



HOTEL SELECTION UTILIZING ONLINE REVIEWS: A NOVEL DECISION SUPPORT MODEL BASED ON SENTIMENT ANALYSIS AND DL-VIKOR METHOD

Xia LIANG, Peide LIU*, Zhihao WANG

*School of Management Science and Engineering, Shandong University
of Finance and Economics, Jinan 250014, China*

Received 22 October 2018; accepted 9 June 2019

Abstract. With the considerable development of tourism market, as well as the expansion of the e-commerce platform scale, increasing tourists often prefer to select tourism products such as services or hotels online. Thus, it needs to provide an efficient decision support model for tourists to select tourism products. Online reviews based on the user experience would help tourists improve decision efficiency on tourism products. Therefore, in this study, a quantitative method for hotel selection with online reviews is proposed. First, with respect this problem with online reviews, by analyzing sentiment words in online reviews, tourists' sentiment preferences are transformed into the format of distribution linguistic with respect to sentiment levels. Second, from a theoretical perspective, we proposed a method to determine the ideal solution and nadir solution for distribution linguistic evaluations. Next, based on the frequency of words for evaluating hotel and the distribution linguistic evaluations, the weight vector of the evaluation features is determined. Further, a novel DL-VIKOR method is developed to rank and then to select hotels. Finally, a realistic case from TripAdvisor.com for selecting hotel is used to demonstrate practically and feasibility of the proposed model.

Keywords: decision making, quantitative method, online reviews, hotel selection, distribution linguistic.

JEL Classification: C15, C44, C80.

Introduction

As the time for the era of big data is coming, E-commerce sites have expanded rapidly. Coupled with the rapid growth of tourism market, a new opportunity is given to the e-commerce platform of tourism industry. For most tourists, choices of accommodation are related to the joy of traveling (Sparks & Browning, 2011; Song, 2015; Sabokbar et al., 2016; Chiu & Lin, 2018). Thanks to the e-commerce platform of tourism industry, tourists can

*Corresponding author. E-mails: peide.liu@gmail.com; liangxia@sdufe.edu.cn

access the tourism websites to obtain the accommodations about their travel destinations so as to find the desirable hotel. Nowadays, many websites, such as TripAdvisor.com (<https://www.tripadvisor.com/>), Tuniu.com (<http://www.tuniu.com/>) and Qunar.com (<https://www.qunar.com/>) can provide channels for tourists to choose their tourism service and post online reviews for trip. For example, a tourist wants to visit Macao for few days. Before arriving in Macao, he can browse online reviews about Macao hotels via certain tourism website, such as TripAdvisor.com. Furthermore, some tourism websites provide tourists with an opportunity to view and collect other tourists' reviews. In Figure 1, it shows a list of tourism hotels in Macao from TripAdvisor.com. Since the complexity and subjectivity of online reviews in e-commerce platform, it is not very easy for a tourist to identify the most desirable tourism hotel. Therefore, the topic orientated online reviews to evaluate and select desirable tourism hotel is significant.

At present, selection problem of tourism hotels, which is oriented to online reviews has been drawn attention to many scholars at home and abroad. Li, Ye, and Law (2013) investigate method for determining the satisfaction of customer oriented on hotel, in which the online reviews have been considered. Many consumers usually look through the reviews provided by other consumers before make a decision in travel trip. Thus, Vermeulen and Seegers (2009) studied the impact of consumers' online reviews for hotels, in which the consideration set theory have been applied. Jacob and Guéguen (2015) investigated the effect about the

The image shows a screenshot of the TripAdvisor.com website displaying three hotel listings in Macao. Each listing includes a photograph of the hotel, its name, a 'Sponsored' badge, a 'SAVE' amount, the current price, the original price, the number of reviews, and a 'View Deal' button. The listings are as follows:

- Wynn Macau** (Sponsored): Price CN¥1,773 (Save CN¥250 from CN¥2,023). 616 reviews. Amenities: Pool, Restaurant, Special offer.
- Hotel Okura Macau**: Price CN¥922 (Save CN¥1,351 from CN¥2,273). 1,627 reviews. #1 Best Value of 94 hotels in Macau. Amenities: Free Wifi, Pool. Savings alert: 49% below average rate.
- Mandarin Oriental Macau**: Price CN¥988 (Save CN¥1,687 from CN¥2,675). 807 reviews. #2 Best Value of 94 hotels in Macau. Amenities: Free Parking, Pool, Special offer.

Figure 1. A list of tourism hotels in Macao from TripAdvisor.com

geographic proximity of products for customer's decision. Sparks, So, and Bradley (2016) investigated views of the potential customers with respect to negative oriented of online review and along with the responses of hotels. Amaro, Duarte, and Henriques (2016) researched the influence of the social media on travel plan, and constructed a cluster analysis. Xiang, Du, Ma, and Fan (2017) used text analysis method to investigate three e-commerce websites, such as Yelp, Expedia and TripAdvisor. With respect to a problem of selecting hotels, Li, Law, Vu, and Rong (2013) proposed a fuzzy decision model by using Choquet Integral (CI) aggregation operator. Ye, Li, and Wang (2012) investigated the impact of price on the service quality and value from the perspective of customer perceptions. The study indicates the price has positive effects on perceived service but has negative effects on perceived value. Later on, based on online reviews, a mathematical model was designed to select appropriate hotels on websites (Yu, J. Wang, J. Q. Wang, & Li, 2018). Bi, Liu, Fan, & Zhang (2019) proposed a methodology for conducting importance-performance analysis (IPA) through online reviews. C. B. Zhang, H. Y. Zhang, and Wang (2018) investigated personalized restaurant recommendation problem, considering group correlations and customer preferences, proposed a new recommendation method.

Besides studies on evaluation and selection of tourism hotels, there are some other studies on processing online reviews still worthy studying for reference. These processing technology of online reviews provides technical support for the subsequent decision research with online reviews. With respect to the problem of selecting mobile service, Daekook and Yongtae (2014) analyzed online reviews provided consumers, and proposed a decision model combined text mining and sentiment analysis technology. In order to support online users making decision, Li and Lai (2014) proposed a social appraisal mechanism, which combines social companionship analysis, collective opinion analysis, and consensus decision analysis with the micro-blogsphere. In order to provide a desirable decision model for consumer, Fan, Xi, and Liu (2018) established a model based on online ratings, in which the stochastic dominance theory and PROMETHEE-II method are used. Later on, Liu, Bi, and Fan (2017a) proposed a decision model through online reviews. In this model, by using the sentiment analysis technique, sentiment orientations of the online reviews were identified so as to select products.

Although existing literature has contributed a lot to the decision making problem with online reviews, the methods of processing online reviews are still in an initial stage. The processing for sentiment words of online reviews were used the assignment scoring and adding method. This is equivalent to take the average of an sentiment words, which cannot fully reflect all the information provided by the online reviews. Actually, due to the complexity, fuzziness and uncertainty of the online reviews, the reasonable quantification of the online reviews is a complex problem. Moreover, it is essential to build a reasonable method to solve the hotel selection problem with the complex format of online reviews.

Compared with existing traditional uncertain problem, this selection hotels problem with online review in e-commerce sites is more complex. Therefore, the research aims at building a quantitative method to select the desirable travel hotel(s). On the one hand, since the different evaluation levels are expressed from different online reviews by different consumers, we derive to distributed evaluation easier. On the other hand, in previous studies, distribution

linguistic evaluation is a kind of effective tool for assessment. So as to accomplish the overall objective of selecting the desirable travel hotel(s), inspired by studies for processing online reviews in e-commerce sites, there need following three tasks. First, by adopting feature extraction and sentiment analysis, the online reviews in format of text are transformed into quantified information, which is distribution linguistic information. Then the basic theories for distribution linguistic should be presented, such as operational rule and distance formula. Additionally, the weight of evaluation feature for selecting hotels should be determined. Further, a novel decision model for ranking the alternative hotels should be proposed. Finally, a realistic selection case is illustrate to explain the practicability and effectiveness of the decision support model in this paper.

The remaining sections of this study are organized as follows. Some basic concepts of VIKOR (VlseKriterijumska Optimizacija I Kompromisno Resenje) method and basic theory for distribution linguistic evaluation are introduced in Section 1. Then we developed a novel method to determine the ideal solution and nadir solution. Further, we construct a method of calculating distance between two distribution linguistic variables in Section 1. In Section 2, the formulation of the problem and the resolution process of ranking and selecting hotels are constructed. In Section 3, we built a novel decision support model. The online reviews in format of text are processed. According to feature extraction and sentiment analysis, the weight vector of evaluation feature is determined. And the distribution linguistic evaluations of the hotels are constructed. Then we proposed a distribution linguistic VIKOR (DL-VIKOR) method to rank the travel hotels. A realistic case of selecting travel hotel(s) is used to explain the practicability and effectiveness of the method in Section 4. And conclusions are drawn in the last Section.

1. Preliminaries

1.1. The VIKOR method

In some practical problems with multiple attribute or criteria, because of some conflicting attribute or criteria, it usually has no solution subject to all criteria, simultaneously. The VIKOR method was originally proposed by Opricovic (Opricovic, 1998; Opricovic & Tzeng, 2002), the Serbian name: VlseKriterijumska Optimizacija I Kompromisno Resenje, means multi-criteria optimization and compromise solution. The method is a classical efficient tool for dealing with multiple criteria decision making (MCDM) problems (Alimardani, Hashemkhani Zolfani, Aghdaie, & Tamošaitienė, 2013; Zhang & Xing, 2017) with conflicting criteria. The VIKOR method considers the distances between both the ideal solution and nadir solution, which can sort the alternatives by using utility values and regret degrees.

Suppose a MCDM problem has m alternatives and n criteria. As a matter of convenience, we suppose that all criteria are of the benefit type. Let $M = \{1, 2, 3, \dots, m\}$ and $N = \{1, 2, 3, \dots, n\}$ be the index set of alternatives and criteria, respectively. Let $A = \{A_1, A_2, \dots, A_m\}$ with $m \geq 2$ be a finite set of alternatives, and let $F = \{F_1, F_2, \dots, F_n\}$ with $n \geq 2$ be a finite set of criteria. Let $\Omega = (\omega_1, \omega_2, \dots, \omega_n)^T$, $\omega_j \geq 0$ for $j \in N$, be the criteria weight vector, where ω_j is the weight assigned to criteria F_j . Let $\Xi = [\xi_{ij}]_{m \times n}$ be the decision making matrix, in which ξ_{ij} is the assessment of alternative A_i with respect to criterion F_j .

The classical VIKOR method is described step by step in the following:

Step 1. Normalize the decision matrix $\Xi = [\xi_{ij}]_{m \times n}$ into matrix $K = [\kappa_{ij}]_{m \times n}$, in which κ_{ij} is determined by

$$\kappa_{ij} = \frac{\xi_{ij} - \min_{i \in M} \{\xi_{ij}\}}{\max_{i \in M} \{\xi_{ij}\} - \min_{i \in M} \{\xi_{ij}\}}, \quad i \in M, \quad j \in N. \tag{1}$$

Step 2. Construct the ideal solution $K^+ = \{\kappa_1^+, \kappa_2^+, \dots, \kappa_n^+\}$ and the nadir solution $K^- = \{\kappa_1^-, \kappa_2^-, \dots, \kappa_n^-\}$ by the following Eqs, i.e.,

$$\kappa_j^+ = \max\{\kappa_{ij} \mid i = 1, 2, \dots, m\}, \quad j \in N; \tag{2}$$

$$\kappa_j^- = \min\{\kappa_{ij} \mid i = 1, 2, \dots, m\}, \quad j \in N. \tag{3}$$

Step 3. Compute the maximum group utility value S_i of alternative A_p , i.e.,

$$S_i = \sum_{j=1}^n \omega_j \frac{\kappa_j^+ - \kappa_{ij}}{\kappa_j^+ - \kappa_j^-}, \quad i \in M, \tag{4}$$

where ω_j is the weight of criteria F_j .

Step 4. Calculate the minimum individual regret value R_i of alternative A_p , i.e.,

$$R_i = \max_j \omega_j \frac{\kappa_j^+ - \kappa_{ij}}{\kappa_j^+ - \kappa_j^-}, \quad i \in M. \tag{5}$$

Step 5. Calculate the composite value Q_i of alternative A_p , i.e.,

$$Q_i = \frac{\eta(S_i - \min_i S_i)}{\max_i S_i - \min_i S_i} + \frac{(1 - \eta)(R_i - \min_i R_i)}{\max_i R_i - \min_i R_i}, \quad i \in M. \tag{6}$$

η is the preference weight of maximum group utility value, while $1 - \eta$ is the preference weight of the individual regret degree.

Step 6. Sort the alternatives by maximum group utility value S_p , individual regret degree R_i and composite value Q_i in an ascending order.

The alternative $A^{(1)}$ is the best solution if the following two principles are satisfied:

D1. Acceptable advantage: $Q(A^{(2)}) - Q(A^{(1)}) \geq DQ$, while $A^{(2)}$ is the 2nd position in the ranking result of alternatives by Q_i and $DQ = 1/m - 1$ ($DQ = 0.25$ if $m \leq 4$).

D2. Acceptable stability in decision making. The alternative $A^{(1)}$ must also be the best ranked by S_i or/and R_i .

If one of the principles is not meet, then a set of compromise solutions is proposed, which consists two principles: (a) Alternatives $A^{(1)}$ and $A^{(2)}$ if only the principle D2 is not meet; or (b) Alternatives $\{A^{(1)}, A^{(2)}, \dots, A^{(k)}\}$ if the principle D1 is not meet, $A^{(k)}$ is determined by the relation $Q(A^{(k)}) - Q(A^{(1)}) \approx DQ$ for maximum k (the positions of these alternatives are “in closeness”).

1.2. The distribution linguistic evaluation

The fuzzy evaluations, stochastic evaluations (Liang, Jiang, & Liu, 2018; Sun, Hrušovský, Zhang, & Lang, 2018) and probability linguistic evaluations are useful uncertainty evalua-

tions in practical decision making (Ji, Zhang, & Wang, 2018; Li & Wang, 2018; Luo, Zhang, Wang, & Li, 2019). As a useful format of uncertain evaluation, linguistic evaluations can represent qualitative information more accurate. Herrera (Herrera, Herrera-Viedma, & Verdegay, 1995; Herrera & Herrera-Viedma, 1996) first put forward a totally ordered and finite linguistic term set $S = \{s_0, s_1, \dots, s_T\}$ with odd cardinality. As a matter of convenience, the terms are placed symmetrically around S . For example, a set S with seven terms could be given as: $S = \{s_0 = \text{nothing}, s_1 = \text{very low}, s_2 = \text{low}, s_3 = \text{medium}, s_4 = \text{high}, s_5 = \text{very high}, s_6 = \text{perfect}\}$. Usually, it is required that the linguistic term set S should be satisfied the following characteristics. (a) There is a negative operator, i.e. $Neg(s_i) = s_{g-i}$, where $g + 1$ is the cardinality of S ; (b) The set is ascending ordered, i.e. $s_i \leq s_j \Leftrightarrow i \leq j$. Thus, there exists a maximization operator, i.e. $\max\{s_i, s_j\} = s_i$ when $s_i \geq s_j$, and a minimization operator, i.e. $\min\{s_i, s_j\} = s_i$ when $s_i \leq s_j$.

Definition 1 (Zhang, Dong, & Xu, 2014; Guo, Huynh, & Sriboonchitta, 2017). Suppose $S = \{s_0, s_1, \dots, s_T\}$ denotes a linguistic term set and p_k be the symbolic proportion of s_k , where $s_k \in S, 0 \leq p_k \leq 1, k = 0, 1, \dots, T$ and $\sum_{k=0}^T p_k = 1$. Then an assessment $\psi = \{(s_k, p_k) | k = 0, 1, \dots, T\}$ is called a distribution linguistic evaluation of linguistic term set S .

In order to solve MCDM problem with distribution linguistic evaluations, we may need to determine the ideal distribution linguistic evaluation ψ^+ and nadir distribution linguistic evaluation ψ^- for a sequence of distribution linguistic evaluations. Inspired by Jiang, Liang, and Sun (2015), we proposed three definitions to construct ideal and nadir distribution linguistic evaluations, as well as the distance between two distribution linguistic evaluations.

Definition 2. Let $\psi_1, \psi_2, \dots, \psi_l$ be l distribution linguistic evaluations with $l \geq 2$, where $\psi_i = \{(s_k, p_k^i) | k = 0, 1, \dots, T\}$. Then the ideal distribution linguistic evaluation ψ^+ is defined as

$$\psi^+ = \left\{ \begin{array}{l} (s_0, p_0^+ = 0), \\ \vdots \\ (s_{k-1}, p_{k-1}^+ = 0), \\ (s_k, p_k^+ = 1 - \sum_{v=k+1}^T p_v^+), \\ (s_{k+1}, p_{k+1}^+ = \max_i \{p_{k+1}^i\}), \\ \vdots \\ (s_T, p_T^+ = \max_i \{p_T^i\}) \end{array} \right\}, \tag{7}$$

such that $\sum_{v=0}^T p_v^+ = 1$.

In definition 2, the index k is the smallest index satisfies $p_k^+ \geq 0$, in which k is determined by $p_v^+ = 0$ for all $v(1 \leq v \leq k-1)$, $p_k^+ = 1 - \sum_{v=k+1}^T p_v^+$ and $p_v^+ = \max_i \{p_v^i\}$ for all $v(k+1 \leq v \leq T)$. Thus, we have $\sum_{v=k+1}^T p_v^+ = 1$.

Definition 3. Let $\psi_1, \psi_2, \dots, \psi_l$ be l distribution linguistic evaluations with $l \geq 2$, where $\psi_i = \{(s_k, p_k^i) | k = 0, 1, \dots, T\}$. Then the nadir distribution linguistic evaluation r^- is defined as

$$\Psi^- = \left\{ \begin{array}{l} (s_0, \quad p_0^- = \max_i \{p_0^i\}), \\ \vdots \\ (s_{k-1}, \quad p_{k-1}^- = \max_i \{p_{k-1}^i\}), \\ (s_k, \quad p_k^- = 1 - \sum_{v=1}^{k-1} p_v^-), \\ (s_{k+1}, \quad p_{k+1}^- = 0), \\ \vdots \\ (s_T, \quad p_T^- = 0) \end{array} \right\}, \tag{8}$$

such that $\sum_{v=0}^T p_v^- = 1$.

In definition 3, the index k is the largest index satisfies $p_k^- \geq 0$, in which k is determined by $p_v^- = \max_i \{p_v^i\}$ for all $v(1 \leq v \leq k-1)$, $p_k^- = 1 - \sum_{v=1}^{k-1} p_v^-$ and $p_{k+1}^- = 0$ for all $v(k+1 \leq v \leq T)$. Thus, we have $\sum_{v=1}^{k+1} p_v^- = 1$. Note that the value of k in definition 3 is independent of that of in definition 3.

Definition 4. Suppose ψ_1 and ψ_2 are two distribution linguistic evaluations. Then the normalized distance between two ψ_1 and ψ_2 can be given as:

$$d(\psi_1, \psi_2) = \sqrt{\frac{1}{2(\max_{ij} q_{ij} - \min_{ij} q_{ij})} [P(\psi_1) - P(\psi_2)] Q [P(\psi_1) - P(\psi_2)]^T}, \tag{9}$$

where $P(\psi_1) = [p_0^1, p_1^1, \dots, p_T^1]^T$, and the matrix Q is defined as $Q = [q_{ij}]_{T \times T}$, $q_{ij} = [f^{-1}(s_T)]^2 - [f^{-1}(s_i) - f^{-1}(s_j)]^2$, $f^{-1}(s_i) = i + 1$. Obviously, matrix Q is a symmetric positive definite matrix. It can be seen that the smaller the difference between ψ_1 and ψ_2 is, the larger the value of q_{ij} is.

For distance between two distribution linguistic evaluations, the following properties are provided.

Property 1 (Non-negativity). For any two distribution linguistics ψ_1 and ψ_2 , $0 \leq d(\psi_1, \psi_2) \leq 1$, and $d(\psi_1, \psi_2) = 0 \Leftrightarrow \psi_1 = \psi_2$.

Property 2 (Symmetry). For any two distribution linguistics ψ_1 and ψ_2 , we have $d(\psi_1, \psi_2) = d(\psi_2, \psi_1)$.

Property 3 (Triangle inequality). For any three distribution linguistics ψ_1, ψ_2 and ψ_3 , we have $d(\psi_1, \psi_2) + d(\psi_2, \psi_3) \geq d(\psi_1, \psi_3)$.

Property 4 (Degeneration). If ψ_1 and ψ_2 are two deterministic linguistics, then the distance between ψ_1 and ψ_2 is degraded into the Euclidean distance.

Example 1. The domain $S = \{s_0, s_1, s_2, s_3, s_4\}$ denotes a linguistic term set. Let $\psi_1 = \{(s_0, 0.2), (s_1, 0.2), (s_2, 0.2), (s_3, 0.2), (s_4, 0.2)\}$, $\psi_2 = \{(s_0, 1), (s_1, 0), (s_2, 0), (s_3, 0), (s_4, 0)\}$ and $\psi_3 = \{(s_0, 0), (s_1, 0), (s_2, 0), (s_3, 0), (s_4, 1)\}$ be three distribution linguistic evaluations, which are shown in Table 1.

Table 1. Three distribution linguistic evaluations

| Distribution linguistic evaluations | s_0 | s_1 | s_2 | s_3 | s_4 |
|-------------------------------------|-------|-------|-------|-------|-------|
| ψ_1 | 0.2 | 0.2 | 0.2 | 0.2 | 0.2 |
| ψ_2 | 1 | 0 | 0 | 0 | 0 |
| ψ_3 | 0 | 0 | 0 | 0 | 1 |

Based on the subjective intuition, with respect to distribution linguistic ψ_1 , the full probability is averaged by each linguistic term. And ψ_3 and ψ_2 are two extreme distribution linguistics. Thus, the distance between ψ_3 and ψ_2 should be the maximum. And the distances between ψ_1 and ψ_2 , as well as ψ_1 and ψ_3 should be equal. Based on Definition 4, we have $P(\psi_1)=[0.2,0.2,0.2,0.2,0.2]^T$, $P(\psi_2)=[1,0,0,0,0]^T$ and $P(\psi_3)=[0,0,0,0,1]^T$. Then based on Eq. (9), we can obtain $d(\psi_1, \psi_2)=d(\psi_1, \psi_3)=0.5$ and $d(\psi_2, \psi_3)=1$, which is satisfied with both triangle inequality and subjective intuition. Meanwhile, according to the distance measure provided by Yu et al. (2018), we can calculate the distance between any two distribution linguistic evaluations, which are shown as $d(\psi_1, \psi_2)=0.05$ and $d(\psi_1, \psi_3)=0.85$, and $d(\psi_2, \psi_3)=1$. Unfortunately, we obtain that $d(\psi_1, \psi_2)+d(\psi_1, \psi_3)<d(\psi_2, \psi_3)$, which is not satisfied with triangle inequality. In addition, the result $d(\psi_1, \psi_2) \neq d(\psi_1, \psi_3)$ is not consistent with our subjective intuition. Therefore, the proposed distance measure is a relatively reasonable formula to measure the separation degree of two distribution linguistic evaluations.

2. Decision making problem with online reviews

2.1. Formulation of the problem

Considering a problem is to select the most reasonable travel hotel(s) in e-commerce platform. By a preliminary screening, some acceptable travel hotels are initially confirmed. But, because of the limited knowledge and expertise, the tourist hesitates among the several alternative hotels. In order to rank and select a reasonable hotel, some evaluation features are considered, which are provided from e-commerce platform. In order to support the consumer's decision, a great many of online reviews about the hotels associated with the evaluation features are crawled from the related e-commerce platform. Therefore, the objective in this paper is how to rank and select the travel hotels with the online reviews.

As a matter of convenience, throughout all of this study, the following notations are defined in the problem. Suppose $M = \{1, 2, \dots, m\}$ and $N = \{1, 2, \dots, n\}$ are two sets of subscript. Suppose $A = \{A_1, A_2, \dots, A_m\}$ with $m \geq 2$ is a set of alternative hotels, where A_i expresses the i th alternative hotel, $i \in M$. Generally, by limiting the location and price from the tourist, the alternative hotels set A can be pre-screened. $F = \{F_1, F_2, \dots, F_n\}$ with $n \geq 2$ is a finite set of evaluation features, where F_j denotes the j th evaluation feature, $j \in N$. We suppose the set F can be obtained from e-commerce platform according to their online reviews. Assume $\Omega = (\omega_1, \omega_2, \dots, \omega_n)^T$ is a vector of feature weights, in which ω_j is the weight given to feature F_j with $\omega_j \geq 0$ for $\forall j \in N$, and $\sum_{j=1}^n \omega_j = 1$. The weight vector of evaluation feature can be

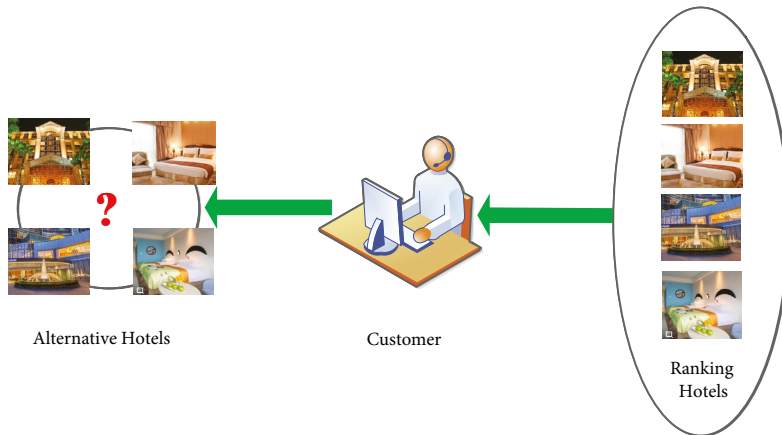


Figure 2. The problem of ranking alternative hotels with online reviews

directly provided by the tourist or indirectly obtained using some classical objective method. $E_i^r = \{e_{ri}^1, e_{ri}^2, \dots, e_{ri}^{Q_i}\}$ is the vector of the online reviews concerning alternative hotels provided by the r th tourist, $r = 1, 2, \dots, R_i$, where e_{ri}^q denotes the q th online review concerning alternative A_i , $q = 1, 2, \dots, Q_i$, Q_i is the number of the online reviews about hotel A_i , $i \in M$.

The problem addressed in this paper is how to rank and select the most reasonable alternative hotel(s) from the alternative set A with the online review information for alternative hotel A_i , the feature weight vector Ω . In Figure 2, the formulation of ranking alternative hotels with online reviews is described concretely.

2.2. Framework and processing for the problem

In order to solve the selection problem above, a process of resolution is set, which is shown in Figure 3. From the Figure 3, we can see that the process of resolution can be divided into two modules, i.e.: (1) crawling the online reviews associated with tourist hotels and identifying sentiment orientations about the online reviews by the sentiment analysis technology; (2) ranking the hotels by using the DL-VIKOR method. The brief description of each module is given below.

Preparatory Phase. In the former module, after determining the initially alternative hotels, as well as the evaluation features associated with hotels provided by the website, the online reviews about the tourist hotels on evaluation features can be crawled by using the crawler software. Then the online reviews are analyzed based on the Stanford parser. Next, by the sentiment dictionary and the online reviews of the hotels, the positive, negative and neutral sentiment dictionaries are constructed. Further, the sentiment orientation of the hotel associated with each evaluation feature in each review is recognized by identifying those of words.

Decision analysis Phase. In the later module, based on the sentiment oriented of reviews associated with different hotels on evaluation features, the assessments of hotels on each evaluation feature can be constructed, which are in format of distribution linguistic evalu-

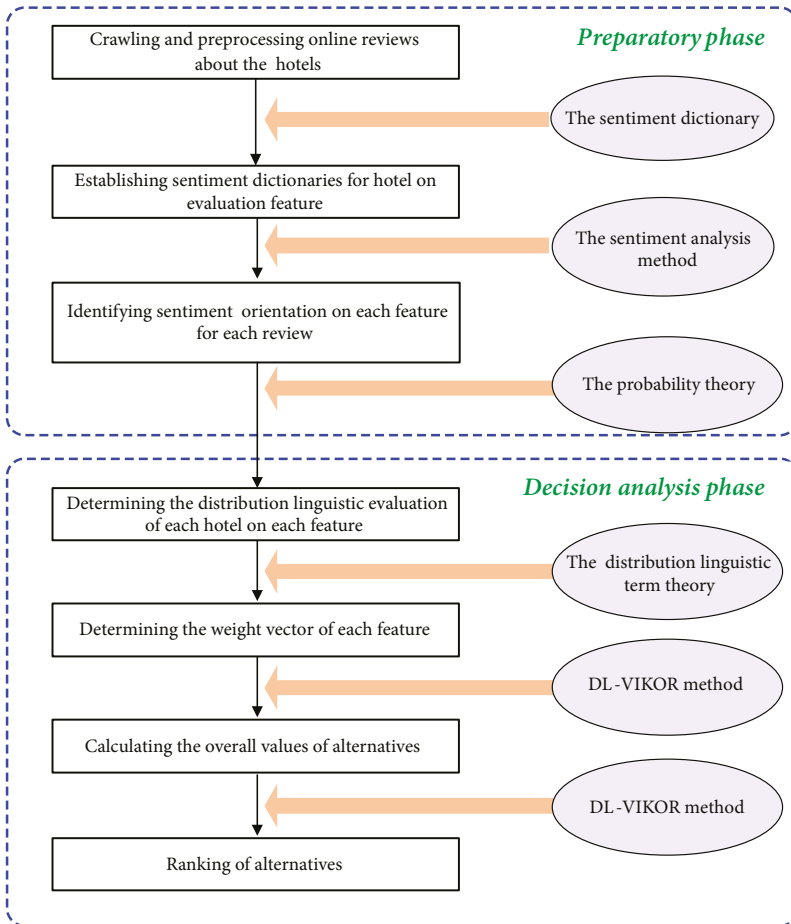


Figure 3. The resolution procedure for the problem of sorting hotels with online reviews

ations. Then, based on numbers of online reviews on each evaluation feature, as well as the evaluation of distribution linguistic evaluation, the weight vector of evaluation features is determined. Further, according to the evaluations of distribution linguistic evaluations for hotels, we proposed a generalized VIKOR method to obtain a ranking of the alternative hotels.

3. The proposed method

Based on Figure 3, the details of the model for ranking and selecting hotels through online reviews are described in the following. In Subsection 3.1, we proposed sentiment analysis technology so as to derive the structural data. Then the method to determine the weight vector of criteria is proposed in Subsection 3.2. Considering the format of distribution linguistic evaluations, a DL-VIKOR method is presented to sort the alternative hotels.

3.1. Determining the structure evaluation information

Based on the sentiment analysis technology, the sentiment orientations on the hotels associated with each evaluation feature in online reviews can be obtained. Thus, this module can be further divided into three parts: (a) Crawling online reviews and preprocessing the related data with respect to the hotels; (b) Establishing sentiment dictionaries for the alternative hotels. Then we identify the positive, negative and neutral sentiment orientation of online reviews. (c) Further, the distribution linguistic evaluation for the hotels on each evaluation feature can be determined. Then the detail of each part is shown as follows.

3.1.1. Crawling and preprocessing online reviews concerning the alternative hotels

In today’s society, some websites support the tourists to public out of their experience feeling online about products (Bi et al., 2019; Liu et al., 2017b, Fan, Che, & Chen, 2017). For example, the TripAdvisor establishes evaluation features of hotels, and encourages the tourists to post their reviews for tourist hotels. Thus, in the following, we suppose that the online review associated with hotels on evaluation features can be crawled from the website TripAdvisor.com. For example, the online reviews provide by any tourist for Okura Macao Hotel from TripAdvisor.com is shown in Figure 4.



Figure 4. An online review for Okura Macao Hotel from TripAdvisor.com

In this subsection, to begin with, we used crawler software, such as Octopus harvester, to derive online reviews. Then we preprocess online reviews according to Stanford parser. Concretely, with respect to all alternative hotels, i.e. $A = \{A_1, A_2, \dots, A_m\}$, the online reviews can be crawled from the related website. Nowadays, some web crawlers have been presented, which can be taken to derive the online reviews. Some concrete introduction on web crawlers may be derived from Wikipedia (https://en.wikipedia.org/wiki/Web_crawler). In fact, an online review posted by a tourist may contain several sentences concerning different features of the hotel. Thus, to identify the customers’ sentiment orientations and sentiment strengths towards the certain feature, the sentences concerning each feature need to be extracted from the online reviews. Specifically, each online review is first divided into several sentences according to the punctuations. As a result, each sentence after divided is considered as a virtual tourist’s online review. Therefore, the online reviews of alternative hotel A_i provided by the r th virtual tourist can be expressed by $E_i^r = \{e_{r_i}^1, e_{r_i}^2, \dots, e_{r_i}^{Q_i}\}$.

3.1.2. Establishing sentiment dictionaries and sentiment analysis

Based on sentiment dictionary HowNet, adding the buzzword, an improved sentiment dictionary is established. Let V^+ , V^- and V' represent the sentiment sets of positive, negative and neutral, which are shown in Table 2.

Table 2. The sentiment words and their attitudes

| The sets of sentiment attitudes | Sentiment words |
|---------------------------------|--|
| V^+ | Satisfy, good, great, high, save, comfortable, excellent, etc. |
| V' | Just so so, neither better nor worse, etc. |
| V^- | Outrageous, shabby, crude, poor, etc. |

In the following, the positive, negative and neutral sentiment orientation of the hotel associated with each evaluation feature in every review can be identified. We suppose the sentiment orientation of a review is mainly embodied in the sentiment words in this statement. If there is only positive or negative sentiment words in this review, then the sentiment orientation of every review is deemed as positive or negative; if there are no obvious sentiment words or both positive and negative sentiment words in this review, then the sentiment orientation of this review is regarded as neutral; if there are negation words in the review, then the sentiment orientation of the review should be reversed.

Let C_{ij}^{rq} represents the sentiment word of for e_{ri}^q on evaluation feature F_j , $P(C_{ij}^{rq})$ represents the oriented of sentiment word C_{ij}^{rq} , i.e.

$$P(C_{ij}^{rq}) = \begin{cases} 1, & \text{if } C_{ij}^{rq} \in V^+; \\ 0, & \text{if } C_{ij}^{rq} \in V'; \quad i=1,2,\dots,m; \quad j=1,2,\dots,n; \quad r=1,2,\dots,R_i; \quad q=1,2,\dots,Q_i. \\ -1, & \text{if } C_{ij}^{rq} \in V^-, \end{cases} \quad (10)$$

The degree words can enhance the expression of sentiment levels. In this paper, based on sentiment dictionary, the degree word is divided into two levels according to the intensity of expression. Let $\text{deg}(C_{ij}^{rq})$ express the degree of adverb which modifies sentiment word C_{ij}^{rq} , where $\text{deg}(C_{ij}^{rq})=1,2,\dots,T/2$ (T is an odd number). The degree words and their scores are shown in Table 3.

If there are negation words with sentiment words, such as “no”, “none” or “really not”, then the sentiment orientation will be reversed. In fact, there are two situations we should consider: (1) If the negation word is negative to another negation word, which expresses positive, then the orientation of sentiment word is constant; (2) If the negation word is

Table 3. The degree words and their scores

| Scores | Degree words |
|--------|--|
| 1 | a bit, a little, slightly, slightly, appreciably..... |
| 2 | Quite, greatly, highly, very, extremely, startlingly, fairly, awfully..... |
| | |

negative to the sentiment word, then its orientation is reversed. Generally, the negation word may appear once or twice. Let N be expressed the number of negation words for sentiment words C_{ij}^{rq} , then $N = 1, 2$.

To sum up, after the treatments of the sentiment orientation, degree adverb and negation word, the sentiment scores Score_{ij}^{rq} of online review e_{ri}^q on feature F_j is

$$\text{Score}_{ij}^{rq} = P(C_{ij}^{rq}) \times \text{deg}(C_{ij}^{rq}) \times [(-1)^{rq}]^N, \quad i = 1, 2, \dots, m; \quad j = 1, 2, \dots, n; \quad q = 1, 2, \dots, Q_i. \quad (11)$$

Therefore, the linguistic term of online review e_i^q on feature F_j is denoted as y_{ij}^q , which is calculated by

$$y_{ij}^{rq} = s_{\text{Score}_{ij}^{rq} + T/2}, \quad i = 1, 2, \dots, m; \quad j = 1, 2, \dots, n; \quad q = 1, 2, \dots, Q_i. \quad (12)$$

3.1.3. Determining the distribution linguistic evaluation of each alternative hotel concerning each feature

We suppose $S = \{H^0 = s_0, H^1 = s_1, \dots, H^T = s_T\}$ is a set of sentiment orientation levels. In order to accurately describe the differences in the level of sentiment orientation, in this study, the evaluations for alternative hotels are format of distribution linguistic evaluations.

Let $\kappa_{ij} = \{(s_k, p_k^{ij}) \mid s_k \in S, 0 \leq p_k^{ij} \leq 1, \sum_{k=1}^T p_k^{ij} = 1, k = 1, 2, \dots, T\}$ be the distribution linguistic evaluation for hotel A_i with respect to evaluation feature F_j . The probability that the hotel A_i on feature F_j expressed as H^k is defined as p_k^{ij} , $i = 1, 2, \dots, m; \quad j = 1, 2, \dots, n; \quad k = 1, 2, \dots, T$.

$$p_k^{ij} = \frac{\sum_{r=1}^{R_i} \sum_{q=1}^{Q_i} \varphi^{rk}(y_{rq}^{ij})}{\sum_{k=1}^T \sum_{r=1}^{R_i} \sum_{q=1}^{Q_i} \varphi^{rk}(y_{rq}^{ij})}, \quad i = 1, 2, \dots, m; \quad j = 1, 2, \dots, n; \quad k = 1, 2, \dots, T, \quad (13)$$

where $\varphi^{rk}(y_{rq}^{ij}) = \begin{cases} 1, & y_{rq}^{ij} = s_k \\ 0, & y_{rq}^{ij} \neq s_k \end{cases}$, $p_k^{ij} \in [0, 1]$ and $\sum_{k=1}^T p_k^{ij} = 1$.

3.2. Determination of evaluation features' weights

The online reviews generalized include two types. One is to express the overall impression of the hotel, such as “One of my favorite hotels in the world”. In this situation, the evaluation of all features is the same and equal to the overall evaluation. The other is to express the evaluation of the hotel with respect to any attribute, such as “Great service”. In order to collect the assessments on all evaluation features for hotels, the desirable weights of evaluation features should be determined. In general, the weights of evaluation features are usually regarded as subjective weights, that is predefined by the decision maker. However, the subjective weights of evaluation features is sometimes completely unknown. In order to determine the desirable features weights that tourists will focus on, we proposed a method based on online reviews.

- 1) The more times a feature appears in online reviews, the higher the weight of the feature is. A statistical method is used to sum up the times of the online reviews of each evaluation feature. Thus, we can obtain the weights of features F_j by the following Equation:

$$\omega_j^1 = \frac{\sum_{i=1}^m \sum_{r=1}^{R_i} \sum_{q=1}^{Q_i} \tau_j(e_{ri}^q)}{\sum_{j=1}^n \sum_{i=1}^m \sum_{r=1}^{R_i} \sum_{q=1}^{Q_i} \tau_j(e_{ri}^q)}, \quad j = 1, 2, \dots, n, \quad (14)$$

in which $\tau_j(e_{ri}^q) = \begin{cases} 1, & y_{ij}^{rq} \neq \phi \\ 0, & y_{ij}^{rq} = \phi \end{cases}$.

2) In the practical process of decision making, the larger the degree of deviation for the evaluation feature, the more important role the evaluation feature plays in the aggregation process of assessments. So we should attach a larger weight toward this evaluation feature. Thus, we can obtain the weight of evaluation feature F_j :

$$\omega_j^2 = \frac{\sum_{i=1}^{m-1} \sum_{l=i+1}^m d(\kappa_{ij}, \kappa_{lj})}{\sum_{j=1}^n \sum_{i=1}^m \sum_{l=i+1}^m d(\kappa_{ij}, \kappa_{lj})}, \quad j = 1, 2, \dots, n, \tag{15}$$

in which $d(\kappa_{ij}, \kappa_{lj})$ is the distance between two distribution linguistic evaluations of hotel A_i and A_l with respect to evaluation feature F_j .

Based on two weight vector for evaluation features, we integrate ω_j^1 and ω_j^2 into combined weight vector of evaluation features, which denoted by

$$\omega_j = \alpha \omega_j^1 + (1 - \alpha) \omega_j^2, \quad j = 1, 2, \dots, n. \tag{16}$$

where α is an important of ω_j^1 for combined weight of evaluation feature F_j , which provided by the decision maker.

3.3. Ranking the hotels based on the distribution linguistic information

After obtaining the evaluations in format of distribution linguistic for alternative hotels and the evaluation features, we propose a novel DL-VIKOR method to ranking the alternative hotels.

First, we need to determine the ideal distribution linguistic solution $K^+ = \{\kappa_1^+, \kappa_2^+, \dots, \kappa_n^+\}$ and nadir distribution linguistic solution $K^- = \{\kappa_1^-, \kappa_2^-, \dots, \kappa_n^-\}$ for all of the alternative hotels by the following Eqs, i.e.,

$$\kappa_j^+ = \left\{ \begin{array}{l} (s_0, \quad p_0^{+j} = 0), \\ \vdots \\ (s_{k-1}, \quad p_{k-1}^{+j} = 0), \\ (s_k, \quad p_k^{+j} = 1 - \sum_{v=k+1}^T p_{jv}^+), \\ (s_{k+1}, \quad p_{k+1}^{+j} = \max_i \{p_{k+1}^{ij}\}), \\ \vdots \\ (s_T, \quad p_T^{+j} = \max_i \{p_T^{ij}\}) \end{array} \right\}, \quad j = 1, 2, \dots, n; \tag{17}$$

$$\kappa_j^- = \left\{ \begin{array}{l} (s_0, \quad p_0^{-j} = \max_i \{p_0^{ij}\}), \\ \vdots \\ (s_{k-1}, \quad p_{k-1}^{-j} = \max_i \{p_{k-1}^{ij}\}), \\ (s_k, \quad p_k^{-j} = 1 - \sum_{v=1}^{k-1} p_v^{-j}), \\ (s_{k+1}, \quad p_{k+1}^{-j} = 0), \\ \vdots \\ (s_T, \quad p_T^{-j} = 0) \end{array} \right\}, \quad j = 1, 2, \dots, n. \tag{18}$$

Then, based on the normalized distance between two distribution linguistic evaluations, the maximum group utility value DLS_i of alternative A_i is calculated, i.e.

$$DLS_i = \sum_{j=1}^n \omega_j \frac{d(\kappa_j^+, \kappa_{ij})}{d(\kappa_j^+, \kappa_j^-)}, \quad i \in M, \tag{19}$$

in which ω_j is the weight of evaluation feature F_j .

Further, the distribution linguistic individual regret degree DLR_i of alternative hotel A_i can be computed by the following Eq.:

$$DLR_i = \max_j \left\{ \omega_j \frac{d(\kappa_j^+, \kappa_{ij})}{d(\kappa_j^+, \kappa_j^-)} \right\}, \quad i \in M. \tag{20}$$

Based on the classical VIKOR method, the compromise solution has a characteristic, i.e. a maximum group utility while a minimum individual regret degree, simultaneously. Therefore, the compromise degree DLQ_i of the alternative hotel A_i can be derived as follows:

$$DLQ_i = \frac{\eta(DLS_i - \min_i DLS_i)}{\max_i DLS_i - \min_i DLS_i} + \frac{(1 - \eta)(DLR_i - \min_i DLR_i)}{\max_i DLR_i - \min_i DLR_i}, \quad i \in M, \tag{21}$$

where $0 \leq \eta \leq 1$ is the preference weight of maximum group utility values, while $1 - \eta$ is the weight of the individual regret degree.

According to the ascending order of DLS_i , DLR_i and DLQ_i , three ranking results for alternative hotels are obtained. The compromise solution is the hotel $A^{(1)}$ (the alternative hotel with the first position in the ranking result by DLQ_i) if the following two principles are satisfied.

D1. $DLQ(A^{(2)}) - DLQ(A^{(1)}) \geq DQ$, whereas $A^{(2)}$ with 2nd position in the ranking result by DLQ_i and $DQ = 1/m - 1$.

D2. The alternative hotel $A^{(1)}$ must also be the best ranked by DLS_i or/and DLR_i .

If one of the principles is not meet, then a set of compromise solutions is determined, which consists two principles: (a) Alternatives hotel $A^{(1)}$ and $A^{(2)}$ if only the principle D2 is not satisfied, or (b) Alternatives $\{A^{(1)}, A^{(2)}, \dots, A^{(m)}\}$ if the principle D1 is not satisfied, $A^{(k)}$ is determined by the relation $DLQ(A^{(k)}) - DLQ(A^{(1)}) \approx DQ$ for maximum k .

According to the above analysis, the procedures of the proposed method for solving the selection problem with online reviews can be summarized as follows:

Step 1. Crawl online review $E_i = \{e_i^1, e_i^2, \dots, e_i^{Q_i}\}$ from the related travel website by a web crawler software. Preprocess the obtained online reviews based on Stanford parser, in which the notional words of each review is denoted by e_i^k .

Step 2. Construct positive, negative and neutral sentiment dictionaries V^+ , V^- and V' concerning feature F_j based on Tables 2–3.

Step 3. Determine distribution linguistic evaluations κ_{ij} of alternative hotel A_i on evaluation feature F_j using Eqs. (10)–(13).

Step 4. Compute the weights of evaluation feature F_j using Eqs (14)–(16).

Step 5. Establish the ideal distribution linguistic solution and nadir distribution linguistic solution $K^+ = \{\kappa_1^+, \kappa_2^+, \dots, \kappa_n^+\}$ and $K^- = \{\kappa_1^-, \kappa_2^-, \dots, \kappa_n^-\}$ using Eqs (17) and (18).

Step 6. Compute the distribution linguistic group utility value DLS_p , distribution linguistic individual regret degree DLR_i and the compromise value DLQ_i of alternative A_i by using Eqs (19)–(21).

Step 7. Determine the compromise solution(s) based on DLS_p , DLR_i and DLQ_i .

4. A case study

A case of selecting the desirable hotel with online reviews in TripAdvisor.com is used to illustrate the proposed model.

A tourist wants to go on a trip to Macao in holiday. However, since the tourist has limited of advance knowledge for foreign hotels, it is difficult to select an optimal hotel in time. Thus, it is need to make decision through a tourism services platform. TripAdvisor.com (<https://www.tripadvisor.com/>) is one of the world's leading tourism websites. TripAdvisor.com compares prices from more than 200 booking sites to help tourists find the lowest price on the right hotel, restaurants and vacation rental. TripAdvisor.com includes with over 200 million online reviews which provide by tourists from all regions. As a popular tourism website, "Latest reviews, lowest prices" is the objective of the TripAdvisor.com. It provides many kinds of reviews about travel text, such as vacation rental, restaurants and hotels, etc.

By limiting the location and price of hotels, five alternative hotels in Macao were initially screened by a tourist in TripAdvisor.com. The tourist needs to select a desirable hotel from the following five hotels.

A_1 : Galaxy Hotel

A_2 : Conrad Macao Cotai Central

A_3 : Hotel Okura Macao

A_4 : Grand Hyatt Macao

A_5 : Sheraton Grand Macao Hotel

Based on the information in this tourism services platform, five features associated with alternative hotels are considered, i.e. Location (F_1), Service (F_2), Sleep Quality (F_3), Cleanliness (F_4) and Rooms (F_5). The weight vector of the five evaluation features is waited for confirmation, which are denoted as $\Omega = (\omega_1, \omega_2, \omega_3, \omega_4, \omega_5)^T$. The proposed method for determining weight vector in details provided in Section 4 is applied to ranking these five alternative hotels. The calculation processes and discussion are expressed as below.

4.1. Preprocessing data

First, based on Octopus harvester, we can crawl the reviews for these five hotels from the website of TripAdvisor.com. Due to the original information obtained from online reviews websites is complex and Textual data, we must further transform the original data into quantifiable evaluations. Then, we can obtain the sentiment words of all reviews by preprocessing the online reviews using Stanford parser. For each online review, the orientations of tourists for hotels are positive sentiment, negative sentiment and neutral sentiment. Thus, we establish positive, negative and neutral sentiment dictionaries V^+ , V^- and V' based on

Tables 2–3. According to the sentiment dictionaries, then the distribution linguistic evaluations κ_{ij} of alternative hotel A_i concerning feature F_j are format of distributed information by using Eqs (10)–(13). In this practical example, we suppose $S = \{s_0, s_1, s_2, s_3, s_4\}$ is a set of sentiment orientation levels. Thus, we construct the distribution linguistic decision matrix of five alternative hotels on five features, which is shown in the following Table 4.

Table 4. The distribution linguistic decision matrix $K = [\kappa_{ij}]_{5 \times 5}$

| | F_1 | F_2 | F_3 |
|-------|--|--|---|
| A_1 | $\left\{ \begin{array}{l} (s_2, 0.24), (s_3, 0.27), \\ (s_4, 0.49) \end{array} \right\}$ | $\left\{ \begin{array}{l} (s_2, 0.29), (s_3, 0.36), \\ (s_4, 0.35) \end{array} \right\}$ | $\left\{ \begin{array}{l} (s_2, 0.02), (s_3, 0.44), \\ (s_4, 0.54) \end{array} \right\}$ |
| A_2 | $\left\{ \begin{array}{l} (s_2, 0.15), (s_3, 0.25), \\ (s_4, 0.6) \end{array} \right\}$ | $\left\{ \begin{array}{l} (s_2, 0.05), (s_3, 0.55), \\ (s_4, 0.40) \end{array} \right\}$ | $\left\{ \begin{array}{l} (s_2, 0.24), (s_3, 0.27), \\ (s_4, 0.49) \end{array} \right\}$ |
| A_3 | $\left\{ \begin{array}{l} (s_2, 0.02), (s_3, 0.33), \\ (s_4, 0.65) \end{array} \right\}$ | $\left\{ \begin{array}{l} (s_0, 0.07), (s_1, 0.3), (s_2, 0.12), \\ (s_3, 0.18), (s_4, 0.33) \end{array} \right\}$ | $\left\{ \begin{array}{l} (s_0, 0.17), (s_2, 0.45), \\ (s_3, 0.24), (s_4, 0.14) \end{array} \right\}$ |
| A_4 | $\left\{ \begin{array}{l} (s_2, 0.39), (s_3, 0.61) \end{array} \right\}$ | $\left\{ \begin{array}{l} (s_1, 0.25), (s_2, 0.02), \\ (s_3, 0.43), (s_3, 0.3) \end{array} \right\}$ | $\left\{ \begin{array}{l} (s_0, 0.2), (s_3, 0.52), \\ (s_4, 0.28) \end{array} \right\}$ |
| A_5 | $\left\{ \begin{array}{l} (s_2, 0.09), (s_3, 0.09), \\ (s_4, 0.82), \end{array} \right\}$ | $\left\{ \begin{array}{l} (s_0, 0.07), (s_1, 0.31), (s_2, 0.24), \\ (s_3, 0.31), (s_4, 0.07) \end{array} \right\}$ | $\left\{ \begin{array}{l} (s_2, 0.43), (s_3, 0.28), \\ (s_4, 0.29) \end{array} \right\}$ |
| | F_4 | F_5 | |
| A_1 | $\left\{ \begin{array}{l} (s_1, 0.14), (s_2, 0.54), \\ (s_3, 0.32) \end{array} \right\}$ | $\left\{ \begin{array}{l} (s_0, 0.29), (s_1, 0.34), \\ (s_3, 0.37) \end{array} \right\}$ | |
| A_2 | $\left\{ \begin{array}{l} (s_2, 0.37), (s_3, 0.33), \\ (s_4, 0.3) \end{array} \right\}$ | $\left\{ \begin{array}{l} (s_1, 0.13), (s_3, 0.39), \\ (s_4, 0.48) \end{array} \right\}$ | |
| A_3 | $\left\{ \begin{array}{l} (s_2, 0.34), (s_3, 0.2), \\ (s_4, 0.46) \end{array} \right\}$ | $\left\{ \begin{array}{l} (s_0, 0.12), (s_1, 0.32), \\ (s_2, 0.28), (s_4, 0.28) \end{array} \right\}$ | |
| A_4 | $\left\{ \begin{array}{l} (s_2, 0.34), (s_3, 0.32), \\ (s_4, 0.34) \end{array} \right\}$ | $\left\{ \begin{array}{l} (s_0, 0.04), (s_1, 0.13), (s_2, 0.05), \\ (s_3, 0.35), (s_4, 0.43) \end{array} \right\}$ | |
| A_5 | $\left\{ \begin{array}{l} (s_0, 0.16), (s_1, 0.33), \\ (s_2, 0.21), (s_3, 0.3) \end{array} \right\}$ | $\left\{ \begin{array}{l} (s_2, 0.19), (s_3, 0.36), \\ (s_4, 0.45) \end{array} \right\}$ | |

4.2. Methodology and results

After processing the evaluations with online reviews, the structure data in format of distribution linguistic evaluations are obtained. Then based on times of online reviews and the distribution linguistic evaluations, the weight vector of evaluations can be determined. By times of features with online reviews and Eq. (14), the weight vector of features can be calculated as $\Omega^1 = (0.2, 0.23, 0.22, 0.17, 0.18)^T$. On the other hand, by the distribution linguistic evaluations for all alternative hotels on each feature, the objective weight vector of evaluation features can be calculated as $\Omega^2 = (0.32, 0.09, 0.2, 0.24, 0.15)^T$. Combined Ω^1 and Ω^2 , then we calculate the overall weight vector of evaluation features, which denoted as $\Omega = (0.26, 0.16, 0.21, 0.21, 0.16)^T$. Here $\alpha = 0.5$ is provided by the tourist.

Further, by the proposed DL-VIKOR method, the ranking result for the alternative hotels can be obtained. Based on the decision making matrix for alternative hotels, the ideal distribution linguistic solution and nadir distribution linguistic solution $K^+ = \{\kappa_1^+, \kappa_2^+, \kappa_3^+, \kappa_4^+, \kappa_5^+\}$ and $K^- = \{\kappa_1^-, \kappa_2^-, \kappa_3^-, \kappa_4^-, \kappa_5^-\}$ are determined by using Eqs (17) and (18). By the distribution linguistic group utility values, distribution linguistic individual regret degrees and the compromise values of five hotels are proposed by Eqs (19)–(21).

Concretely, the distribution linguistic group utility values of five alternative hotels are computed based on Eq. (19):

$$DLS_1 = 0.322, DLS_2 = 0.188, DLS_3 = 0.352, DLS_4 = 0.234, DLS_5 = 0.205.$$

The distribution linguistic individual regret degrees of these five tourist hotels are calculated by using Eq. (20):

$$DLR_1 = 0.129, DLR_2 = 0.051, DLR_3 = 0.145, DLR_4 = 0.128, DLR_5 = 0.116.$$

Using Eq. (21), the distribution linguistic compromise measures of these five tourist hotels are calculated as follows:

$$DLQ_1 = 0.824, DLQ_2 = 0, DLQ_3 = 1, DLQ_4 = 0.554, DLQ_5 = 0.398.$$

Based on the decreasing orders of DLS_i, DLR_i, DLQ_i , three kinds of ranking results for these five alternative hotels are derived as:

$$A_2 \succ_{DLS} A_5 \succ_{DLS} A_4 \succ_{DLS} A_1 \succ_{DLS} A_3; A_2 \succ_{DLR} A_5 \succ_{DLR} A_4 \succ_{DLR} A_1 \succ_{DLR} A_3;$$

$$A_2 \succ_{DLQ} A_5 \succ_{DLQ} A_4 \succ_{DLQ} A_1 \succ_{DLQ} A_3.$$

Obviously, the compromise solution in this decision making process for five alternative hotel is $A_2 \succ_{DLQ} A_5 \succ_{DLQ} A_4 \succ_{DLQ} A_1 \succ_{DLQ} A_3$. Because it meets the following two principles:

(a) principle D1, i.e., $DLQ(A_5) - DLQ(A_2) = 0.398 \geq 0.25$; (b) principle D2, i.e., hotel A_2 is also the hotel with the first position in the ranking results derived by DLR_i . Therefore, the best hotel for this trip is A_2 , i.e. Conrad Macao Cotai Central.

4.3. Discussion and comparison

In the subsection, we design a sensitive analysis and a comparison of our proposal with the methods proposed by Daekook and Yongtae (2014) and Yu et al. (2018) so as to represent the advantages of the model in this study.

Let the parameter η be assigned different numbers, such as $\eta = 0$, $\eta = 0.5$ and $\eta = 1$. Based on the proposed method, different parameters η yield different results, which can be shown in the following Table 5.

Table 5. The different ranking results with different parameters η

| Hotel | DLS_i | DLR_i | $DLQ_i(\eta = 0)$ | ranking | $DLQ_i(\eta = 0.5)$ | ranking | $DLQ_i(\eta = 1)$ | ranking |
|-------|---------|---------|-------------------|---------|---------------------|---------|-------------------|---------|
| A_1 | 0.322 | 0.129 | 0.832 | 4 | 0.824 | 4 | 0.816 | 2 |
| A_2 | 0.188 | 0.051 | 0 | 1 | 0 | 1 | 0 | 1 |
| A_3 | 0.352 | 0.145 | 1 | 5 | 1 | 5 | 1 | 5 |
| A_4 | 0.234 | 0.128 | 0.827 | 3 | 0.554 | 3 | 0.282 | 3 |
| A_5 | 0.205 | 0.116 | 0.696 | 2 | 0.398 | 2 | 0.101 | 2 |

In Table 5, the ranking results under the different parameter η . If $\eta = 0$, the ranking result is $A_2 \succ_{DLQ} A_5 \succ_{DLQ} A_4 \succ_{DLQ} A_1 \succ_{DLQ} A_3$. In the case of $\eta = 0.5$, the ranking result is $A_2 \succ_{DLQ} A_5 \succ_{DLQ} A_4 \succ_{DLQ} A_1 \succ_{DLQ} A_3$. Finally, in the case of $\eta = 1$, the ranking result is $A_2 \succ_{DLQ} A_5 \succ_{DLQ} A_4 \succ_{DLQ} A_1 \succ_{DLQ} A_3$. However, based on the rules of VIKOR method, we have $A_2 \succ_{DLQ} A_5 \succ_{DLQ} A_4 \succ_{DLQ} A_1 \succ_{DLQ} A_3$ when $\eta = 0$ or $\eta = 0.5$ while $\{A_2, A_5, A_4\} \succ_{DLQ} A_1 \succ_{DLQ} A_3$ when $\eta = 1$. Thus, the weight of the maximum group utility reflects decision maker’s preference, which makes a direct contribution to the similar ranking results.

In the following, a comparison of the results obtained using different three methods to solve the case is in Table 6 ($\eta = 0.5$).

Table 6. Comparison of ranking results by using three methods

| Methods | Ranking results |
|--|---|
| The DL-VIKOR method in this study | $A_2 \succ A_5 \succ A_4 \succ A_1 \succ A_3$ |
| The method presented by Daekook and Yongtae (2014) | $\{A_2, A_5, A_4\} \succ A_1 \succ A_3$ |
| The method presented by Yu et al. (2018) | $\{A_1, A_2\} \succ A_5 \succ A_4 \succ A_3$ |

According to Table 6, the major trend of ranking results by using these three methods is basically in line. To our delight, the alternative hotel A_2 , which is the best alternative by our proposed method, has a top scores of 5 points in TripAdvisor.com. And the ranking of other hotels by the proposed method has the similar ranking result in top 15 hotels of Macao in TripAdvisor.com.

However, it can be seen that a slightly difference is obtained using these three methods. Since the two principles of the VIKOR method, although the ranking of group utility values and compromise measures are similar based on these three methods, the ranking of five alternative hotels is different. The ranking of alternative hotels A_2 , A_4 and A_5 cannot be distinguished using the methods presented by Daekook and Yongtae (2014). The main reason for this is that the evaluations which are constructed by online reviews were a little coarse. And some evaluation information about different levels is partially ignored and cancelled. For ex-

ample, in this proposed method, the evaluations of A_5 concerning feature F_2 , F_3 and F_4 are in format of $\{(s_0, 0.07), (s_1, 0.31), (s_2, 0.24), (s_3, 0.31), (s_4, 0.07)\}$, $\{(s_2, 0.43), (s_3, 0.28), (s_4, 0.29)\}$ and $\{(s_0, 0.16), (s_1, 0.33), (s_2, 0.21), (s_3, 0.3)\}$ respectively. These evaluations all have a similar character, which is both the probabilities of low linguistic term and high linguistic term are large. Since the compensatory of evaluations among different linguistic terms, it partially leads to a loss of evaluations information for alternative hotels A_3 , A_4 and A_5 by using the method provided by Daekook and Yongtae (2014).

On the other hand, the similar problem arises in the method provided by Yu et al. (2018). Although the full evaluations of alternative hotels are preserved so as to avoid information loss, the ranking result is also different from that of other method. The main reason is that the distance measure in this generalized VIKOR method is different from that of the proposed method. This was specifically contrasted for these two methods in the Subsection 1.2. Therefore, through the comparison values of results, the proposed DL-VIKOR method can avoid information loss when handling the online reviews. Furthermore, more accurate overall compromise values of alternative hotels are derived, which leads to a clear ranking result of the alternative hotels. What is worth mentioning is that the method provided by Yu et al. (2018) can only solve the problem with linguistic or crisp evaluations from each customer. In fact, some customers usually give complex text online reviews instead of crisp evaluations to express their feelings about consumption. In this study, the proposed method can evaluate and select products with online reviews effectively.

Conclusions

The problem of selecting tourism products with online reviews has extensive practical application background. With respect to evaluating and selecting travel hotels, a decision support model is provide in this study, based on data processing method and distribution linguistic VIKOR (DL-VIKOR) method. In the proposed method, by sentiment analysis, we processed the text data into distribution linguistic evaluations. With respect of assessment in format of distribution linguistic, the approach to determining weight vector of evaluation feature is presented. Then the method of determining ideal solution and nadir solution, and the distance between two distribution linguistic evaluations are proposed. Further, the ranking method for alternative hotels is presented based on DL-VIKOR.

Compared to previous works, this study has some characteristics. From the realistic perspective, the problem of selecting tourism service product with online reviews is investigated so as to propose a novel decision support model. Moreover, from a theoretical point of view, in order to take advantage of textual data as possible, we processed the text data into distribution linguistic evaluation. This link can avoid information loss or distortion compared to the previous models. Then some basic theories about distribution linguistic are introduced, such as ideal solution and nadir solution, distance formula and the DL-VIKOR method, in favor of the selection.

For future researches, it is worth saying that the attainment for evaluation feature associated with alternatives could be further investigated from online reviews. Further, the evaluation or selection problem with online reviews from different regions could be investigated,

such as e-commerce platform, micro-blog and the public service platforms. Additionally, so as to promote the choice of decision makers in the real problems, it is essential to apply the proposed method to solve some real decision problems (Guan, Zhao, & Du, 2017; Sun, Li, Liang, & Zhang, 2019).

Acknowledgements

This work was jointly supported by the National Natural Science Foundation of China (NSFC, Project Nos 71801142, 71771140, 71701111), the Humanities and Social Sciences Research Project of Ministry of Education of China (No. 17YJC630077, 17YJA630065), A Project of Shandong Province Higher Educational Science and Technology Program (No. J17KA189).

References

- Alimardani, M., Hashemkhani Zolfani, S., Aghdaie, M. H., & Tamošaitienė, J. (2013). A novel hybrid SWARA and VIKOR methodology for supplier selection in an agile environment. *Technological and Economic Development of Economy*, 19(3), 533-548. <https://doi.org/10.3846/20294913.2013.814606>
- Amaro, S., Duarte, P., & Henriques, C. (2016). Travelers' use of social media: A clustering approach. *Annals of Tourism Research*, 59, 1-15. <https://doi.org/10.1016/j.annals.2016.03.007>
- Bi, J. W., Liu, Y., Fan, Z. P., & Zhang, J. (2019). Wisdom of crowds: Conducting importance-performance analysis (IPA) through online reviews. *Tourism Management*, 70, 460-478. <https://doi.org/10.1016/j.tourman.2018.09.010>
- Chiu, S. H., & Lin, T. Y. (2018). Performance evaluation of Taiwanese international tourist hotels: Evidence from a modified NDEA model with ICA technique. *Technological and Economic Development of Economy*, 24(4), 1560-1580. <https://doi.org/10.3846/tede.2018.3116>
- Daekook, K., Yongtae, P. (2014). Review-based measurement of customer satisfaction in mobile service: Sentiment analysis and VIKOR approach. *Expert Systems with Applications*, 41(4), 1041-1050. <https://doi.org/10.1016/j.eswa.2013.07.101>
- Fan, Z. P., Che, Y. J., & Chen, Z. Y. (2017). Product sales forecasting using online reviews and historical sales data: A method combining the Bass model and sentiment analysis. *Journal of Business Research*, 74, 90-100. <https://doi.org/10.1016/j.jbusres.2017.01.010>
- Fan, Z. P., Xi, Y., & Liu, Y. (2018). Supporting consumer's purchase decision: A method for ranking products based on online multi-attribute product ratings. *Soft Computing*, 22(16), 5247-5261. <https://doi.org/10.1007/s00500-017-2961-4>
- Guan, H. J., Zhao, A. W., & Du, J. G. (2017). *Enterprise green technology innovation behavior*. Beijing, China: Economic Science Press, 4.
- Guo, W. T., Huynh, V. N., & Sriboonchitta, S. (2017). A proportional linguistic distribution based model for multiple attribute decision making under linguistic uncertainty. *Annals of Operations Research*, 256(2), 305-328. <https://doi.org/10.1007/s10479-016-2356-4>
- Herrera, F., Herrera-Viedma, E., & Verdegay, L. (1995). A sequential selection process in group decision making with linguistic assessment approach. *Information Sciences*, 85, 223-239. [https://doi.org/10.1016/0020-0255\(95\)00025-K](https://doi.org/10.1016/0020-0255(95)00025-K)
- Herrera, F., & Herrera-Viedma, E. (1996). A model of consensus in group decision making under linguistic assessments. *Fuzzy Sets and System*, 78, 73-87. [https://doi.org/10.1016/0165-0114\(95\)00107-7](https://doi.org/10.1016/0165-0114(95)00107-7)
- Jacob, C., & Guéguen, N. (2015). Does the geographic proximity of products influence a consumer's decision? An evaluation in a restaurant. *Annals of Tourism Research*, 52, 169-172. <https://doi.org/10.1016/j.annals.2015.03.001>

- Jiang, Y. P., Liang, H. M., & Sun, M. H. (2015). A method for discrete stochastic MADM problems based on the ideal and nadir solutions. *Computers & Industrial Engineering*, 87(1), 114-125. <https://doi.org/10.1016/j.cie.2015.04.019>
- Ji, P., Zhang, H. Y., & Wang, J. Q. (2018). A fuzzy decision support model with sentiment analysis for items comparison in e-commerce: The case study of PConline.com. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 2018, 1-12. <https://doi.org/10.1109/TSMC.2018.2875163>
- Liang, X., Jiang, Y. P., & Liu, P. D. (2018). Stochastic multiple criteria decision making with 2-tuple aspirations: a method based on disappointment stochastic dominance. *International Transactions in Operational Research*, 25(1), 913-940. <https://doi.org/10.1111/itor.12430>
- Li, G., Law, R., Vu, H. Q., & Rong, J. (2013). Discovering the hotel selection preferences of Hong Kong inbound travelers using the Choquet Integral. *Tourism Management*, 36(3), 321-330. <https://doi.org/10.1016/j.tourman.2012.10.017>
- Li, H., Ye, Q., & Law, R. (2013). Determinants of customer satisfaction in the hotel industry: An application of online review analysis. *Asia Pacific Journal of Tourism Research*, 18(7), 784-802. <https://doi.org/10.1080/10941665.2012.708351>
- Li, J., & Wang, J. (2018). Multi-criteria decision-making with probabilistic hesitant fuzzy information based on expected multiplicative consistency. *Neural Computing and Applications*, 2018, 1-19. <https://doi.org/10.1007/s00521-018-3753-1>
- Li, Y. M., & Lai, C. Y. (2014). A social appraisal mechanism for online purchase decision support in the micro-blogsphere. *Decision Support Systems*, 59(1), 190-205. <https://doi.org/10.1016/j.dss.2013.11.007>
- Liu, Y., Bi, J. W., & Fan, Z. P. (2017a). Ranking products through online reviews: A method based on sentiment analysis technique and intuitionistic fuzzy set theory. *Information Fusion*, 36, 149-161. <https://doi.org/10.1016/j.inffus.2016.11.012>
- Liu, Y., Bi, J. W., & Fan, Z. P. (2017b). A method for ranking products through online reviews based on sentiment classification and interval-valued intuitionistic fuzzy TOPSIS. *International Journal of Information Technology & Decision Making*, 16(06), 1497-1522. <https://doi.org/10.1142/S021962201750033X>
- Luo, S., Zhang, H., Wang, J., & Li, L. (2019). Group decision-making approach for evaluating the sustainability of constructed wetlands with probabilistic linguistic preference relations. *Journal of the Operational Research Society*, 2019, 1-17. <https://doi.org/10.1080/01605682.2018.1510806>
- Opricovic, S. (1998). *Multi-criteria optimization of civil engineering systems* (PhD Thesis). Faculty of Civil Engineering, Belgrade.
- Opricovic, S., & Tzeng, G. H. (2002). Multicriteria planning of post-earthquake sustainable reconstruction. *Computer-Aided Civil and Infrastructure Engineering*, 17(3), 211-220. <https://doi.org/10.1111/1467-8667.00269>
- Sabokbar, H. F., Ayashi, A., Hosseini, A., Banaitis, A., Banaitienè, N., & Ayashi, R. (2016). Risk assessment in tourism system using a fuzzy set and dominance-based rough set. *Technological and Economic Development of Economy*, 22(4), 554-573. <https://doi.org/10.3846/20294913.2016.1198840>
- Song, H. (2015). The structure of customer satisfaction with cruise-line services: An empirical investigation based on online word of mouth. *Current Issues in Tourism*, 18(5), 450-464. <https://doi.org/10.1080/13683500.2013.776020>
- Sparks, B. A., & Browning, V. (2011). The impact of online reviews on hotel booking intentions and perception of trust. *Tourism Management*, 32(6), 1310-1323. <https://doi.org/10.1016/j.tourman.2010.12.011>
- Sparks, B. A., So, K. K. F. K., & Bradley, G. L. (2016). Responding to negative online reviews: The effects of hotel responses on customer inferences of trust and concern. *Tourism Management*, 53, 74-85. <https://doi.org/10.1016/j.tourman.2015.09.011>

- Sun, Y., Hrušovský, M., Zhang, C., & Lang, M. X. (2018). A time-dependent fuzzy programming approach for the green multimodal routing problem with rail service capacity uncertainty and road traffic congestion. *Complexity*, 2018, 8645793. <https://doi.org/10.1155/2018/8645793>
- Sun, Y., Li, X., Liang, X., & Zhang, C. (2019). A bi-objective fuzzy credibilistic chance-constrained programming approach for the hazardous materials road-rail multimodal routing problem under uncertainty and sustainability. *Sustainability*, 11(9), 2577. <https://doi.org/10.3390/su11092577>
- Vermeulen, I. E., & Seegers, D. (2009). Tried and tested: The impact of online hotel reviews on consumer consideration. *Tourism Management*, 30(1), 123-127. <https://doi.org/10.1016/j.tourman.2008.04.008>
- Xiang, Z., Du, Q., Ma, Y., & Fan, W. (2017). A comparative analysis of major online review platforms: Implications for social media analytics in hospitality and tourism. *Tourism Management*, 58, 51-65. <https://doi.org/10.1016/j.tourman.2016.10.001>
- Ye, Q., Li, H. Y., & Wang, Z. S. (2012). The influence of hotel price on perceived service quality and value in e-tourism: An empirical investigation based on online traveler reviews. *Journal of Hospitality & Tourism Research*, 38(1), 23-39. <https://doi.org/10.1177/1096348012442540>
- Yu, S., Wang, J., Wang, J. Q., & Li, L. (2018). A multi-criteria decision-making model for hotel selection with linguistic distribution assessments. *Applied Soft Computing*, 67, 741-755. <https://doi.org/10.1016/j.asoc.2017.08.009>
- Zhang, C. B., Zhang, H. Y., & Wang, J. Q. (2018). Personalized restaurant recommendation method combining group correlations and customer preferences. *Information Sciences*, 454-455, 128-143. <https://doi.org/10.1016/j.ins.2018.04.061>
- Zhang, G., Dong, Y., & Xu, Y. (2014). Consistency and consensus measures for linguistic preference relations based on distribution assessments. *Information Fusion*, 17(1), 46-55. <https://doi.org/10.1016/j.inffus.2012.01.006>
- Zhang, X. L., & Xing, X. M. (2017). Probabilistic linguistic VIKOR method to evaluate green supply chain initiatives. *Sustainability*, 9(7), 12-31. <https://doi.org/10.3390/su9071231>