

Article

Research on a Hotel Collaborative Filtering Recommendation Algorithm Based on the Probabilistic Language Term Set

Erwei Wang ¹, Yingyin Chen ^{2,*} and Yumin Li ¹

¹ Zhuhai Campus, Beijing Institute of Technology, Zhuhai 519088, China; 08119@bitzh.edu.cn (E.W.); 08067@bitzh.edu.cn (Y.L.)

² Faculty of Business Administration, University of Macau, Macau, China

* Correspondence: yc37045@connect.um.edu.mo; Tel.: +86-138-2231-2427

Abstract: In the face of problems such as information overload and the information cocoon resulting from big data, it is a key point of current research to solve the problem of semantic fuzziness of online reviews and improve the accuracy of personalized recommendation algorithms by using online reviews. Based on the advantage of the probabilistic language term set to deal with fuzzy information and the historical data of online hotel reviews, this paper proposes a collaborative filtering recommendation algorithm for hotels. Firstly, the text data of hotel online reviews are crawled by a crawler and processed by jieba and TF-IDF tools. Secondly, the hotel evaluation attribute set is constructed, and the sentiment analysis of the review statements is carried out with the help of the HowNet sentiment dictionary and manual annotation method. The probabilistic language term set is used to classify the data and derive statistics, and the maximum deviation method is used to determine the weight of each attribute. Then, the cosine similarity formula is fused with the modified cosine similarity formula to calculate the similarity and construct the decision matrix. Finally, combined with the historical data of the user's hotel selection, the hotel recommendation results are generated. This paper collected review data from 10 hotels in Macau from the official "Ctrip" website. The proposed recommendation algorithm model was then applied to process and analyze the data, resulting in the generation of a ranked list of hotel recommendations. To validate the accuracy and effectiveness of this research, the recommendation results were compared with those produced by other algorithms.



Citation: Wang, E.; Chen, Y.; Li, Y. Research on a Hotel Collaborative Filtering Recommendation Algorithm Based on the Probabilistic Language Term Set. *Mathematics* **2023**, *11*, 4106. <https://doi.org/10.3390/math11194106>

Keywords: hotel recommendation; probabilistic language term set; modified cosine similarity; cosine similarity; online review

MSC: 68T09; 94A16

Academic Editors: Chao Gao, Zhen Wang and Peican Zhu

Received: 4 August 2023

Revised: 20 September 2023

Accepted: 26 September 2023

Published: 28 September 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

With the continuous development of Internet technology and the widespread application of e-commerce, people's consumption patterns have gradually shifted from offline to online. As the pandemic subsides, there is a growing demand for travel, and users can book hotels online through various platforms [1]. However, the explosive growth of data has led to the increasingly serious problem of "information overload," making the extraction of useful information from massive data increasingly important [2]. Online reviews can assist users in hotel selection, but when faced with a large amount of hotel information and online reviews, users often find themselves in a dilemma. Personalized recommendation systems can help users choose suitable hotels by mining user preferences and hotel features. Research has shown that collaborative filtering recommendation algorithms are more suitable and widely used in hotel recommendations [3].

Online reviews contain information such as user preferences and hotel features and have strong interpretability. They have been widely used in hotel recommendation algorithms. Determining how to extract user preferences and hotel features from online reviews

is crucial to improving the accuracy of hotel recommendation algorithms [4–6]. Tao et al. [7] conducted an instance analysis of hotel online reviews using the λ -fuzzy measure and the Choquet integral operator, considering the correlation between different features, and generated hotel recommendation rankings for travelers. Wang et al. [8] used statistical methods to construct a recommendation index based on online reviews to generate recommendation results. Cui et al. [9] built a scenario-based user profile, calculated similarity using the improved P-J correlation coefficient, and proposed a collaborative filtering recommendation algorithm. Qian et al. [10] applied uncertain variables and uncertain sets to design a sentiment analysis model, analyzed user preferences, and designed a new nearest-neighbor search method based on the sentiment analysis model to generate recommendation results. Li et al. [11] proposed a hybrid approach that used the LSTM (long short-term memory) algorithm to calculate user preference information from online reviews and combined it with matrix decomposition recommendation algorithms to improve recommendation accuracy. Cao et al. [12] established a CNN sentiment classifier to classify emotions in online review text, then merged the classification results into user ratings, improving collaborative filtering algorithms.

However, due to the strong semantic ambiguity of online review information and the presence of malicious reviews, the above methods have their limitations. In the field of natural language processing, scholars such as Pang [13] have proposed the concept of the probabilistic linguistic term set (PLTS) as an effective tool for handling fuzzy information. PLTS can describe the importance of different evaluation terms and has been used to solve problems such as sentiment analysis, text classification, and information retrieval, and has been widely applied in hotel reviews [14], engineering evaluation, logistics management, medical services [15], and other fields. Some researchers have combined PLTS with recommendation algorithms. Zhou et al. [16] used PLTS to extract user emotions from review information, proposed a VIKOR-based probabilistic linguistic multi-criteria decision-making method, and conducted experimental analysis using movie information recommendations. Chen et al. [17] proposed a user-friendly function to transform personalized semantics of online reviews into probability distributions corresponding to user ratings. They constructed a probability language rating matrix and used integrated probabilities to represent the fuzziness of trust, proposing a user collaborative filtering recommendation algorithm based on probability language and dual trust networks. Cui et al. [18] used PLTS to aggregate user preferences from online reviews while eliminating ambiguity in online reviews. They calculated similarity using the cosine similarity formula and generated recommendation results. In the above studies, there is room for improvement in the richness of the user sentiment hierarchy extracted from online reviews (using five-granularity linguistic terms), and the similarity calculation formulas used did not consider the influence of the evaluation scale, resulting in limited differentiation of evaluation results.

To address the above issues, this paper proposes a collaborative filtering recommendation algorithm based on a probabilistic linguistic term set (PLTS). When extracting user sentiment from online reviews, a seven-granularity linguistic term is used to more accurately reflect user preferences in online reviews and improve the accuracy of recommendation results. In attribute weight calculation, the maximum deviation method is used to determine attribute weights. In similarity calculation, given that a single cosine similarity calculation tends to be optimistic [19], while corrected cosine similarity calculation tends to be pessimistic, a fusion similarity calculation formula is used, taking into account the influence of the evaluation scale, to improve the quality of the recommendation algorithm.

2. Preliminaries

2.1. Probabilistic Language Term Set

In 2016, Pang et al. [13] proposed the probabilistic linguistic term set (PLTS), which uses probability distributions to represent different levels of importance of linguistic evaluation.

Lin and Xu [20] proposed the standardization of PLTS, the distance formula, and other related theories.

Definition 1. Assuming that $S = \{s_\alpha | \alpha = -\tau, \dots, -1, 0, 1, \dots, \tau\}$ is a Linguistic Term Set (LTS) where τ is a positive integer, then the probabilistic linguistic term set $L(p)$ can be defined as:

$$L(p) = \left\{ L^k(p^k) \mid L^k \in S, p^k \geq 0, k = 1, 2, \dots, \#L(p), \sum_{k=1}^{\#L(p)} p^k \leq 1 \right\} \tag{1}$$

where L^k represents the k th term, p^k is the probability of the k th term, and $L^k(p^k)$ represents the language term L^k with a probability value p^k . $\#L(p)$ represents the number of terms in the probabilistic language term set $L(p)$.

With reference to [18], it is assumed that there is a linguistic term set:

$$S = \left\{ \begin{array}{l} s_{-3} : \text{Extremely poor}, s_{-2} : \text{Very poor}, s_{-1} : \text{Poor}, \\ s_0 : \text{Generally}, s_1 : \text{Good}, s_2 : \text{Very good}, s_3 : \text{Extremely good} \end{array} \right\}$$

If 40% of the users give a ‘‘Very good’’ rating of the hotel, 45% give a ‘‘Generally’’ rating of the hotel, and 15% give a ‘‘Poor’’ rating of the hotel, then there is $L(p) = \{s_{-1}(0.15), s_0(0.45), s_2(0.4)\}$.

The sum of the term probabilities in PLTS is $\sum_{k=1}^{\#L(p)} p^k$. When $\sum_{k=1}^{\#L(p)} p^k = 1$, this PLTS is called the complete probabilistic language term set; when $\sum_{k=1}^{\#L(p)} p^k < 1$, this PLTS is called the incomplete probabilistic language term set. In PLTS operation or information aggregation, the existence of an incomplete PLTS will lead to the loss of the final information and the inaccuracy of the result, so it is necessary to standardize the incomplete PLTS before the operation or information aggregation. Pang’s standardized formula is as follows:

$$\tilde{L}(p) = \left\{ L^k(\tilde{p}^k) \mid k = 1, 2, \dots, \#L(p) \right\} \tag{2}$$

where $\tilde{p}^k = \frac{p^k}{\sum_{k=1}^{\#L(p)} p^k}$.

Definition 2 [13]. When the number of terms in two probabilistic language term sets is different, it will increase the difficulty of calculation and comparison. Therefore, it is necessary to standardize the number of terms in the two probabilistic language term sets, fill in the missing terms, and make the number of terms in the two probabilistic language term sets the same. The supplementary process is as follows:

Suppose there are two arbitrary sets of probabilistic language terms $L_n(p)$ and $L_m(p)$. $L_n(p) = \left\{ L_n^k(p_n^k) \mid k = 1, 2, \dots, \#L_n(p) \right\}$, $L_m(p) = \left\{ L_m^k(p_m^k) \mid k = 1, 2, \dots, \#L_m(p) \right\}$, $\#L_n(p)$ and $\#L_m(p)$ the number of terms for $L_n(p)$ and $L_m(p)$, respectively. When the number of the two is not equal, the missing parts needs to be completed. If $\#L_n(p) > \#L_m(p)$, then the language terms $\#L_n(p) - \#L_m(p)$ are added to $L_m(p)$, and the added element is the smallest term in $L_m(p)$, and its probability is 0, and vice versa.

Definition 3 [13]. Assume that the probability of two arbitrary linguistic term set $L_n(p)$ and $L_m(p)$, $L_n(p) = \left\{ L_n^k(p_n^k) \mid k = 1, 2, \dots, \#L_n(p) \right\}$, $L_m(p) = \left\{ L_m^k(p_m^k) \mid k = 1, 2, \dots, \#L_m(p) \right\}$, and $\#L_n(p) = \#L_m(p)$, the distance $L_n(p)$ and $L_m(p)$ as follows:

$$d(L_n(p), L_m(p)) = \sqrt{\sum_{k=1}^{\#L_n(p)} (p_n^k r_n^k - p_m^k r_m^k)^2 / \#L_n(p)} \tag{3}$$

where r_n^k and r_m^k represent subscripts of the language terms L_n^k and L_m^k , respectively, and satisfy the following properties:

- (1) $d(L_n(p), L_n(p)) = 0$;
- (2) $d(L_n(p), L_m(p)) = d(L_m(p), L_n(p))$.

2.2. Cosine Similarity

Cosine similarity is used to calculate the distance between two vectors, mainly by calculating the cosine value of the angle between two vectors to calculate the difference between them [21].

Definition 4 [22]. Assume that there exists an arbitrary set of probabilistic language terms $L_i(p) = L_i^k(p_i^k) | k = 1, 2, \dots, \#L(p), (i = 1, 2, \dots, n)$, where $L_n^k, L_m^k \in S$. Set C represents the properties of the hotel, $C = \{c_j | c_1, c_2, \dots, c_n\}$; then, according to the cosine similarity formula, under the same attribute c_j , the formula for the similarity between $L_n(p), L_m(p)$ two probabilistic language term sets $Sim_{cs}(L_n(p), L_m(p))$ is as follows:

$$Sim_{cs}(L_n(p), L_m(p)) = \frac{\sum_{k=1}^{\#L(p)} \left(\left(\tau^{(L_n^k p_n^k)/\tau} \right) \left(\tau^{(L_m^k p_m^k)/\tau} \right) \right)}{\sqrt{\sum_{k=1}^{\#L(p)} \left(\tau^{(L_n^k p_n^k)/\tau} \right)^2} \sqrt{\sum_{k=1}^{\#L(p)} \left(\tau^{(L_m^k p_m^k)/\tau} \right)^2}} \tag{4}$$

where $\tau^{(L_n^k p_n^k)/\tau}$ and $\tau^{(L_m^k p_m^k)/\tau}$ represent the respective score sets of hotel n and hotel m under the attribute c_j , respectively. The cosine similarity satisfies the following properties:

- (1) $Sim_{cs}(L_n(p), L_m(p)) = 1$;
- (2) $Sim_{cs}(L_n(p), L_m(p)) = Sim_{cs}(L_m(p), L_n(p))$.

2.3. Modified Cosine Similarity

Modified cosine similarity is a variant of cosine similarity. Although cosine similarity can be used to measure the difference between two vectors [21], it pays more attention to the relationship between vector angles and is insensitive to the absolute values of specific data such as size and length. Therefore, in 2001, Sarvar et al. [23] proposed the concept of modified cosine similarity, and subtracted the average value of corresponding features in the calculation to reduce the problems caused by these deficiencies. Considering the influence of the degree of scoring bias on the score [24], the calculation of similarity is more reasonable.

Definition 5 [23]. Assume the existence of an arbitrary set of probabilistic language terms $L_i(p) = L_i^k(p_i^k) | k = 1, 2, \dots, \#L(p), (i = 1, 2, \dots, n)$, where $L_n^k, L_m^k \in S$. Set C represents the properties of the hotel, $C = \{c_j | c_1, c_2, \dots, c_n\}$. Under the same attribute c_j , according to the modified cosine similarity formula, $L_n(p)$ and $L_m(p)$ can obtain the similarity between the two probabilistic language term sets $Sim_{acs}(L_n(p), L_m(p))$; the formula is as follows:

$$Sim_{acs}(L_n(p), L_m(p)) = \frac{\sum_{k=1}^{\#L(p)} \left(\left(\tau^{(L_n^k p_n^k)/\tau} - \bar{\tau}_n \right) \left(\tau^{(L_m^k p_m^k)/\tau} - \bar{\tau}_m \right) \right)}{\sqrt{\sum_{k=1}^{\#L(p)} \left(\tau^{(L_n^k p_n^k)/\tau} - \bar{\tau}_n \right)^2} \sqrt{\sum_{k=1}^{\#L(p)} \left(\tau^{(L_m^k p_m^k)/\tau} - \bar{\tau}_m \right)^2}} \tag{5}$$

where $\bar{\tau}_n$ and $\bar{\tau}_m$ represent the respective average score sets of hotel n and hotel m , respectively, and the modified cosine similarity satisfies the following properties:

- (1) $Sim_{acs}(L_n(p), L_m(p)) = 1$;
- (2) $Sim_{acs}(L_n(p), L_m(p)) = Sim_{acs}(L_m(p), L_n(p))$.

3. Materials and Methods

3.1. Model Proposal

With the widespread development and adoption of various travel apps and websites, tourists often refer to evaluations of hotels by other users. They combine these evaluations with their own preferences for hotel attributes such as service, price, and location when assessing and choosing hotels. However, due to time constraints and the abundance of hotel review information, tourists can only access a portion of the information, leading to

a lack of comprehensive and objective assessments of hotels. Addressing the challenge of providing valuable information to tourists from vast amounts of data is essential, and recommendation algorithms are a commonly used approach to solve this problem.

Considering the aforementioned issues, current research on hotel recommendation algorithms rarely integrates users' historical data, such as textual reviews. Furthermore, probabilistic linguistic term sets (PLTSs) exhibit favorable characteristics. They can effectively capture users' perceptions and intentions from textual data, enabling precise recommendations. Hence, this paper proposes a model that combines PLTS with online reviews to enhance the accuracy of hotel recommendation algorithms. This model consists of three main modules.

Module 1: Sentiment Analysis.

Using the jieba tool for tasks such as word segmentation and part-of-speech tagging, the original review text is processed. Sentiment words and intensity adverbs are analyzed to construct the PLTS for the review text.

Module 2: Attribute Weight Calculation and Similarity Computation.

Based on the PLTS of the review text, attribute weights and similarities are computed to obtain a hotel similarity evaluation matrix.

Module 3: Prediction and Recommendation.

Utilizing the hotel similarity evaluation matrix based on the probabilistic linguistic dataset and incorporating user historical data, the model generates a recommendation list.

3.2. Material and Methods

With the large-scale development and popularization of various tourism software packages and websites, tourists prefer to read other users' evaluation of hotels and evaluate and choose hotels based on their own preferences for hotel attributes, such as service, price, and transportation. However, due to the limited time and the large amount of hotel review information, tourists can only obtain partial information and lack a comprehensive and objective evaluation of the hotel. Determining how to provide tourists with valuable information from a large number of data is an urgent problem to be solved, and recommendation algorithms are common used to solve this problem.

In this paper, the set $H = \{h_i|h_1, h_2, \dots, h_m\}$ will be employed to represent the set of hotels, while the set $\{c_j|c_1, c_2, \dots, c_n\}$ will represent the hotel attributes that users consider when referencing other users' review information during the hotel selection process. A large number of online review texts are processed based on the probabilistic language term set, and after obtaining useful information, the system will recommend multiple hotels according to the historical data of hotel selection by tourists. The specific idea is shown in Figure 1.

3.3. Data Acquisition and Preprocessing

After determining the research content, the data will be crawled and processed to prepare for the subsequent recommendation algorithm. The specific steps are shown in Figure 2.

3.3.1. Data Crawling

In this paper, Python 3.6 is used to crawl the hotel review data of a tourism website in the same period of time, and the set H represents the hotel set, that is, $H = \{h_i|h_1, h_2, \dots, h_m\}$, where m is the number of hotels. Each hotel is crawled for l review information, which is exported to a csv file.

3.3.2. Online Comment Text Processing

After obtaining the review data, the jieba tool performs word segmentation and parts-of-speech tagging processing on the original review text. In addition, we select and refer to the stop word table UTF-8 provided by IKAnalyzer to delete words with little or no value, such as modal words and function words, so as to reduce the interference of stop words on

sentence analysis, and the sentences with ambiguous semantics are processed manually. After word segmentation and stop word processing, the words describing hotel attributes and evaluation in the review are obtained. In this paper, $Z_i^k = z_i^{k1}, z_i^{k2}, \dots, z_i^{km}, \dots, z_i^{kp}$ represents the word set of the k comment statement of the i hotel, and z_i^{km} represents the m word in the word set Z_i^k .

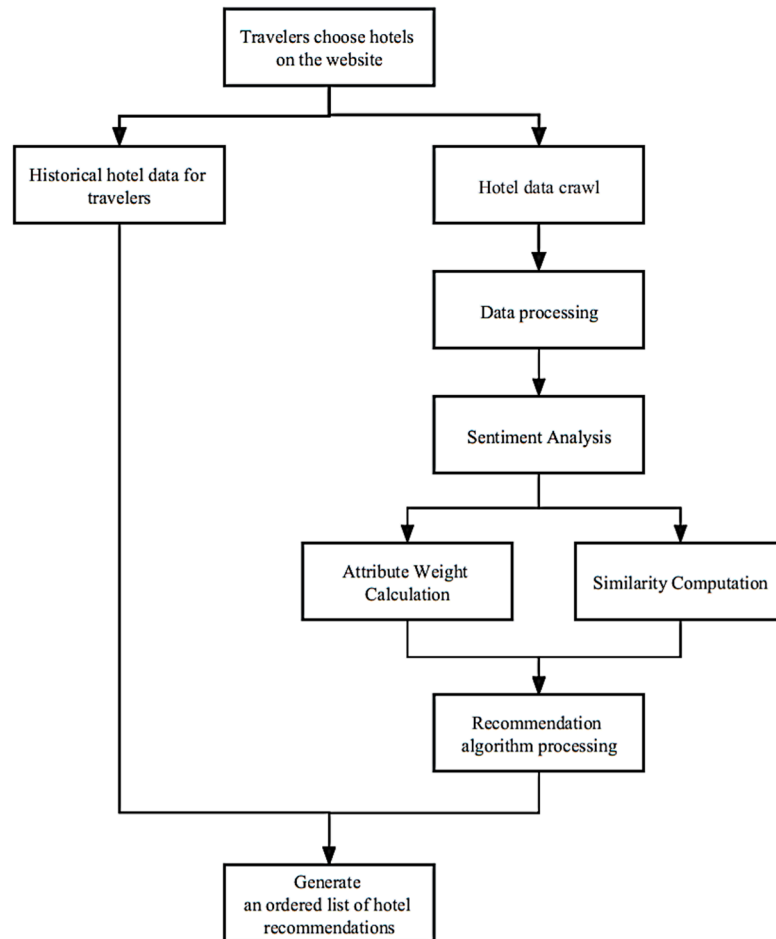


Figure 1. Research idea.

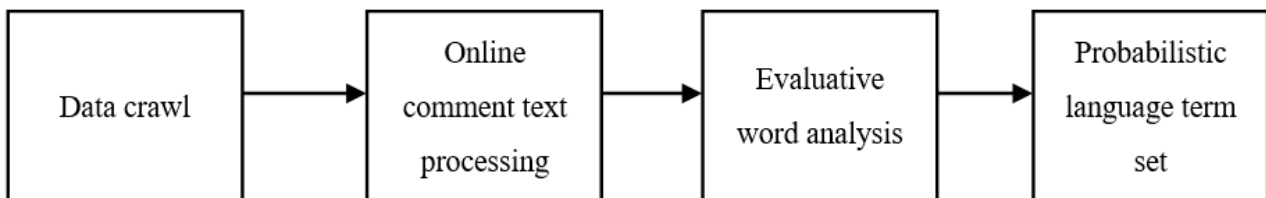


Figure 2. Idea of data acquisition and preprocessing.

3.3.3. Word Analysis of Hotel Evaluation

(1) Construct word knowledge base

As there is currently no sentiment dictionary specifically for hotel review analysis, this paper developed its own vocabulary for this model, which was constructed by a professional familiar with the web terminology of online hotel reviews, by manually examining the review content. Tourists’ online evaluations of hotels generally involve multiple attributes of the hotel and represent different attitudes, such as the attitude towards hotel facilities and services, as shown in Table 1. By looking at online reviews of

various hotels, we can identify the attribute objects evaluated in the reviews and judge their tendency fields. After consulting experts, we finally built a word knowledge base for analyzing hotel reviews [25].

Table 1. Categories of words for hotel evaluation.

Category	Definition	Example
Attribute words	Words that describe some aspect of a hotel	“service”, “facilities”, “price”, etc.
Emotional words	Words that express emotional attitudes towards hotel properties	“comfortable”, “convenient”, “affordable”, etc.
Adverb of degree	Words that modify emotional words and express emotional levels	“very”, “very”, “quite”, etc.

(2) Establish standard attribute words

In the model, the attribute words in the reviews reflect the tourists’ attention to some aspects of the hotel when choosing the hotel, but the expression of the attributes is slightly different. Therefore, before the analysis and processing of the attribute words, the word frequency of the reviews is analyzed, the standard attribute words are selected, and the common attribute words are classified. The properties of the hotel are represented by the set $C, C = \{c_j | c_1, c_2, \dots, c_n\}$, and Z_{ij}^k is used to represent the set of attribute words involved in the comment k of the i hotel, $Z_{ij}^k = z_{i1}^k, z_{i2}^k, \dots, z_{in}^k$. Assuming that z_j^c represents the common attribute words in the comment, and z_j^s represents the selected standard attribute words, the similarity between z_j^c and z_j^s is calculated according to the calculation method of semantic similarity in the Synonymous Word Forest [19], and the common attribute words are transformed into standard attribute words.

(3) Determine sentiment analysis rules

Since the expression of comments is highly flexible, the HowNet sentiment dictionary and manual annotation methods are used for sentiment analysis of comment statements [22]. In the analysis of hotel emotion words, in addition to the positive and negative emotion analysis, 7-granularity is also adopted to distinguish the degree of emotion. As shown in Table 2 below, the positive and negative of emotion words express the tourists’ positive, neutral, or even negative attitudes towards the hotel, while the degree adverbs describing emotion are mainly used to describe the intensity of the attitude, such as “extremely”, “very” and other degree adverbs.

Table 2. Reference for emotion analysis and intensity ranking of emotion word classification scale.

Example of Expression	Scale Expression and Intensity Ranking	Example
Positive emotion	$s_1 < s_2 < s_3$	“Clean”, “comfortable”, “cost-effective”, etc.
Neutral emotion	s_0	“Ok”, “average”, etc.
Negative emotion	$s_{-3} > s_{-2} > s_{-1}$	“Dirty”, “noisy”, “disorderly”, “remote”, etc.

In this paper, we construct a more detailed 7-granularity classification of emotion words, transform emotion adverbs into 7-granularity language terms, and explore the intensity of emotion words. Suppose S is a 7-granularity size scale set, $S = \{s_{-3}, s_{-2}, s_{-1}, s_0, s_1, s_2, s_3\}$, where the positive and negative signs of the s subscript represent positive and negative emotions, respectively, s_0 represents a neutral attitude, and the number of subscripts represents the degree of attitude. The larger the number, the stronger the attitude. s_{-3} represents “very poor”, s_{-2} represents “very poor”, s_{-1} represents “bad”. s_0 represents “general”, s_1 represents “good”, s_2 represents “very good”, s_3 represents “very good”.

In order to make the degree of positive and negative emotions more accurate, the division method follows the following rules:

- (a) Positive or negative words without degree adverbs are classified as s_{-1} or s_1 ;

- (b) Other degree adverbs are uniformly classified as s_{-2} or s_2 according to the classification of emotion words;
- (c) Emotion intensity words containing “extreme”, “invincible”, “only”, and “very” are classified as s_{-3} or s_3 according to the positive and negative direction of emotion words.

3.3.4. Conversion of Online Comments

After the hotel reviews are preprocessed, the data are described and counted using probabilistic language term sets (PLTSs). After data crawling, the attributes in the q review data of each hotel are transformed with language terms, and all attributes and the frequency of language terms are counted. This is represented by a probabilistic language term set, which is divided into the following three steps, as shown in Figure 3:

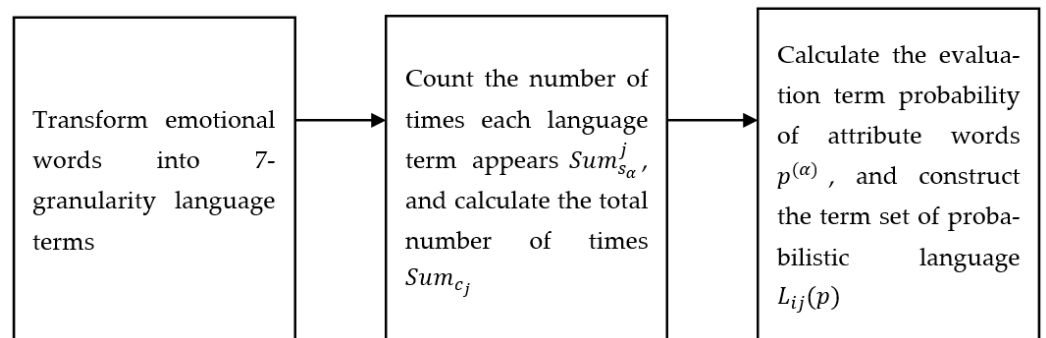


Figure 3. Steps for converting comments into a probabilistic language term set.

Step 1: Assume that hotel k exists, analyze its review text set $Z_i^k = z_i^{k1}, z_i^{k2}, \dots, z_i^{km}, \dots, z_i^{kp}$, attribute word set $Z_{ij}^k = z_{i1}^k, z_{i2}^k, \dots, z_{in}^k$, and then transform the emotion words in the evaluation text into 7-granularity language terms according to the rules.

Step 2: Calculate the frequency $Sum_{s_alpha}^j$ of each evaluation language term in each attribute word, and count the total frequency Sum_{c_j} of each attribute c_j .

Step 3: Calculate the evaluation term probability $p^{(\alpha)}$ of attribute words. The calculation formula is shown in Formula (6). After calculating the $p^{(\alpha)}$ of all attribute words, use the probabilistic language term set $L_{ij}(p)$ to describe the evaluation set of the j th attribute of the i hotel.

$$p^{(\alpha)} = \frac{Sum_{s_alpha}^j}{Sum_{c_j}} \tag{6}$$

3.4. Steps for Recommendation Algorithms

After the comments are converted into a probabilistic language term set, the recommendation algorithm will be constructed. The specific steps are shown in Figure 4.

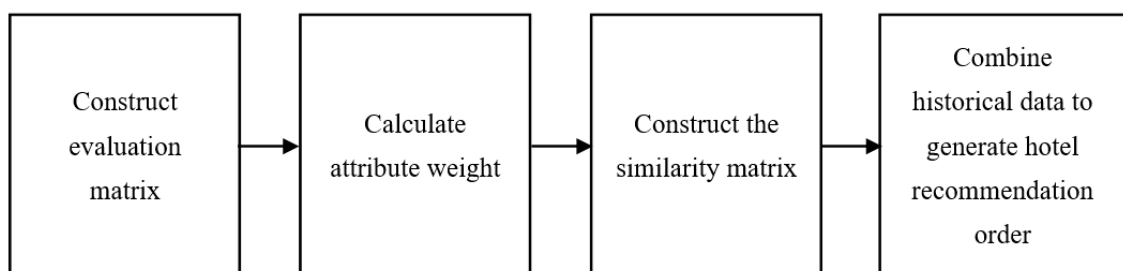


Figure 4. Idea of recommendation algorithm.

Step 1: Construct evaluation matrix

The evaluation matrix P of the hotel is constructed to describe the obtained probabilistic language term set. L_{ij} is used to represent the evaluation information about the hotel h_i attribute c_j . $i \in [1, m], j \in [1, n], i$ and j are integers.

$$P = \begin{bmatrix} L_{11}(p) & L_{12}(p) & \cdots & L_{1n}(p) \\ L_{21} & \cdots & \cdots & L_{2n}(p) \\ \vdots & \vdots & \vdots & \vdots \\ L_{m1} & L_{m2} & \cdots & L_{mn}(p) \end{bmatrix}$$

Step 2: Determine the attribute weights

Each tourist has different needs for hotels, so they will pay different degrees of attention to the attributes of hotels [18]. In this paper, the maximum deviation method is adopted to determine the attribute weight. In general, the greater the attribute weight, the greater the deviation degree between hotels, and the stronger their attribute differentiation degree [20]. Assuming that there is a hotel h_i , under the same attribute c_j , the deviation degree between the hotel h_i and other hotels is first calculated, and then the total deviation degree between the hotel h_i and other hotels is calculated. Secondly, the total deviation of attributes in the evaluation matrix is calculated, and the maximum deviation optimization model is constructed. Finally, the Lagrange function is used to solve the problem, and the standardized attribute weights are obtained. Specific calculations are as follows:

Suppose that there is an attribute weight set $w = \{w_j | j = 1, 2, \dots, n\}$. According to Formula (3), under attribute c_j , the deviation degree between hotel h_i and other hotels is:

$$d_{ij}(w) = \sum_{l=1, l \neq i}^m d(L_{ij}(p), L_{lj}(p)) \tag{7}$$

where $d(L_{ij}(p), L_{lj}(p)) = \sqrt{\frac{\sum_{k=1}^{\#L_{ij}(p)} (p_{ij}^{(k)} r_{ij}^{(k)} - p_{lj}^{(k)} r_{lj}^{(k)})^2}{\#L_{ij}(p)}}$.

The total deviation degree between hotel h_i and each hotel is:

$$d_j(w) = \sum_{i=1}^m d_{ij} \tag{8}$$

In the evaluation matrix P , the total deviation degree between all attributes is:

$$d_P(w) = \sum_{j=1}^n w_j d_j \tag{9}$$

The maximum deviation optimization model is constructed as follows:

$$\begin{cases} \max d_P(w) = \sum_{j=1}^n w_j \sum_{i=1}^m \sum_{l=1, l \neq i}^m d(L_{ij}(p), L_{lj}(p)) \\ \sum_{j=1}^n (w_j)^2 = 1, w_j \geq 0 \end{cases} \tag{10}$$

Then, the Lagrange function is constructed to solve the model:

$$L(w, \lambda) = \sum_{j=1}^n w_j \sum_{i=1}^m \sum_{l=1, l \neq i}^m d(L_{ij}(p), L_{lj}(p)) + \frac{\lambda}{2} (\sum_{j=1}^n (w_j)^2 - 1) \tag{11}$$

Finally, the normalized attribute weights are obtained:

$$w_j = \frac{\sum_{i=1}^m \sum_{l=1, l \neq i}^m d(L_{ij}(p), L_{lj}(p))}{\sum_{j=1}^n \sum_{i=1}^m \sum_{l=1, l \neq i}^m d(L_{ij}(p), L_{lj}(p))} \tag{12}$$

Step 3: Construct the similarity matrix

The method of weighted similarity is used to calculate the similarity between hotels, that is, the similarity of two hotels under the same attributes is first calculated, and then the weight of attributes is added to calculate the weighted similarity, and finally the similarity matrix between hotels is constructed, from which the similarity between the two matrices can be directly observed.

In view of the fact that the cosine similarity calculation results tend to be optimistic [19], while the modified cosine similarity calculation results tend to be pessimistic, this paper adopts a compromise and an improved similarity calculation formula, as shown in Formula (13), where α is the adjustment coefficient:

$$Sim = \alpha Sim_{cs} + (1 - \alpha) Sim_{acs} \tag{13}$$

- (1) According to Formula (13), calculate the similarity of h_i and h_l of two hotels under the same attribute c_j , $\#L_{ij}(p) = \#L_{lj}(p)$,

$$Sim(L_{ij}(p), L_{lj}(p)) = \frac{1}{2} Sim_{cs}(L_{ij}(p), L_{lj}(p)) + \frac{1}{2} Sim_{acs}(L_{ij}(p), L_{lj}(p)) \tag{14}$$

- (2) The weighted similarity between the two hotels is obtained according to the attribute weights.

$$Sim(h_i, h_l) = \sum_{j=1}^n w_j \times Sim(L_{ij}(p), L_{lj}(p)) \tag{15}$$

- (3) Construct the hotel similarity matrix M :

$$M = \begin{bmatrix} Sim(h_1, h_1) & Sim(h_1, h_2) & \cdots & Sim(h_1, h_m) \\ Sim(h_2, h_1) & Sim(h_2, h_2) & \cdots & Sim(h_2, h_m) \\ \vdots & \vdots & \vdots & \vdots \\ Sim(h_m, h_1) & Sim(h_m, h_2) & \cdots & Sim(h_m, h_m) \end{bmatrix}$$

Step 4: Hotel recommendation ranking

According to the similarity matrix obtained in Step 3 and the historical hotel stays of tourists, the hotels that reach the threshold are recommended and sorted according to the similarity size, and the hotel recommendations are made to tourists. Assume that h_i is the hotel where the visitor has previously stayed, i is the hotel subscript, and the recommended threshold is Δ . When $Sim(h_i, h_j) \geq \Delta$, the system will recommend the hotel h_j to the tourist; otherwise, the hotel h_j will not be recommended.

4. Case Application

4.1. Data Crawling

In the case analysis, this paper randomly selected 10 hotels in Macau according to their popularity, including five-star, four-star, and three-star hotels. In accordance with the principle of priority of recent reviews, Python was used to crawl 200 pieces of recent online review information, and “hotel name” and “review content” were selected as valid fields for subsequent analysis. Table 3 shows the 10 hotels selected, namely two three-star hotels, three four-star hotels and five five-star hotels. The hotel collection can be used as $H = \{h_1, h_2, h_3, h_4, h_5, h_6, h_7, h_8, h_9, h_{10}\}$.

4.2. Definition of Hotel Standard Attributes

Based on the word frequency analysis of reviews and the analysis of historical residents’ concerns about hotels, hotel attribute words can be divided into the following five attribute sets $C = \{c_1, c_2, c_3, c_4, c_5\}$, as shown in Table 4, from which it can be seen that there are five standard attribute words, such as \bar{z}_1^c for “service”, \bar{z}_2^c for “location”, etc.

Table 3. Selected hotels in Macau.

Hotel Code (h_i)	Hotel Name	Hotel Class
h_1	Emperor Hotel	Three-star level
h_2	Fu Hua Hotel	Three-star level
h_3	Hotel Beverly Plaza	Four-star
h_4	Casa Real Hotel	Four-star
h_5	Hotel Golden Dragon	Four-star
h_6	The Parisian Macao	Five-star
h_7	Riviera Hotel Macau	Five-star
h_8	Hotel Lisboa	Five-star
h_9	The Venetian Macao	Five-star
h_{10}	Sheraton Grand Macao	Five-star

Table 4. Standard hotel property categories.

Standard Attribute Word Character (\bar{z}_j^c)	Standard Attribute Word Name	Definition	Example
\bar{z}_1^c	Service	Refers to the necessary labor provided to meet the material or spiritual needs of the guests, including the hotel’s service content, mode, attitude, speed, and efficiency.	“Check-in”, “Room type upgrade”, “late check-out”, “Reception”, “drop-off service”, etc.
\bar{z}_2^c	Position	Refers to the geographical location and traffic conditions of the hotel, considering the distance from the airport, station, business center, tourist attractions and other environmental factors, as well as whether it is easy to find and other factors.	“Bus station”, “walking”, “attractions”, “airport”, “shopping”, “transportation”, etc.
\bar{z}_3^c	Facility	Refers to the hardware and software facilities provided by the hotel, mainly including the necessary accommodation facilities, hot water supply facilities, communication facilities, fire-fighting facilities, etc.	“WIFI”, “projection”, “bed”, “TV”, “bathtub”, “Restaurant”, “gym”, etc.
\bar{z}_4^c	Price	Refers to the price of staying at the hotel and the cost performance.	“Cost performance”, “cheap price”, etc.
\bar{z}_5^c	Environment	Refers to the hotel’s environmental comfort and cleanliness, mainly including the two aspects of the internal and external environment.	“Quiet”, “clean”, “comfortable”, “bright”, etc.

4.3. Data Preprocessing

Firstly, word segmentation of the online review data retrieved by jieba is performed, and then the stop words are deleted to obtain the attribute words, emotion words, and degree adverbs of emotion in the review, and the attribute set $Z_i^k = \{z_i^{k1}, z_i^{k2}, \dots, z_i^{km}, \dots, z_i^{kp}\}$ is formed. Then, according to the method of synonym merging, attribute words are converted into standard attribute words, and emotion words and emotion degree adverbs are analyzed and converted into 7-granularity language terms. The number of occurrences $Sum_{s_\alpha}^j$ of each evaluation term in each standard attribute word is counted, and the total number of occurrences Sum_{c_j} of each attribute c_j is counted. For example, for the hotel h_1 , Table 5 shows the analysis results.

Thus, its probabilistic language term set L_{1j} is represented as follows:

$$L_{11} = \{s_3(0.125), s_2(0.470), s_1(0.110), s_0(0.200), s_{-1}(0.035), s_{-2}(0.035), s_{-3}(0.025)\}$$

$$L_{12} = \{s_3(0.030), s_2(0.295), s_1(0.190), s_0(0.485)\}$$

$$L_{13} = \{s_2(0.070), s_1(0.055), s_0(0.755), s_{-1}(0.060), s_{-2}(0.055), s_{-3}(0.005)\}$$

$$L_{14} = \{s_3(0.065), s_2(0.095), s_1(0.245), s_0(0.550), s_{-1}(0.010), s_{-2}(0.005), s_{-3}(0.030)\}$$

$$L_{15} = \{s_3(0.030), s_2(0.185), s_1(0.495), s_0(0.200), s_{-1}(0.040), s_{-2}(0.035), s_{-3}(0.015)\}$$

Table 5. Statistics of the occurrence of evaluation terms of hotel h_1 .

Hotel (h_i)	Standard Attribute (\bar{z}_j^c)	s_{-3}	s_{-2}	s_{-1}	s_0	s_1	s_2	s_3	Sum c_j
h_1	Service (\bar{z}_1^c)	5	7	7	40	22	94	25	200
	Position (\bar{z}_2^c)	0	0	0	97	38	59	6	200
	Facility (\bar{z}_3^c)	1	11	12	151	11	14	0	200
	Price (\bar{z}_4^c)	6	1	2	110	49	19	13	200
	Environment (\bar{z}_5^c)	3	7	8	40	99	37	6	200

4.4. Hotel Recommendation Result Generation

Step 1: After the preliminary data processing, the evaluation matrix of 10 hotels can be constructed according to the probabilistic language term set.

$$P = \begin{bmatrix} \left\{ \begin{matrix} s_3(0.125), s_2(0.470), \\ s_1(0.110), s_0(0.200), \\ s_{-1}(0.035), s_{-2}(0.035), \\ s_{-3}(0.025) \end{matrix} \right\} & \left\{ \begin{matrix} s_3(0.030), s_2(0.295), \\ s_1(0.190), s_0(0.485) \end{matrix} \right\} & \cdots & \left\{ \begin{matrix} s_3(0.030), s_2(0.185), \\ s_1(0.495), s_0(0.200), \\ s_{-1}(0.040), s_{-2}(0.035), \\ s_{-3}(0.015) \end{matrix} \right\} \\ \left\{ \begin{matrix} s_3(0.005), s_2(0.120), \\ s_1(0.185), s_0(0.660), \\ s_{-1}(0.025), s_{-2}(0.005) \end{matrix} \right\} & \left\{ s_0(0.554), s_1(0.446) \right\} & \cdots & \left\{ \begin{matrix} s_3(0.005), s_2(0.051), \\ s_1(0.198), s_0(0.633), \\ s_{-1}(0.103), s_{-2}(0.009) \end{matrix} \right\} \\ \vdots & \vdots & \ddots & \vdots \\ \left\{ \begin{matrix} s_3(0.070), s_2(0.500), \\ s_1(0.085), s_0(0.280), \\ s_{-1}(0.030), s_{-2}(0.010), \\ s_{-3}(0.025) \end{matrix} \right\} & \left\{ \begin{matrix} s_3(0.030), s_2(0.130), \\ s_1(0.400), s_0(0.435), \\ s_{-2}(0.005) \end{matrix} \right\} & \cdots & \left\{ \begin{matrix} s_3(0.029), s_2(0.166), \\ s_1(0.275), s_0(0.483), \\ s_{-1}(0.026), s_{-2}(0.015), \\ s_{-3}(0.006) \end{matrix} \right\} \end{bmatrix}_{10 \times 5}$$

Step 2: First, the probabilistic language term set L_{ij} is standardized, and then the maximum deviation method is used to calculate the weight of each attribute.

For example, for hotel h_1 and h_2 , for the same standard properties, the standardized treatments of \bar{L}_{11} and \bar{L}_{21} are as follows:

$$L_{11} = \{s_2(0.470), s_3(0.125), s_1(0.110), s_0(0.200), s_{-1}(0.035), s_{-2}(0.035), s_{-3}(0.025)\}$$

$$L_{21} = \{s_2(0.120), s_1(0.185), s_3(0.005), s_0(0.660), s_{-2}(0.000), s_{-2}(0.005), s_{-1}(0.025)\}$$

According to Formula (7), under the standard attribute \bar{z}_1^c , the difference between hotel h_1 and h_2 is $d(L_{11}(p), L_{21}(p)) = 0.3185$. In the same way, the mutual difference between different attributes of each hotel can be calculated, and finally the deviation degree of each attribute can be obtained: $d_1(w) = 13.3258, d_2(w) = 14.1691, d_3(w) = 3.9918, d_4(w) = 6.0544, d_5(w) = 7.3635$. Then, according to Formula (8), the total deviation degree can be obtained as follows:

$$d_P(w) = 13.3258w_1 + 14.1691w_2 + 3.9918w_3 + 6.0544w_4 + 7.3635w_5$$

The maximum deviation degree optimization model is constructed and solved using the Lagrange function $L(w, \lambda)$. The normalized attribute weights are as follows:

$$w_1 = 0.3035, w_2 = 0.3227, w_3 = 0.0909, w_4 = 0.1151, w_5 = 0.1677$$

Step 3: Calculate the similarity between pairwise matrices and construct the similarity matrix.

Taking hotels h_1 and h_2 as an example, the cosine similarity and modified cosine similarity of the two hotels can be calculated according to Formulas (4) and (5), and then the final similarity of the two hotels can be calculated according to Formula (13) (Table 6).

Table 6. Similarity calculation process of hotel h_1 and h_2 for the same attribute.

	$L_{11}(p), L_{21}(p)$	$L_{12}(p), L_{22}(p)$	$L_{13}(p), L_{23}(p)$	$L_{14}(p), L_{24}(p)$	$L_{15}(p), L_{25}(p)$	$L_{11}(p), L_{21}(p)$
Sim_{cs}	0.9988	0.9960	0.9999	0.9995	0.9999	0.9988
Sim_{acs}	0.9996	0.2460	0.8983	0.8731	0.9944	0.9996
Sim	0.9992	0.6210	0.9491	0.9363	0.9972	0.9992

Based on the formula and attribute weights, the final similarity between the two hotels is 0.8650. Similarly, the similarity between the two hotels is calculated, and the similarity matrix M is constructed.

$$M = \begin{bmatrix} 1 & 0.865 & 0.874 & 0.858 & 0.970 & 0.850 & 0.846 & 0.810 & 0.965 & 0.874 \\ 0.865 & 1 & 0.995 & 0.959 & 0.848 & 0.977 & 0.951 & 0.940 & 0.838 & 0.977 \\ 0.874 & 0.995 & 1 & 0.965 & 0.859 & 0.978 & 0.958 & 0.926 & 0.842 & 0.980 \\ 0.858 & 0.959 & 0.965 & 1 & 0.837 & 0.973 & 0.979 & 0.906 & 0.849 & 0.971 \\ 0.970 & 0.848 & 0.859 & 0.837 & 1 & 0.857 & 0.853 & 0.806 & 0.956 & 0.855 \\ 0.850 & 0.977 & 0.978 & 0.973 & 0.857 & 1 & 0.987 & 0.960 & 0.860 & 0.981 \\ 0.846 & 0.951 & 0.958 & 0.979 & 0.853 & 0.987 & 1 & 0.940 & 0.861 & 0.981 \\ 0.810 & 0.940 & 0.926 & 0.906 & 0.806 & 0.960 & 0.940 & 1 & 0.854 & 0.953 \\ 0.965 & 0.838 & 0.842 & 0.849 & 0.956 & 0.860 & 0.869 & 0.854 & 1 & 0.877 \\ 0.874 & 0.977 & 0.980 & 0.971 & 0.855 & 0.981 & 0.981 & 0.953 & 0.877 & 1 \end{bmatrix}$$

Step 4: Set a threshold Δ to generate the hotel recommendation order based on the historical check-in data of the passenger. If there are two hotels with the same similarity, the recommendation order is the same.

Assuming that the preset recommendation threshold is $\Delta = 0.805$, if the passenger has used hotel h_1 before, the sequence of hotel recommendation is $h_5 \succ h_9 \succ h_3 \succ h_{10} \succ h_2 \succ h_4 \succ h_6 \succ h_7 \succ h_8$, and the same can be obtained in other cases.

Assuming that the preset recommendation threshold is $\Delta = 0.805$, if the passenger has used hotel h_8 before, the sequence of hotel recommendation is: $h_6 \succ h_{10} \succ \{h_2, h_7\} \succ h_3 \succ h_4 \succ h_9 \succ h_1 \succ h_5$, and the same can be obtained in other cases. For hotel h_8 , since the similarity between h_2 and h_7 is the same, hotel h_2 and hotel h_7 are recommended in the same order. Similarly, for hotel h_{10} , there is the same recommendation ranking situation.

5. Comparative Analysis

In this paper, the hotel recommendation algorithm based on the probabilistic language term set can make use of the online review information of users, intelligently provide users with hotel recommendations that match their preferences, and at the same time sort multiple recommended hotels. In order to highlight the excellent performance of this algorithm, we conducted an in-depth analysis and comparison with the recommended algorithms proposed by Cui et al. [19] and Zhou et al. [16], as shown in Table 7. In order to prove the feasibility of the algorithm, we performed the calculations. According to the law, for the h_2 hotel, the recommended hotels are ranked. To rank the recommended hotels, $h_6 \succ h_{10} \succ \{h_2, h_7\} \succ h_3 \succ h_4 \succ h_9 \succ h_1 \succ h_5$, the results are consistent with the outcomes obtained using the algorithm proposed in this paper. According to the attribute weight formula ($w_j = \frac{N_j}{\sum_{j=1}^n N_j}$) and the VIKOR method criteria from the literature, the recommended hotel ranking is $h_6 \succ h_{10} \succ \{h_2, h_7\} \succ h_3 \succ h_4 \succ h_9 \succ h_1 \succ h_5$, which aligns with the results obtained by the algorithm presented in this paper. Except for hotel

h_2 , the optimal hotel recommendations are the same as those in this paper, effectively validating the feasibility of the proposed method.

Table 7. Algorithm comparison.

Recommended Algorithm	Attribute Weight	Attribute Similarity	Ranking Method	Recommended Method
Textual algorithm	Maximum deviation degree method	$Sim = \alpha Sim_{cs} + (1 - \alpha) Sim_{acs}$	The similarity between hotels	Combined with historical data, based on hotel similarity
Cui et al. [19], literature algorithm	Maximum deviation degree method	Sim_{cs}	The similarity between products	Based on similarities between products
Zhou et al. [16], literature algorithm	Number of occurrences of subject terms	-	VIKOR method	Based on user product evaluation score

As observed from the table above, this paper employs the maximum deviation method, consistent with Cui’s literature algorithm, for handling weights. Meanwhile, Zhou’s study primarily relies on the frequency of occurrence of keywords, which can be easily influenced by the quantity of keywords. Regarding attribute similarity, this paper combines cosine similarity with modified rectangular similarity, while Cui’s article relies solely on cosine similarity. Concerning the sorting method, this paper primarily ranks the hotels based on their similarity to generate a recommendation list. In terms of the recommendation approach, this paper combines user’s historical hotel selection data with the generated hotel similarity list for recommendations. In summary, the method proposed in this paper exhibits superiority.

6. Conclusions and Future Research Directions

In response to the application challenges of a large volume of hotel online review data in recommendations, this paper employs a probabilistic language terminology set to express hotel attributes and user preferences within online reviews. It uses a seven-level granularity language terminology to more precisely capture the fuzzy emotional semantics in online reviews and proposes a hotel collaborative filtering recommendation algorithm based on PLTS. We utilize jieba for text segmentation, conduct sentiment analysis using the HowNet sentiment dictionary, and effectively transform the review terms into PLTS using the TF-IDF statistical method. We compile probability language information by constructing an evaluation matrix, determine attribute weights using the maximum deviation method, calculate the weighted similarity between hotels using an improved fusion similarity computation formula, and present the recommendation results. Via a case analysis with examples from Ctrip, the effectiveness of the recommendation algorithm proposed in this paper is validated. Furthermore, via a comparative analysis with other methods, the relative advancement of the recommendation approach presented in this paper is demonstrated.

Limitation and Future Work: The data acquisition and attributes in the case study are relatively small, and the selection of hotel attributes relies on subjective experience. In the future, these shortcomings will be addressed; we plan to extract hotel attributes from online reviews based on the model, and use larger-scale data to optimize the algorithm.

Author Contributions: Conceptualization, E.W.; Methodology, E.W. and Y.C.; Software, Y.L.; Investigation, Y.C. and Y.L.; Data curation, Y.C.; Writing—original draft, Y.C.; Writing—review & editing, E.W.; Project administration, E.W. All authors have read and agreed to the published version of the manuscript.

Funding: This study was supported by the Research on Tourism Product Recommendation Algorithm in the Key Fields of Guangdong General Universities (Digital Economy; No. 2021ZDZX3010).

Data Availability Statement: Both data and algorithms are listed in the article.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Government of the Macao Special Administrative Region. Bureau of Statistics and Census [DB/OL]. Available online: <https://www.dsec.gov.mo/zh-CN/Statistic?id=401> (accessed on 10 January 2023).
2. Shokeen, J.; Rana, C. A study on features of social recommender systems. *Artif. Intell. Rev.* **2020**, *53*, 965–988. [\[CrossRef\]](#)
3. Liu, S.; Chen, Z.; Li, X. Time-semantic-aware Poisson tensor factorization approach for scalable hotel recommendation. *Inf. Sci.* **2019**, *504*, 422–434. [\[CrossRef\]](#)
4. Linden, G.; Smith, B.; York, J. Amazon.com recommendations: Item-to-item collaborative filtering. *IEEE Internet Comput.* **2003**, *7*, 76–80. [\[CrossRef\]](#)
5. Shi, Y.; Larson, M.; Hanjalic, A. Collaborative filtering beyond the user-item matrix: A survey of the state of the art and future challenges. *ACM Comput. Surv. (CSUR)* **2014**, *47*, 1–45. [\[CrossRef\]](#)
6. Zhu, P.; Cheng, L.; Gao, C.; Wang, Z.; Li, X. Locating multi-sources in social networks with a low infection rate. *IEEE Trans. Netw. Sci. Eng.* **2022**, *9*, 1853–1865. [\[CrossRef\]](#)
7. Tao, L.; You, T.; Yuan, Y. Hotel Selection Based on Hotel Feature Information and Online Review Information. *J. Northeast Univ. Nat. Sci.* **2019**, *40*, 1667–1672.
8. Wang, S.; Wu, S. Personalized Attraction Recommendation Using Online Reviews. *J. Huaqiao Univ. Nat. Sci.* **2018**, *39*, 467–472.
9. Cui, C.; Wang, X.; Li, W. Research on Contextual Environment-Based User Profile Tourism Product Recommendation Algorithm. *Pract. Underst. Math.* **2019**, *49*, 122–131.
10. Lihua, S. Advanced collaborative filtering recommendation model based on sentiment analysis of online review. *J. Shandong Univ. Eng. Sci.* **2019**, *49*, 47–54.
11. Li, X.J.; Deng, G.S.; Wang, X.Z.; Wu, X.L.; Zeng, Q.W. A hybrid recommendation algorithm based on user comment sentiment and matrix decomposition. *Inf. Syst.* **2023**, *117*, 102244. [\[CrossRef\]](#)
12. Cao, H.; Kang, J. Study on improvement of recommendation algorithm based on emotional polarity classification. In Proceedings of the 2020 5th International Conference on Computer and Communication Systems (ICCCS), Shanghai, China, 15–18 May 2020; IEEE: Piscataway, NJ, USA, 2020; pp. 182–186.
13. Pang, Q.; Wang, H.; Xu, Z. Probabilistic linguistic term sets in multi-attribute group decision making. *Inf. Sci.* **2016**, *369*, 128–143. [\[CrossRef\]](#)
14. Zhang, Y.; Liang, D.; Xu, Z. Cross-platform hotel evaluation by aggregating multi-website consumer reviews with probabilistic linguistic term set and Choquet integral. *Ann. Oper. Res.* **2022**, 1–35. [\[CrossRef\]](#) [\[PubMed\]](#)
15. Celotto, A.; Loia, V.; Senatore, S. Fuzzy linguistic approach to quality assessment model for electricity network infrastructure. *Inf. Sci.* **2015**, *304*, 1–15. [\[CrossRef\]](#)
16. Zhou, H.; Ma, H.; Liu, J. Research on Film and Television Recommendation Algorithm Integrating Sentiment Analysis and Probabilistic Language. *Inf. Stud. Theory Pract.* **2020**, *43*, 180–186.
17. Chen, S.; Zhang, C.; Zeng, S.; Wang, Y.; Su, W. A probabilistic linguistic and dual trust network-based user collaborative filtering model. *Artif. Intell. Rev.* **2023**, *56*, 429–455. [\[CrossRef\]](#)
18. Cui, C.; Wang, X.; Li, W. Research on Tourism Attraction Recommendation Algorithm Based on User Online Reviews. *Syst. Sci. Math.* **2020**, *40*, 1103.
19. Cui, C.; Wei, M.; Che, L.; Wu, S.; Wang, E. Hotel recommendation algorithms based on online reviews and probabilistic linguistic term sets. *Expert Syst. Appl.* **2022**, *210*, 118503. [\[CrossRef\]](#)
20. Lin, M.; Xu, Z. Probabilistic linguistic distance measures and their applications in multi-criteria group decision making. *Soft Comput. Appl. Group Decis.-Mak. Consens. Model.* **2018**, *357*, 411–440.
21. Chen, Z.; Huang, C.; Zhou, R.; Wang, Z.; Cao, F. Crowdsourcing Task Recommendation Algorithm Based on Collaborative Filtering. *Inf. Technol. Informatiz.* **2021**, *8*, 119–121.
22. Luo, S.Z.; Zhang, H.Y.; Wang, J.Q.; Li, L. Group decision-making approach for evaluating the sustainability of constructed wetlands with probabilistic linguistic preference relations. *J. Oper. Res. Soc.* **2019**, *70*, 2039–2055. [\[CrossRef\]](#)
23. Sarwar, B.; Karypis, G.; Konstan, J.; Riedl, J. Item-based collaborative filtering recommendation algorithms. In Proceedings of the 10th international conference on World Wide Web, Hong Kong, China, 1–5 April 2001; pp. 285–295.
24. Chu, H.; Liu, Q.; Mu, C. A Collaborative Filtering Recommendation Algorithm Improved for Modified Cosine Similarity. *J. Yantai Univ. Nat. Sci. Eng. Ed.* **2021**, *34*, 330–336.
25. Jiang, L.; Zhang, Q. Research on Personalized Recommendation Strategy Based on Sentiment Analysis of Reviews—A Case Study of Douban Movie Reviews. *Inf. Stud. Theory Pract.* **2017**, *40*, 99–104.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.