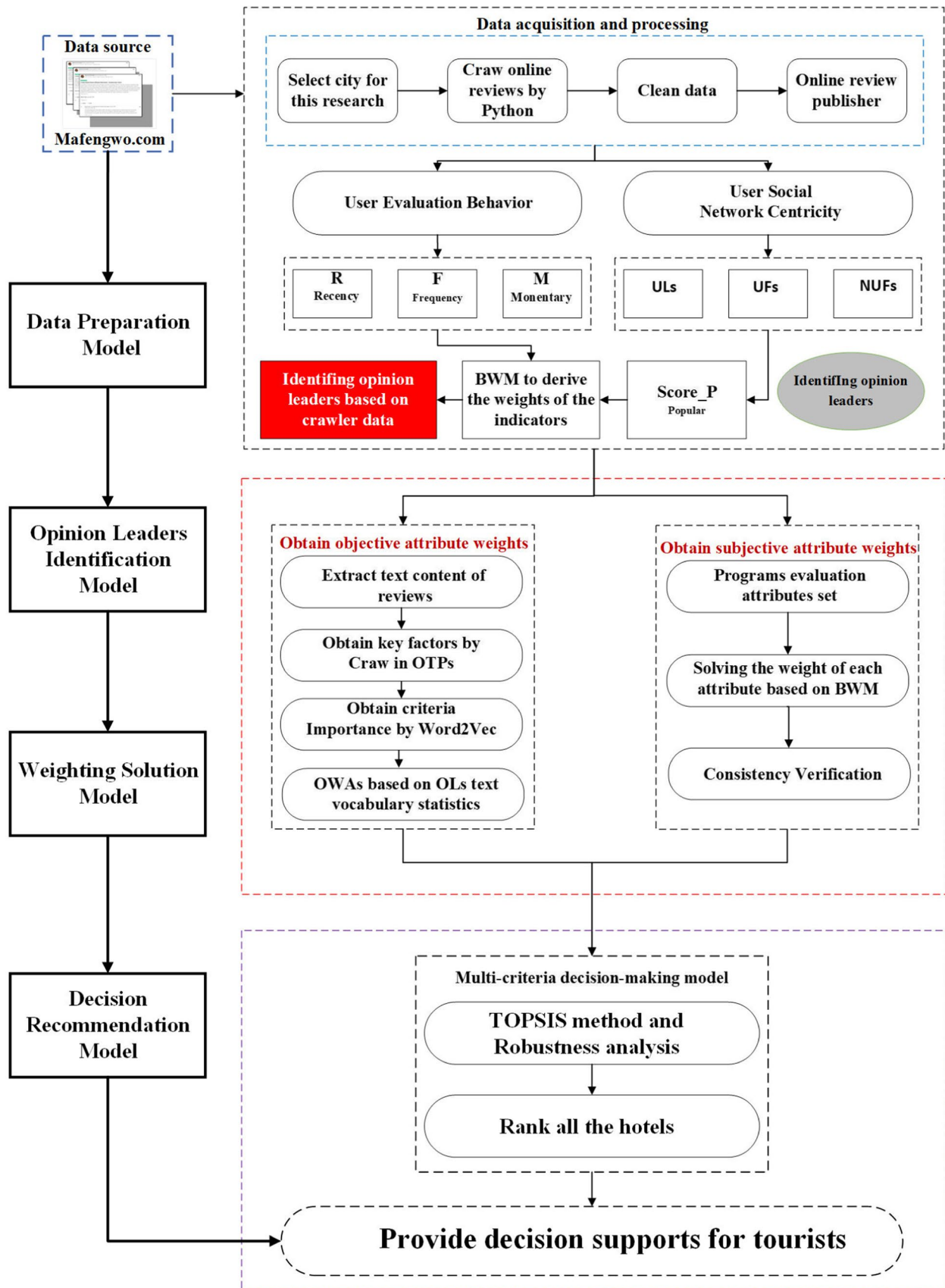
Fig. 2 The structure of the proposed decision support model for hotels

Fig. 3 A screen shot from Mafengwo.com

Table 3 Partial Evaluation Information in a hotel review on Mafengwo.com

H_{ID}	U_{ID}	Contents
A_1	$User_1^{A_1}$	Excellent as always, the location is superior, 5 minutes walk to street, to the Chenghuang Temple and People's Square and other places are very convenient$\cdots$
A_1	$User_1^{A_1}$	The biggest highlight is that it is located in $\cdots$
A_1	$User_2^{A_1}$	This is the first time I stayed at the Expo Intercontinental$\cdots$
A_1	$User_3^{A_1}$	The location of the hotel is very good,$\cdots$
A_1	$User_4^{A_1}$	The hotel is located on Jiujiang Road, next to Wang $\cdots$
...	...	...

total postings on various topics. A growing number of scholars have conducted research based on the online review of Mafengwo.com and can confirm the stability and reliability of the website data.

Firstly, we retrieved all available information from Mafengwo.com listed Shanghai City hotels in April 28, 2020 by Python. To ensure the credibility of our research sample, we crawled online reviews for hotels with at least 6137 reviews.Partial Evaluation Information in a hotel review on Mafengwo.com are shown in Table 3. The data of other hotels A_2 $A_3...A_8$ are the same as A_1, not shown here for now. Here are some categories of information (name, online reviews, OET ES ET NEMs IVORs UFs and NUFs) were crawled, multi-criteria (value, rooms, location, cleanliness, food quality, comfort and service) ratings can download from eight hotels $\{A_1, A_2, A_3, A_4, A_5, A_6, A_7, A_8\}$,and The set of evaluation criteria for each hotel is $C = \{C_1, C_2, C_3, C_4, C_5, C_6\}$ where $C_1, C_2, C_3, C_4, C_5, C_6$ represent position, service, cleanliness, comfort, facility and

food, respectively. The set of users is $U\_ID = \left\{ User_h^{A_j} \right\}_{k \times n}$. The $User_h^{A_j}$ denotes the data of the h-th user in A_j hotels, $j = 1, 2, ..., 8$.

The data of Hotel A_1 is shown in Table 4:

The data of other hotels A_2 $A_3...A_8$ are the same as A_1, not shown here for now.

For example, firstly we found a user named ZYM in A_1, who wrote a online review after booking a hotel on OTPs: "Great location, deep culture, comfortable and elegant facilities, attentive service. The service charge is confusing: it is known that no waiter receives this charge. Then it can be said that the money is unreasonably taken by the hotel. Including the service charge of the hotel, it can be said that it is an unreasonable charge under the name of the hotel. Shanghai is probably the only place in the world that does this. I think that the hotel can increase the fee, but do not increase the price in disguise". Secondly, we visit the user's ZYM profile and download data such as the user's ULs, UFs,

Table 4 Original data sheet

H_{ID}	U_{ID}	⋯	R	F	M	ULs	UFs	$NUFs$
A_1	$User_1^{A_1}$	⋯	1208	1459	100000	45	6853	559
A_1	$User_2^{A_1}$	⋯	1200	826	353391	45	2503	2876
A_1	$User_3^{A_1}$	⋯	98	723	940242	45	6841	135
A_1	$User_4^{A_1}$	⋯	70	252	630816	20	1217	2696
A_1	$User_5^{A_1}$	⋯	1461	13	2599	18	1190	2511
⋯	⋯	⋯	⋯	⋯	⋯	⋯	⋯	⋯

Table 5 Consistency index (CI) table

a_{BW}	1	2	3	4	5	6	7	8	9
Consistency index ($max\tilde{A}\tilde{A}$)	0.00	0.44	1.00	1.63	2.30	3.00	3.73	4.47	5.23

NUFs, and all rating information for that user over time on the platform.

4.1 The Dataset and Data Pre-Process

In this paper, based on an extension of the RFM, the RFMP is used to measure the influence scores of publisher of online review and to select OLs. The P scores of publisher of online review are calculated based on three criteria: UL, UFs, NUFs, and the **Algorithm 1** is used to measure the P scores of publishers of online reviews.

The steps for using **Algorithm 1** are as follows:

Step 1: Define the decision criteria set { ULs,UFs,NUFs} and identify the best(UFs) and worst(ULs) criteria respectively.

Step 2: Determine the BOV and OWV vectors. Calculate BOV = {8,1,2} and OWV = {1,8,5} based on the preference relationship of multiple experts for the indicator

Step 3: Calculate the optimal weights of the criteria Eq. 21.

$$\min \zeta$$

$$s.t. \begin{cases} \left| \frac{W_{UFS}}{W_{ULS}} - 8 \right| \leq \zeta \\ \left| \frac{W_{UFS}}{W_{NUFs}} - 2 \right| \leq \zeta \\ \left| \frac{W_{NUFs}}{W_{ULS}} - 5 \right| \leq \zeta \\ \sum_j W_j = 1 \\ W_j \geq 0, \text{ for all j} \end{cases} \quad (21)$$

Solving the problem, the optimal weights $W^* = \{W_{ULs}^*, W_{UFs}^*, W_{NUFs}^*\} = \{0.072, 0.589, 0.339\}$ are obtained,

Step 4: Calculate the consistency ratio

For the consistency ratio, as the consistency index for this problem is 4.47 (see Table 5), the consistency ratio can

be calculated by Eq. 7. Thus, $CR = 0.058$, which implies a very good consistency.

Remark 1 The consistency index for BWM presents in Table 3 only including the a_{BW} numbers range from 1 to 9.

Next, according to the calculated optimal weights of the three indicators, the P of a score of all publishers of online reviews is calculated, and the calculation result is shown in Table 6.

Similarly, the score of P of rest alternative hotels $\{A_2, A_3, A_4, A_5, A_6, A_7, A_8\}$ can be calculated.

Calculate the influence score of publishers of online reviews using the four criteria of R, F, M and P. Furthermore, M and R is determined as the best and worst criteria, respectively. BOV = { 9,8,1,3} and OWV = {1,3,9,8} were determined by expert'PRs.

Calling **Algorithm 1** to solve for the optimal weights as follows:

$$\min \mu$$

$$s.t. \begin{cases} \left| \frac{w_M}{w_R} - 8 \right| \leq \mu \\ \left| \frac{w_M}{w_F} - 8 \right| \leq \mu \\ \left| \frac{w_M}{w_P} - 6 \right| \leq \mu \\ \left| \frac{w_F}{w_R} - 3 \right| \leq \mu \\ \left| \frac{w_P}{w_R} - 5 \right| \leq \mu \\ \sum_j w_j = 1 \\ w_j \geq 0, \text{for all } j \end{cases} \quad (22)$$

The optimal weights $W* = \{w_R *, w_F *, w_M *, w_P *\}$ = {0.064, 0.113, 0.649, 0.174} and $\mu* = 2.26$ are obtained. As $a_{BW} = a_{MR} = 9$ the consistency index for this problem is 4.47 (see Table 3),the consistency ratio is 2.26/4.47 = 0.505 using Eq. 24, which also is a good consistency.

Table 6 User popularity score P

H_{ID}	U_{ID}	···	ULs	S_{ULs}	UFs	S_{UFs}	NUFs	S_{NUFs}	P
A_1	$User_1^{A_1}$	···	45	1.0	6853	0.267	559	0.186	0.292
A_1	$User_2^{A_1}$	···	45	1.0	2503	0.098	2876	0.958	0.454
A_1	$User_3^{A_1}$	···	45	1.0	6841	0.267	135	0.045	0.245
A_1	$User_4^{A_1}$	···	20	0.44	1217	0.047	0.2908	0.896	0.612
A_1	$User_5^{A_1}$	···	18	0.40	1190	0.046	2383	0.894	0.603
...	...	...	...	...	...	...			

Table 7 Standardized data sheet

H_{ID}	···	S_R	S_F	S_M	S_{ULs}	S_{UFs}	S_{NUFs}	S_P
A_1	···	0.0301	0.2376	0.0223	1.00	0.267	0.186	0.292
A_1	···	0.0300	0.1403	0.0787	1.00	0.098	0.958	0.454
A_1	···	0.0224	0.1530	0.2063	1.00	0.267	0.045	0.245
A_1	···	0.0364	0.0409	0.1405	0.44	0.047	0.896	0.612
A_1	···	0.0364	0.0033	0.0564	0.40	0.046	0.894	0.603
...	...	...	...	...	...	...	...	...

Table 8 Opinion leader data sheet

H_{ID}	U_{ID}	···	S_R	S_F	S_M	S_P	U_{score}
A_1	$User_1^{A_1}$	···	0.0301	0.2376	0.0223	0.292	0.126
A_1	$User_2^{A_1}$	···	0.0300	0.1403	0.0787	0.454	0.140
A_1	$User_3^{A_1}$	···	0.0224	0.1530	0.2063	0.245	0.241
A_1	$User_4^{A_1}$	···	0.0364	0.0409	0.1405	0.612	0.142
A_1	$User_5^{A_1}$	···	0.0364	0.0033	0.0564	0.603	0.131
...	...	...	...	...	...	...	...

Before calculating the user influence score, this article first standardizes the users' various parameter values. The standardized formula is:

$$Std_{A_i} = \frac{A_i - A_{min}}{A_{max} - A_{min}} \tag{23}$$

After the RFMP model is trained, the influence of the publisher is predicted by inputting test samples, and then the influence ranking is performed. The standardized data table of R, F, M, and S of the publishers of online reviews publishers are shown in the Table 7.

Similarly, the value of RFMP of rest alternative hotels $\{A_2, A_3, A_4, A_5, A_6, A_7, A_8\}$ can be calculated.

Set the threshold to 0.074 and select OLs. The selected OLs are shown in the Table 8:

The user score value calculation formula is:

$$U_{Score_h} = W_R \times S_{R_h} + W_F \times S_{F_h} + W_M \times S_{M_h} + W_P \times S_{P_h} \tag{24}$$

As a result, 333 OLs were selected based on the RFMP model. In the next subsection, a detailed analysis of opinion leaders comments and preferred opinions.

4.2 Obtaining Weight Vector of Hotel Selection

4.2.1 Obtaining Objective Weight Vector

In the previous subsection, 366 OLs had been screened out and this section focuses on processing and analyzing OLs online reviews.Calling **Algorithm 2** to process the text data of the OLs, after classifying the vocabulary, the Table 8 shows:

By analyzing the comment texts of 333 opinion leaders, a total of 7935 words were counted, of which 6967 unique words were counted. In this paper, 968 invalid words were removed, counting six categories of words, Position, Service, Cleanliness, Comfort, Facility, and Food. Cleanliness appeared 950 times, comfort 955 times, facility 1299 times, food 878 times, location 1545 times, and service 1340 times. The feature words are grouped as shown in the Table 9. The

Table 9 Hotel selection criteria and feature words

Criteria	Feature words
Position	Convenience, Shanghai, position, Nanjing Road, The Bund, Location traffic Walking street, Travel, subway, Very close, Nanjing East Road, Entrance, Pujiang North, wing, museum, Kerry, park, train, Expo, intercontinental, city, location, Number line, Export, Subway, entrance, Located, High-speed rail, Security, check, train station, road, ten minutes outside···
Service	Cost, quality, Good service, Discount, Baggage, Service in place, Price, service, Warm and thoughtful, intimate, initiative, Cheap, attitude, Service, attitude enthusiasm, Five star, Cost performance···
Cleanliness	Surroundings, clean, health, tidy, clean, Wash···
Comfort	Comfortable be quiet Warm enjoy temperature mood Soundproof superior Atmosphere···
Facility	Facility convenient Decoration suite building Room hardware bed bathroom bar Bathtub, Lounge mattress equipment swimming pool the film TV Interior···
Food	Breakfast restaurant buffet Affordable, taste, Self-help, rich, Variety, Meal delivery···

words were plotted into a word cloud as shown in the figure 4. In this paper, the frequencies are used to estimate the OWAs, which could be estimated as { 0.222,0.192,0.136,0 .137,0.186,0.126}.

4.2.2 Obtaining Subjective Weight Vector

Considering the data provided in Table 10 results in model named BWM for this problem, as follows:

Fig. 4 Classification results for sub-criteria

Table 10 Pairwise comparison vector based on PRs

Criteria	Position	Service	Cleanliness	Comfort	Facility	Food
Best Criterion: Cleanliness	4	3	1	2	4	8
Worst Criterion: Food	3	2	8	4	4	1

$\min \lambda$

$$s.t. \begin{cases} \left|\frac{W_3}{W_1} - 4\right| \le \lambda \\ \left|\frac{W_3}{W_2} - 3\right| \le \lambda \\ \left|\frac{W_3}{W_4} - 2\right| \le \lambda \\ \left|\frac{W_3}{W_5} - 4\right| \le \lambda \\ \left|\frac{W_3}{W_6} - 8\right| \le \lambda \\ \left|\frac{W_1}{W_6} - 3\right| \le \lambda \\ \left|\frac{W_2}{W_6} - 2\right| \le \lambda \\ \left|\frac{W_4}{W_6} - 4\right| \le \lambda \\ \left|\frac{W_5}{W_6} - 1\right| \le \lambda \\ \sum_j W_j = 1 \\ W_j \ge 0, \text{for all } j \end{cases} \quad (25)$$

Solving the problem, the optimal weights $W^{SWAs} = \{0.116, 0.133, 0.387, 0.232, 0.083, 0.050\}$ are obtained, for the consistency ratio, as $W^{SWAs}_{Best}/W^{SWAs}_{Worst} = W^{SWAs}_3/W^{SWAs}_6 = 8$ the consistency index for this problem is Table 3 (see Table 3), which is a standard comparison table calculated by experts and the consistency ratio is 0.149, and implies a very good consistency.

4.2.3 Obtaining Combined Weight Vectors(CWAs) of Hotel Selection

suppose that the parameter $\beta = 0.5$, which means that objective weight and subjective weight are equally important to

the traveler, the combined weight vector can be calculated by Eq. 14:

$W = \{ 0.169, 0.163, 0.261, 0.185, 0.135, 0.088\}$

4.3 Ranking of Hotel

It is shown in Table 11 of the initial matrix created, however we must normalize the original data in the analysis before.

We use formula 16 to evaluate indicators with different units together before TOPSIS analysis, normalization process was applied . The decision matrix after standardization is shown in Eq. 26.

$$\begin{bmatrix} 0.151 & 0.143 & 0.144 & 0.140 & 0.148 & 0.140 \\ 0.152 & 0.143 & 0.148 & 0.143 & 0.144 & 0.143 \\ 0.149 & 0.144 & 0.141 & 0.146 & 0.143 & 0.140 \\ 0.154 & 0.143 & 0.140 & 0.140 & 0.141 & 0.149 \\ 0.148 & 0.140 & 0.144 & 0.140 & 0.148 & 0.146 \\ 0.149 & 0.144 & 0.141 & 0.146 & 0.143 & 0.140 \\ 0.149 & 0.146 & 0.149 & 0.148 & 0.148 & 0.144 \\ 0.149 & 0.146 & 0.140 & 0.133 & 0.141 & 0.140 \end{bmatrix} \quad (26)$$

According to Eq. 19, the ideal solution R^+ and negative ideal solution R^- are determined, and the calculation results are shown in follow:

$R^+ = \{0.154, 0.146, 0.149, 0.148, 0.148, 0.149\}$
$R^- = \{0.148, 0.140, 0.140, 0.133, 0.141, 0.140\}$ $\quad (27)$

According to Eq. 20 ideal discrimination S_1^+ and negative ideal discrimination S_1^- are determined, and the calculation results are shown in follow:

$$S_1^+ = \sqrt{\sum_{j=1}^{n} \left[r_{1j}w_j - r_j^+ w_j\right]^2} = \sqrt{\sum_{j=1}^{n} w_j^2 \left[r_{1j} - r_j^+\right]^2}$$

$$= \frac{\sqrt{(2756786\beta^2 - 6697564\beta + 7669687}}{1000000}$$

$$S_1^- = \sqrt{\sum_{j=1}^{n} \left[r_{ij}w_j - r_j^- w_j\right]^2} = \sqrt{\sum_{j=1}^{n} w_j^2 \left[r_{ij} - r_j^-\right]^2}$$

$$= \frac{\sqrt{(2106387\beta^2 - 4065696\beta + 5649467}}{1000000}$$

$$C_1^* = \frac{\sqrt{(2106387\beta^2 - 4065696\beta + 5649467}}{\sqrt{(2106387\beta^2 - 4065696\beta + 5649467} + \sqrt{(2756786\beta^2 - 6697564\beta + 7669687}}$$

$\quad (28)$

Table 11 Comprehensive Hotel Score

HotelID	Position	Service	Cleanliness	Comfort	Facility	Food
A_1	9.4	8.9	9.0	8.7	9.2	8.7
A_2	9.6	8.9	8.7	8.7	8.8	9.3
A_3	9.4	9.0	8.9	9.0	8.8	9.3
A_4	9.5	8.9	9.2	8.9	9.0	8.9
A_5	9.2	8.7	9.0	8.7	9.2	9.1
A_6	8.0	8.7	9.0	8.9	9.0	8.8
A_7	9.2	8.8	8.5	8.9	9.0	8.8
A_8	9.0	8.8	8.7	8.9	8.7	8.7

Table 12 Sensitive analysis with parameters

Score Value	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
A1	0.4619	0.4639	0.4664	0.4693	0.4726	0.4763	0.4804	0.4848	0.4894	0.4941	0.4987
A2	0.7398	0.7317	0.7223	0.7114	0.6991	0.6851	0.6696	0.6526	0.6343	0.6151	0.5954
A3	0.4875	0.4907	0.4936	0.4961	0.4979	0.4987	0.4982	0.4961	0.4919	0.4854	0.4762
A4	0.3174	0.3284	0.341	0.3551	0.3709	0.3882	0.407	0.4269	0.4476	0.4686	0.489
A5	0.4475	0.4473	0.4473	0.4474	0.4478	0.4485	0.4495	0.4509	0.4526	0.4546	0.4568
A6	0.4875	0.4907	0.4936	0.4961	0.4979	0.4987	0.4982	0.4961	0.4919	0.4854	0.4762
A7	0.8891	0.8738	0.8576	0.8404	0.8224	0.8036	0.7839	0.7636	0.7429	0.722	0.7014
A8	0.0902	0.0991	0.1089	0.1195	0.1312	0.1439	0.1576	0.1723	0.188	0.2044	0.221

Similarly, $S_2^+ \cdots S_6^+$, $S_2^- \cdots S_6^-$ and $C_2^* \cdots C_6^*$ can be solved in turn. The results of the calculations are shown in the following table. Here, the weighting of the subjective and objective weights is chosen to be 0.5 each, and the final ranking score results in:

$A_1 = 0.477$; $A_2 = 0.685$; $A_3 = 0.499$; $A_4 = 0.388$; $A_5 = 0.449$; $A_6 = 0.499$; $A_7 = 0.804$; $A_8 = 0.144$

Then the final ranking of the hotel is: $A_7 > A_2 > A_6 > A_3 > A_1 > A_5 > A_4 > A_8$.

4.4 Comparison Experiments

MCDM based on BWM and fuzzy TOPSIS methods has been proposed and widely used by traditional research cite Himanshu Gupta. However, hotel selection using online reviews is a lay decision problem. Filtering out useful online reviews from the large number of online reviews is a problem that needs to be solved. To illustrate the effect of objective weight vector on the final hotel selection results, we conducted a comparative experiment between the BWM and Fuzzy TOPSIS based MCDM proposed by Gupta and the method proposed in this study, and when the proportion of objective weights reached 0.7 and above, the solution ranking showed a significant change, and the larger the value of the weight of objective weights, the larger the solution ranking change,as shown in Fig. 5.

Thus, when the percentage of objective weight is less than 0.9 the final ranking of the hotel by Gupta's method is: $A_7 > A_2 > A_6 > A_3 > A_1 > A_5 > A_4 > A_8$. Meanwhile, the final ranking of the hotel by the proposed method is $A_7 > A_2 > A_1 > A_3 > A_6 > A_4 > A_5 > A_8$.

4.5 Computational Complexity

Since the BMW has fewer comparisons than the AHP, and only linear calculation is needed, the simplicity of weight obtaining is an advantage of this method. the RFMP proposed in this study filters out online reviews of opinion leaders and reduces the amount of online reviews during data processing.

5 Result and Discussion

In the last content of the previous chapter, we equalize the influence utility of the subjective and objective weights and obtain the alternative hotel solutions in the following order: $A_7 > A_2 > A_6 > A_3 > A_1 > A_5 > A_4 > A_8$. Since tourism evaluation is a non-expert system, the customer has not been to the destination, it is said that some OLs are needed to give objective weights to help the decision-maker to make a decision, and the existing literature does not consider the influence of OLs on the decision, so our method is more reasonable. Sensitivity is now widely used by researchers for studies presenting a hybrid model to confirm the validity of the results and eliminate any chance of biasness by
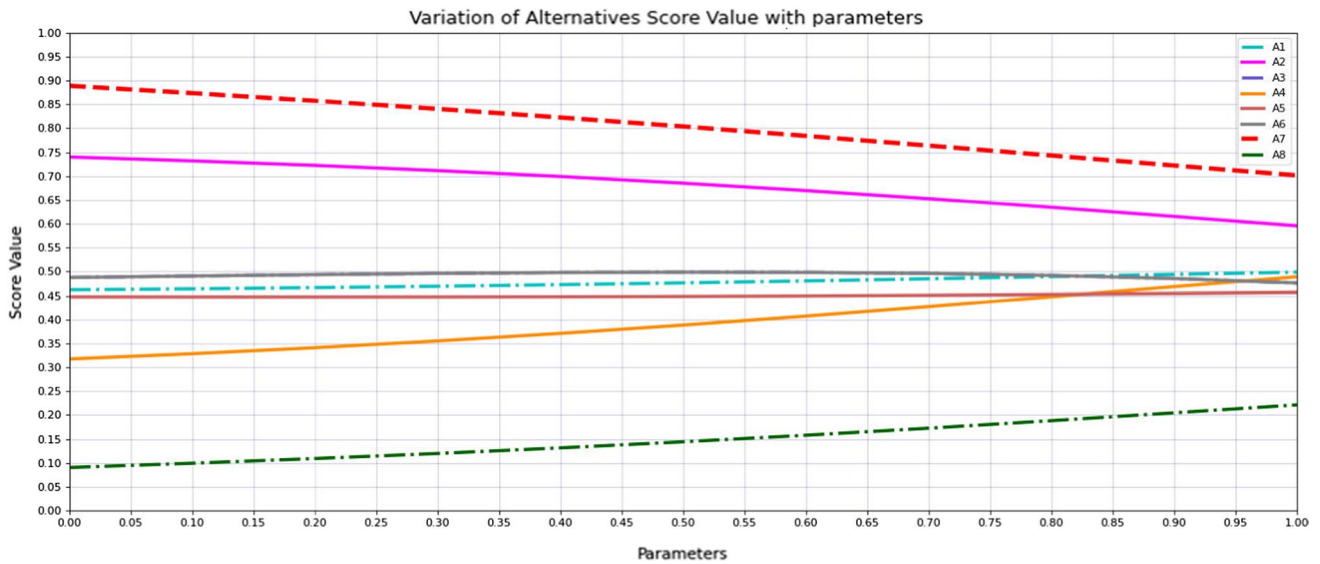
Fig. 5 Robust analysis results of hotel ranking with parameters

subjectivity.To perform sensitivity analysis the criteria obtaining highest weight in subjective and objective is varied from 0.1 to 0.9 and consequently combined weights of all the criteria are varied.

This section focuses on the impact of the objective weights represented by opinion leaders on the overall ranking of hotel alternatives.

Based on the above solution, the trend of the alternatives scores value posting progress with parameters is plotted using mathematical tools as shown in Fig. 5. Table 12 displays the expected results for each hotel under different contexts, whereas Fig. 6 visualizes the robustness analysis results of these experiments.

As the Fig. 5 shows, the objective weight vectors affect the ranking result of the scheme to some extent, but on the whole, the trend of the scheme is stable and variable. This shows that the stability of the online decision support model proposed in this paper considering the subjective and objective weight set is reasonable.

6 Conclusion and Future Work

This article aims to provide a solution for the non-expert decision-making method such as hotel selection. To achieve this goal, a novel hotel selection decision support model based on the online reviews by opinion leaders. In this method, It has the following main advantages with respect to conventional models proposed in the literature.

1. To utilized the helpful information rather than the massive information with fake and useless information on the OTPs, an RFMP model is proposed to extract the

online reviews of opinion leaders, and the Word2vec method is used to extract the criteria from the online review of opinion leaders. Due to the BWM method performs better than the AHP [27] method in terms of consistency and number of comparisons, it is employed to aggregate the RFMP values.

2. Since hotel selection is a typical non-expert decision, the objective weights derived from the online reviews helps the decision maker know which factors are important in places they have not been, and the subjective preferences of decision makers have not been ignored. The method of combining subjective and objective weights proposed in this study considers more comprehensive factors.
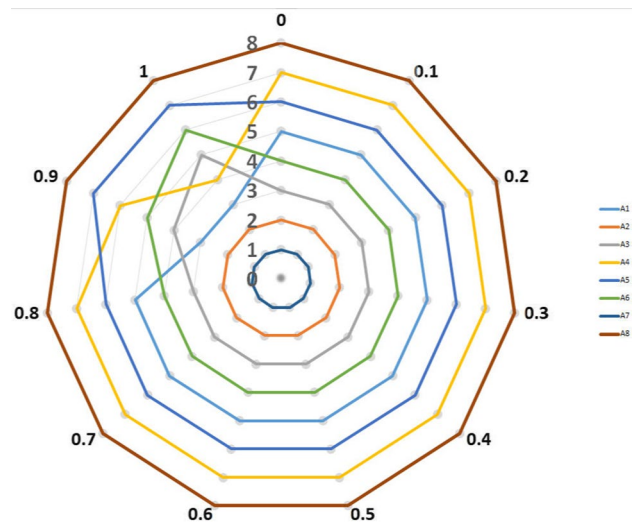

Fig. 6 Robust analysis results of hotel ranking

This study makes the above contributes but also has some limitations, which may serve as avenues for future research. First, Due to the different types of data displayed on OTPs, the data set in this study crawled from only one website. On the premise that data is available, data cross-validation of multiple websites is an effective way to verify this method. Second, As group travel is increasingly popular in people's leisure life, The opinion leaders in group tourism will undoubtedly have an impact on the decision-making results. the group analysis of opinion leaders and group consensus reaching mechanism could be a future research field [39–44]. Besides, the sentiment analysis of online reviews to obtain the objective preference for decision-makers could be considered in the future work.

Declarations

Conflict of interest The authors declare that they have no competing financial interests.

Ethics approval and consent to participate This paper is an original work and has not been published nor has it been submitted simultaneously elsewhere. No data, text, or theories by others who are not list as coauthors are presented.

Consent for publication Not applicable.

References

1. Baek, H., Ahn, J., Choi, Y.: Helpfulness of online consumer reviews: readers objectives and review cues. Int. J. Electron. Commer **17**(2), 99–126 (2012)
2. Kuan, K.Y., Hui, K.L., Prasarnphanich, P., Lai, H.Y.: What makes a review voted? An empirical investigation of review voting in online review systems. J. Assoc. Inf. Syst. **16**(1), 48–71 (2015)
3. Wu, J., Hong, Q., Cao, M., Liu, Y., Fujita, H. A group consensus-based travel destination evaluation method with online reviews. Appl. Intell. (2021) 1-19
4. Park, D.H., Lee, J., Han, I.: The effect of online consumer reviews on consumer purchasing intention: the moderating role of involvement. Int. J. Electron. Commer. **11**(4), 125–148 (2007)
5. Abubakar, A.M., Ilkan, M.: Impact of online WOM on destination trust and intention to travel: a medical tourism perspective. J. Destin Mark. Manag. **5**(3), 192–201 (2018)
6. Zhang, H., Ji, P., Wang, J., Chen, X. A novel decision support model for satisfactory restaurants utilizing social information: a case study of TripAdvisor. com Tour. Manag. (2017) 281–297
7. Huang, G.-H., Chang, C.-T., Bilgihan, A., Okumus, F. Helpful or harmful? A double-edged sword of emoticons in online review helpfulness. Tour. Manag. (2020) 104–135
8. Sun, X., Han, M., Feng, J.: Helpfulness of online reviews: examining review informativeness and classification thresholds by search products and experience products. Decis. Support Syst. **124**, 113099 (2019)
9. Shan, G., Zhou, L., Zhang, D.: From conflicts and confusion to doubts: Examining review inconsistency for fake review detection. Decis. Support Syst. **144**, 113513 (2021)
10. Yu, S., Wang, J., Wang, J., Li, L.: A multi-criteria decision-making model for hotel selection with linguistic distribution assessments. Appl. Soft Comput. **67**, 741–755 (2018)
11. Nie, R., Tian, Z., Wang, J., Chin, K.S.: Hotel selection driven by online textual reviews: Applying a semantic partitioned sentiment dictionary and evidence theory. Int. J. Hosp. Manag. **88**, 102459 (2020)
12. Zhang, C., Xu, Z., Gou, X., Chen, S.: An online reviews-driven method for the prioritization of improvements in hotel services. Tour. Manag. **87**, 104382 (2021)
13. Bi, J.-W., Liu, Y., Fan, Z.-P., Zhang, J.: Exploring asymmetric effects of attribute performance on customer satisfaction in the hotel industry. Tour. Manag. **77**, 104006 (2020)
14. Rezaei, J.: Best-worst multi-criteria decision-making method, *omega,* (2015)
15. Chua, A.Y.K., Banerjee, S.: Understanding review helpfulness as a function of reviewer reputation, review rating, and review depth. J. Assoc. Inf. Sci. Technol. **122**, 354–362 (2015)
16. Diaz, G.O., Ng, V.: Modeling and prediction of online product review helpfulness. Annual Meeting of the Association for Computational Linguistics. Proc. Conf. **122**, 698–708 (2018)
17. Krishnamoorthy, S.: Linguistic features for review helpfulness prediction. Expert Syst. Appl. **42**(7), 3751–3759 (2015)
18. Malik, M.S.I.: Hussain Helpfulness of product reviews as a function of discrete positive and negative emotions. Comput. Hum. Behav. **73**, 290–302 (2017)
19. Saumya, S., Singh, J.P., Baabdullah, A.M., Rana, N.P., Dwivedi, Y.K.: Ranking online consumer reviews. Expert Syst. Appl. **29**, 78–89 (2019)
20. Korfiatis, N., GarcÂÂ, a-Bariocanal, E. Evaluating content quality and helpfulness of online product reviews: The interplay of review helpfulness vs. review content. *Electron. Commer. Res. Appl.,* (2019) 205-217
21. Fan, M., Feng, C., Guo, L., Sun, M.: Product-aware helpfulness prediction of online reviews. In The Web Conference 2019

- Proceedings of the World Wide Web Conference, 122, 2715-2721 (2019)

22. Chen, M.J.: Examining the influence of emotional expressions in online consumer reviews on perceived helpfulness. Inf. Process. Manag. **6**, 57 (2020)

23. Huang, G.H., Chang, C.T., Bilgihan, A.: Helpful or harmful? A double-edged sword of emoticons in online review helpfulness. Tour. Manag. **122**, 81 (2020)

24. Zhou, Y., Yang, S., li, yixiao, chen, Y., Yao, J.: Does the review deserve more helpfulness when its title resembles the content? Locating helpful reviews by text mining. Inf. Process. Manag. 7(2020)102179

25. Hu, Y.H., Yeh, T.W.: Discovering valuable frequent patterns based on RFM analysis without customer identification informationm. Knowl.-Based Syst. **61**, 76–88 (2014)

26. Khajvand, M., Zolfaghar, K., Ashoori, K.: Estimating customer lifetime value based on RFM analysis of customer purchase behavior: Case study. Procedia Comput. Sci. **3**, 57–63 (2011)

27. Jain, L., Katarya, R.: Discover opinion leader in online social network using firefly algorithm. Expert Syst. Appl. **122**, 1–15 (2019)

28. Zhao, Y., Xu, X.: Wang, M, Predicting overall customer satisfaction: big data evidence from hotel online textual reviews. Expert Syst. Appl. **76**, 111–121 (2019)

29. Xu, X., Li, Y.: The antecedents of customer satisfaction and dissatisfaction toward various types of hotels: a text mining approach. Int. J. Hosp. Manag. **55**, 57–69 (2016)

30. Yen, C.A., Tang, C.H.: The effects of hotel criteria performance on electronic word-of-mouth (eWOM) behaviors. Int. J. Hosp. Manag. **76**, 9–18 (2019)

31. Kwok, P.K., Lau, H.Y.K.: Hotel selection using a modified TOPSIS-based decision support algorithm. Decis. Support Syst. **120**(2019), 95–105 (2019)

32. Adal, E.A.: A new integrated decision making approach based on SWARA and OCRA methods for the hotel selection problem. Int. J. Adv. Oper. **8**(2), 140–151 (2016)

33. Yu, S.M.: Wang, A multi-criteria decision-making model for hotel selection with linguistic distribution assessments. Appl. Soft Comput. **67**, 739–753 (2018)

34. Kumar, A., Gupta, H.: Evaluating green performance of the airports using hybrid BWM and VIKOR methodology. Tour. Manag. **76**, 16–24 (2020)

35. Rezaei, J., Kothadiya, O., Tavasszy, L., Kroesen, M.: Quality assessment of airline baggage handling systems using SERVQUAL and BWM. Tour. Manag. **66**, 85–93 (2014)

36. Wu, J., Wang, S., Chiclana, F. Herrera-Viedma, E.: Twofold Personalized Feedback Mechanism for Social Network Consensus by Uninorm Interval Trust Propagation IEEE Transactions on Cybernetics, https://doi.org/10.1109/TCYB.2021.3076420 (2021)

37. Sun, Q., Wu, J., Chiclana, F., Fujita, H. Herrera-Viedma, E.: A dynamic feedback mechanism with attitudinal consensus threshold for minimum adjustment cost in group decision making. IEEE Transactions on Fuzzy Systems, https://doi.org/10.1109/TFUZZ.2021.3057705 (2021)

38. Liu, Y.J., Liang, C.Y., Chiclana, F., Wu, J.: A trust induced recommendation mechanism for reaching consensus in group decision making. Knowl. Based Syst. **119**, 221–231 (2017)

39. Wu, J., Sun, Q., Fujita, H., Chiclana, F.: An attitudinal consensus degree to control the feedback mechanism in group decision making with different adjustment cost. Knowl.-Based Syst. **164**, 265–273 (2019)

40. Cao, M.S., Wu, J., Chiclana, F., Herrera-Viedma, E.: A bidirectional feedback mechanism for balancing group consensus and individual harmony in group decision making. Inf. Fusion **76**, 133–144 (2021)

41. Wu, J., Zhao, Z.W., Sun, Q., Fujita, H.: A maximum self-esteem degree based feedback mechanism for group consensus reaching with the distributed linguistic trust propagation in social network. Inf. Fusion **67**, 80–93 (2021)

42. Dong, Y. C, Zha, Q. B., Zhang,H. J. Kou, G., Fujita, H., Chiclana, F., Herrera-Viedma, E. Consensus reaching in social network group decision making: Research paradigms and challenges,*Knowl.-Based Syst.*, 162, 3-13 (2018)

43. Zhang, Z., Gao, Y., Li, Z.L.: Consensus reaching for social network group decision making by considering leadership and bounded confidence. Knowl.-Based Syst. **204**, 106240 (2020)

44. Dong, Q. X., Zhou, X., Martínez, L. A hybrid group decision making framework for achieving agreed solutions based on stable opinions. Infor Sci, 490, 227-243 (2019)

45. Hwang, C-L, Y Kwangsun. Methods for multiple criteria decision making *Springer*, 1981, 58-191

46. Gupta, H.: Assessing organizations performance on the basis of GHRM practices using BWM and Fuzzy TOPSIS. J. Environ. Manag. **226**, 201–216 (2018)