Journal of Systems Science and Information
Feb., 2024, Vol. 12, No. 1, pp. 96–112
DOI: 10.21078/JSSI-2023-0135

Research on Recommendation Algorithms Based on Cloud Models in Probabilistic Linguistic Environments

Peng YANG
College of Computer and Information Engineering, Henan University of Economics and Law, Zhengzhou 450046*, China*
E-mail: 804604955@*qq.com*

Xifeng MA*
College of Computer and Information Engineering, Henan University of Economics and Law, Zhengzhou 450046*, China*
E-mail: 876397199@*qq.com*

Meng WEI
College of Computer and Information Engineering, Henan University of Economics and Law, Zhengzhou 450046*, China*
E-mail: *m*1587592449@163*.com*

Chunsheng CUI
College of Computer and Information Engineering, Henan University of Economics and Law, Zhengzhou 450046*, China*
E-mail: *traition*@126*.com*

Libin CHE
College of Computer and Information Engineering, Henan University of Economics and Law, Zhengzhou 450046*, China*
E-mail: *c*19545656515@163*.com*

Abstract To solve the problem that the traditional cloud model can't directly process the textual review information in the recommendation algorithm, this paper combines the merits of the cloud model in transforming qualitative and quantitative knowledge with the multi-granularity advantages of probabilistic linguistic term sets in representing uncertain information, and proposes a recommendation algorithm based on cloud model in probabilistic language environment. Initially, this paper quantifies the attributes in the review text based on the probabilistic linguistic term set. Subsequently, the maximum deviation method is used to determine the weight of each attribute in the evaluation information of the product to be recommended, and the comprehensive evaluation number and attribute weight are

Received September 4, 2023, accepted October 19, 2023
Supported by the Humanities and Social Sciences Research Planning Fund Project of the Ministry of Education (23YJA860004), the Major Basic Research Project of Philosophy and Social Sciences in Higher Education Institutions in Henan Province (2024-JCZD-27) and 2021 Project of Huamao Financial Research Institute of Henan University of Economics and Law (HCHM-2021YB001)
* Corresponding author

converted into the digital characteristic value of the cloud model by using the backward cloud generator. Finally, the products are recommended and sorted based on the digital characteristic value of the cloud model. The algorithm is applied to the recommendation of 10 hotels, and the results show that the method is effective and practical, enriching the application of cloud models in the recommendation field.

Keywords recommendation algorithm; cloud model; probabilistic linguistic term set; text reviews

1 Introduction

With the development of e-commerce, people are becoming increasingly accustomed to booking hotels online when they travel[1, 2]. In the whole process of booking a hotel online, the reviews of other historical users of the target hotel provide great reference value for potential users[3, 4]. Most of the existing online hotel booking platforms use star reviews as the main source of user feedback information. However, direct star reviews are not as valuable as text reviews because of problems such as random ratings by users. Although some platforms use the combination of ratings and reviews as a reference, the problem of insufficient accuracy in text information processing has always existed. In recent years, research on recommendation algorithms based on review information has gradually become a hot topic. Pahari, et al.[5] proposed a hotel recommendation ranking for online reviews based on an intuitive fuzzy algorithm. Zhou, et al.[6] put patient consultation text and doctor labels as samples into the Word2vec model and LDA model for training and put forward a doctor recommendation algorithm based on label and patient consultation text, which effectively solved the problems of the loose structure of online medical information and insufficient accuracy of doctor recommendation on medical platforms. Chen, et al.[7] proposed a recommendation algorithm that integrates user text comments and score interactions to better characterize user preferences and product characteristics and alleviate the impact of sparse score data in collaborative filtering recommendation algorithms on recommendation results. Cui, et al.[8] obtained user preferences based on user online comment information and proposed a tourist attraction recommendation algorithm based on user online comments, which effectively improved the accuracy of tourist attraction recommendations and met the personalized needs of users. To meet the real-time needs of food recommendation, Deng, et al.[9] used web content extraction and multi text automatic summary and other technologies to extract online comments and built a food recommendation algorithm based on online comments. By combining the research on many existing recommendation algorithms, it can be seen that recommendation based on review information occupies a considerable proportion of various types of recommendation algorithms, which can also reflect that many recommendation algorithms rely on user comments to a large extent.

Cloud model theory, combined with the relevant theories of fuzzy mathematics and statistics, can effectively solve the problem of fuzziness in qualitative linguistics. It can determine the relationship between fuzziness and randomness to better reflect the quantitative characteristics of qualitative linguistics. Cloud models have been gradually applied to research on recommendation systems[10]. Wang, et al.[11] proposed a web service recommendation algorithm based on the user spatial location scoring cloud model to solve the problem of inaccurate recommendations caused by the imbalance of user similarity calculation in web service recommendations. Gao, et al.[12] used cloud model theory to quantitatively describe the qualitative indicators of line scheme measurement to reflect the attribute information of the indicators more accurately, and based on this, the line scheme was selected and sorted. Xiao, et al.[13] adopted a comprehensive cloud model to generate scoring standards and transform user ratings into a recommendation system prediction method, which improved the accuracy deviation of traditional forecasting methods in the case of sparse user rating data. Liu, et al.[14] selected the similarity of the cloud model based on the project score and the similarity of the cloud model based on the user score to synthesize the final prediction score, thus constructing the recommendation algorithm based on the two-dimensional cloud model, which improved the accuracy of the recommendation system to a certain extent. In summary, recommendation algorithms based on cloud models are mostly developed based on score data in the literature, and there is insufficient research on qualitative descriptive data such as text comments. In addition, previous studies have not considered the weight ratio between different attributes in reviews and the uncertainty and fuzziness of review statements[15, 16]. The innovation of this paper is that it combines the cloud model and the probabilistic linguistic term set to process the text through the probabilistic linguistic term set, which solves the weight relationship of multiple hotel attributes well. At the same time, by using the cloud model to recommend text processed by the probabilistic linguistic term set, the ambiguity and uncertainty of the review text are fully considered.

2 Materials and Methods

2.1 Probabilistic Linguistic Term Set

In a complex decision-making environment and with limited thinking ability, decision-makers often prefer to use qualitative descriptions to express evaluations[17, 18]. To effectively solve this problem, the concept of linguistic variables was developed[19]. With the increasing complexity of decision-making problems, it is increasingly difficult for a single linguistic variable to adequately express decision-makers' preferences toward the problem. To overcome this deficiency, Rodriguez, et al.[20] proposed the concept of HFLTSs. The linguistic terms in HFLTSs have the same weight. However, in actual situations, due to the different probability distributions, importance, and credibility of information, the linguistic terms provided by decision-makers usually have different weights[21]; that is, the probability of linguistic terms is not equal.

In 2016, Pang, et al.[22] put forward the concept of the probabilistic linguistic term set (PLTS) and gave relevant theories such as PLTS standardization and distance equations.

Definition 1 Suppose S is a linguistic term set (LTS) and $S=\{s_\alpha|\alpha=-\tau,\cdots,-1,0,1,\cdots,\tau\}$ and τ are positive integers; then, the probabilistic linguistic term set can be expressed as:

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

where L^k represents the kth term and p^k is the probability of the term k. $\#L(p)$ represents the number of terms in the probabilistic linguistic term set $L(p)$.

When the sum of the term probabilities in a PLTS is equal to 1, immediately $\sum_{k=1}^{\#L(p)}p^k=1$, the PLTS is said to be a complete probabilistic linguistic term set. When the sum of the term probabilities in a PLTS is less than 1, immediately $\sum_{k=1}^{\#L(p)} p^k < 1$, the PLTS is called an incomplete probabilistic linguistic term set. In the PLTS operation or information assembly, the existence of incomplete PLTS will lead to the loss of the final information and the inaccuracy of the result, so before the operation or information assembly, incomplete PLTS should be standardized. Pang gives the following equation for standardization:

$$\dot{L}(p) = \{L^k(\dot{p}^k) | k = 1, 2, \cdots, \#L(p)\},$$

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

Let $L_1(p) = \{L_1^k(p_1^k) | k = 1, 2, \cdots, \#L_1(p)\}$ and $L_2(p) = \{L_2^k(p_2^k) | k = 1, 2, \cdots, \#L_2(p)\}$ be two probabilistic linguistic term sets, and $L_1(p) = L_2(p)$; then, the distance between $L_1(p)$ and $L_2(p)$ is:

$$d(L_1(p), L_2(p)) = \sqrt{\sum_{k=1}^{\#L_1(p)} (p_1^k r_1^k - p_2^k r_2^k)^2 / \#L_1(p)}. \tag{1}$$

where r_1^k and r_2^k represent the subscripts of the linguistic terms L_1^k and L_2^k respectively.

Definition 2 Probabilistic linguistic averaging operator (PLA)

$$\begin{aligned} \text{PLA}(L_1(p), L_2(p), \cdots, L_i(p)) &= \frac{1}{n}(L_1(p) \bigoplus L_2(p) \bigoplus \cdots \bigoplus L_n(p)) \\ =& \frac{1}{n}\left(\bigcup_{L_1^k \in L_1(p), L_2^k \in L_2(p), \cdots, L_n^k \in L_n(p)} p_1^k L_1^k \bigoplus p_2^k L_2^k \bigoplus \cdots \bigoplus p_n^k L_n^k \right). \end{aligned} \tag{2}$$

2.2 Maximum Deviation Method

In this paper, the maximum deviation degree method is used to determine the attribute weight. Under a certain attribute, the greater the deviation degree between schemes, the stronger the differentiation of the attribute, and the greater the attribute weight. Let $w = \{w_j | j = 1, 2, \cdots, n\}$ be the attribute weight set and calculate the deviation degree with the equation. Under the attribute c_j, the deviation degree between hotel x_i and other hotels is:

$$d_{ij}(w) = \sum_{l=1, l \neq i}^{m} d(L_{ij}(p), L_{lj}(p)), \tag{3}$$

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

Under the attribute c_j, the total deviation degree between the schemes is:

$$d_j(w) = \sum_{i=1}^{m} d_{ij}. \tag{4}$$

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

$$d_P(w) = \sum_{j=1}^{n} w_j d_j. \tag{5}$$

The maximum deviation optimization model is constructed as follows:

$$\begin{cases} \max d_P(w) = \sum_{j=1}^{n} w_j \sum_{i=1}^{m} \sum_{l=1,l\neq i}^{m} d(L_{ij}(p), L_{lj}(p)) \\ \sum_{j=1}^{n} (w_j)^2 = 1,\ w_j \geq 0. \end{cases} \tag{6}$$

Construct the Lagrange function to solve the model:

$$L(w,\lambda) = \sum_{j=1}^{n} w_j \sum_{i=1}^{m} \sum_{l=1,l\neq i}^{m} d(L_{ij}(p), L_{lj}(p)) + \frac{\lambda}{2} \left(\sum_{j=1}^{n} (w_j)^2 - 1 \right). \tag{7}$$

Finally, obtain the standardized attribute weight:

$$w_j = \frac{\sum_{i=1}^{m} \sum_{l=1,l\neq i}^{m} d(L_{ij}(p), L_{lj}(p))}{\sum_{j=1}^{n} \sum_{i=1}^{m} \sum_{l=1,l\neq i}^{m} d(L_{ij}(p), L_{lj}(p))}. \tag{8}$$

2.3 Cloud Model and Related Theories

Li Deyi, academician of the Chinese Academy of Engineering, proposed the concept of a cloud in 1995 based on probability theory and fuzzy mathematics and studied the correlation[23].

Definition 3 Let X be an ordinary set, $X = [1]$ is called the domain of discourse, and T is the linguistic value associated with X. The fuzzy set A in the domain X refers to the existence of A random number $\mu_A(x)$ with a stable tendency for any X, which is called the membership of X to A. If the elements in the domain of discourse are simply ordered, then X can be regarded as the underlying variable. If the elements in the discourse domain are not simply ordered, X can be mapped to another ordered section X' according to some law f, one and only one of X' corresponds to X, then X' is the foundation variable, and the distribution of membership degrees on X' is called the (membership) cloud.

According to the definition of a cloud, the membership degree of a certain point in the discourse domain is not constant but always changes slightly, but this change is not drastic and will not affect the overall characteristics of the membership cloud. Without the overall shape and condensation characteristics of the subordinate cloud, it is meaningless to discuss the membership degree of a certain point separately. We cannot determine the membership degree of a point in isolation.

Definition 4 The digital characteristics of the cloud are expressed by the three values of expected *Ex*, entropy *En* and super entropy *He*[24].

Expected *Ex*: The point that can best represent the qualitative concept A in the number domain space or the most typical sample point of the quantization of this concept.

Entropy *En*: Reflects the uncertainty of qualitative concept A, which manifests itself in three aspects. On the one hand, entropy reflects the size of the range of the cloud droplet group that can be accepted by linguistic value A in the number domain space, that is, the ambiguity, which is a measure of the qualitative concept. On the other hand, entropy also reflects the probability that the cloud droplet group in the number domain can represent the linguistic value and represents the randomness of the cloud droplet appearing on behalf of the qualitative concept. In addition, entropy also reveals the correlation between fuzziness and randomness and can also be used to represent the granularity of a qualitative concept. In general, the higher the entropy is, the more macroscopic the concept, the greater the fuzziness and randomness, and the more difficult it is to quantify with certainty.

He is the uncertainty measure of entropy, i.e., the entropy of entropy, reflecting the cohesion of the uncertainty of all points representing the linguistic value in the number domain space, i.e., the cohesion of the cloud droplet, the greater the super entropy, the greater the dispersion of the cloud droplet, the greater the randomness of the membership degree, and the greater the thickness of the cloud.

$$Ex = \bar{X},$$

$$En = \sqrt{\frac{\pi}{2}} \cdot \frac{1}{n} \sum_{i=1}^{n} |x_i - Ex|,$$

$$He = \sqrt{S^2 - En^2}.$$

2.4 Comprehensive Evaluation Method Based on the Cloud Model

This algorithm uses the cloud model to replace the membership function in the fuzzy comprehensive evaluation method and then calculates the weight matrix and comprehensive evaluation matrix. In this paper, the weight coefficient is calculated and obtained by the method of maximum separation, and the characteristic parameters of the cloud model (*Ex*, *En*, *He*) are obtained by using the reverse cloud generator[25]. The weight coefficient and fuzzy comprehensive evaluation matrix are shown below:

$$AW = [a_{w1}, a_{w2}, \cdots, a_{wq}] = \begin{bmatrix} Ex_{aw1} & En_{aw1} & He_{aw1} \\ Ex_{aw2} & En_{aw2} & He_{aw2} \\ \vdots & \vdots & \vdots \\ Ex_{awq} & En_{awq} & He_{awq} \end{bmatrix}^{\mathrm{T}}. \tag{9}$$

$$Rx = [rx_1, rx_2, \cdots, rx_q]^T = \begin{bmatrix} Ex_1 & En_1 & He_1 \\ Ex_2 & En_2 & He_2 \\ \vdots & \vdots & \vdots \\ Ex_q & En_q & He_q \end{bmatrix}^{\mathrm{T}}. \tag{10}$$

The comprehensive evaluation results are as follows:

$$\begin{aligned} B = AW \circ Rx &= \begin{bmatrix} Ex_{aw1} & En_{aw1} & He_{aw1} \\ Ex_{aw2} & En_{aw2} & He_{aw2} \\ \vdots & \vdots & \vdots \\ Ex_{awq} & En_{awq} & He_{awq} \end{bmatrix}^{\mathrm{T}} \circ \begin{bmatrix} Ex_1 & En_1 & He_1 \\ Ex_2 & En_2 & He_2 \\ \vdots & \vdots & \vdots \\ Ex_q & En_q & He_q \end{bmatrix}^{\mathrm{T}} \\ &= (Ex, En, He). \end{aligned} \tag{11}$$

Of which,

$$En = \sqrt{\sum_{i=1}^{q}\left(\left|Ex_{awi} \times Ex_i\sqrt{\left(\frac{En_{awi}}{Ex_{awi}}\right)^2 + \left(\frac{En_i}{Ex_i}\right)^2}\right|\right)^2},$$

$$He = \sqrt{\sum_{i=1}^{q}\left(\left|Ex_{awi} \times Ex_i\sqrt{\left(\frac{He_{awi}}{Ex_{awi}}\right)^2 + \left(\frac{He_i}{Ex_i}\right)^2}\right|\right)^2}, \tag{12}$$

$$Ex = \sum_{i=1}^{q}(Ex_{awi} \times Ex_i). \tag{13}$$

According to the above equation, the comprehensive evaluation score of each evaluation object can be obtained, and the analysis of its ranking, stability, randomness, and other characteristics can be continued according to the characteristic value of the cloud model. It is generally believed that the priority ranking relationship is sorted first according to the size of *Ex*, and the smaller the *Ex* is, the better the ranking. If the *Ex* of the two is the same, the smaller the *En* ranking is. If both *Ex* and *En* are the same, the smaller *He* is, the better the ranking is. It can be understood that the expected score is sorted first, and if the expectation is the same, the stability of the two is compared, and the ranking of the better stability is preferred. If the expectation and stability are the same, the randomness is compared, and the less randomness is ranked first.

3 Model Approach

Step 1 This article takes hotel recommendations as an example and uses a crawler to obtain hotel reviews. Word segmentation and part-of-speech tagging are carried out by word segmentation tools. In addition, stop words are deleted by reference to the stop words table.

Step 2 After preprocessing the comment information, use the probabilistic linguistic term set tool to describe and count the comment information. First, for hotel x, analyze each review statement set $Z_i^k = \{z_i^{k1}, z_i^{k2}, \cdots, z_i^{km}, \cdots, z_i^{kp}\}$ and attribute word set $Z_{ij}^k = \{z_{i1}^k, z_{i2}^k, \cdots, z_{in}^k\}$, and transform the emotion words in the statement into 5-granularity linguistic terms, expressed as $S = \{s_{-2}, s_{-1}, s_0, s_1, s_2\}$. Second, the number of occurrences of each evaluation term in each attribute word is counted $Q_{s_\alpha}^j$, and the total number of occurrences c_j of each attribute Q_{c_j} is counted. Finally, the evaluation term probability $p^{(\alpha)}$ of the attribute word is calculated:

$$p^{(\alpha)} = \frac{Q_{s_\alpha}^j}{Q_{c_j}}. \tag{14}$$

After the evaluation term probability of all attribute words is calculated, the probabilistic linguistics term set is used to describe $L_{ij}(p)$, representing the evaluation set of the jth attribute of the ith hotel.

Step 3 According to the probabilistic linguistic average aggregation operator in Definition 2, the probabilistic linguistic term set obtained in Step 2 is transformed into a specific value that can comprehensively reflect the user comment information. Then, it is brought into Equation (6) to determine the weight of the comment attributes.

Step 4 Use the reverse cloud generator to convert the comment information values and the weight attribute values into the form expressed by the digital features of the cloud.

Step 5 Construct the weight matrix AW and comprehensive evaluation matrix Rx.

Step 6 Use the fuzzy synthesis operator (multiplication and sum operator) to calculate the comprehensive evaluation results and make recommendations according to the results.

4 Case Application

This paper selects 10 hotels in Beijing $(x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8, x_9, x_{10})$ and uses Python to crawl the relevant basic hotel information and online reviews from the Ctrip website.

The review information fields are "User nickname", "hotel name", "travel type", "room type", "review time", "rating", "review content" and so on. To simplify the calculation, only 200 online reviews for each hotel in 2021 were extracted using the field information of "Hotel name" and "review content".

The hotel set is represented as $X = \{x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8, x_9, x_{10}\}$, corresponding to the 10 hotels above. Five standard attributes are preset: Service, location, environment, transportation, and price. Therefore, the set of attributes is $C = \{c_1, c_2, c_3, c_4, c_5\}$, and the meaning of each attribute is shown in Table 1.

Table 1 Meaning of attributes

Attributes	Meaning
Service (c_1)	The hospitality of the hotel staff before, during, and after check-in generally includes precheck-in consultation services, postcheck-in to solve any problems encountered by the guests, and postcheck-out questions to answer.
Location (c_2)	The geographical location of the hotel and the surrounding geographical environment, some hotels may be in a better location, convenient transportation, close to shopping malls, commercial streets, etc.
Environment (c_3)	The internal environment of the hotel, generally includes hygiene, safety, facilities, etc.
Transportation (c_4)	Is there a convenient transportation station near the hotel, including a bus station, subway station, train station, etc.
Price (c_5)	The price of a hotel stay. The factor varies by room type.

First, data preprocessing is performed on each review, word segmentation is performed with the Jieba program, and stop words are deleted. Expressed in set form as $Z_i^k = \{z_i^{k1}, z_i^{k2}, \cdots, z_i^{km}, \cdots, z_i^{kp}\}$, it is expressed as extraction and analysis of the attribute words, emotion words, and emotion intensity words involved in the comment statements. According to the hotel attributes, we set the standard attribute words $\bar{z}_1^c =$ "service", $\bar{z}_2^c =$ "location", $\bar{z}_3^c =$ "environment". According to the synonym merge method, the attribute words z_j^c in each comment statement are converted to the standard attribute words, and the attribute words set $Z_{ij}^k = \{z_{i1}^k, z_{i2}^k, \cdots, z_{in}^k\}$ is constructed. Emotion words and emotion intensity words are analyzed and converted into 5-granularity linguistic term $Q_{s_\alpha}^j$ representations. Count the number of occurrences of each evaluation term in each attribute word, and count the total number of occurrences Q_{c_j} of each attribute c_j.

Table 2 Statistical table of the occurrence times of evaluation terms

Hotels	Attributes	Reviews					Total
		s_{-2}	s_{-1}	s_0	s_1	s_2	Q_{c_j}
x_1	c_1	0	1	7	60	132	200
	c_2	0	0	110	44	46	200
	c_3	0	0	12	105	83	200
	c_4	0	3	135	27	35	200
	c_5	0	1	123	51	25	200
x_2	c_1	1	3	85	58	53	200
	c_2	1	6	59	47	87	200
	c_3	1	6	47	92	54	200
	c_4	0	2	154	24	20	200
	c_5	2	3	185	9	1	200
x_3	c_1	4	3	37	39	117	200
	c_2	0	2	135	14	49	200
	c_3	3	6	41	75	75	200
	c_4	0	2	154	24	20	200
	c_5	2	3	167	7	21	200
x_4	c_1	17	20	84	32	47	200
	c_2	1	1	43	53	102	200
	c_3	34	50	51	44	21	200
	c_4	1	3	74	38	84	200
	c_5	1	15	156	16	12	200
x_5	c_1	5	11	106	33	45	200
	c_2	0	1	82	50	67	200
	c_3	15	31	102	32	20	200
	c_4	0	0	134	34	32	200
	c_5	2	10	144	28	16	200
x_6	c_1	3	4	35	41	117	200
	c_2	0	0	60	53	87	200
	c_3	24	30	58	48	40	200
	c_4	4	0	135	40	21	200
	c_5	0	2	160	26	12	200
x_7	c_1	9	11	23	73	84	200
	c_2	0	3	156	23	18	200
	c_3	9	19	121	23	28	200
	c_4	0	2	157	29	12	200
	c_5	4	6	163	20	7	200

Table 2 (Continued)

Hotels	Attributes	Reviews					Total
		s_{-2}	s_{-1}	s_0	s_1	s_2	Q_{c_j}
x_8	c_1	5	7	47	82	59	200
	c_2	12	6	57	48	77	200
	c_3	6	10	82	58	44	200
	c_4	0	2	86	40	72	200
	c_5	4	6	153	18	19	200
x_9	c_1	12	7	115	42	24	200
	c_2	3	1	95	70	31	200
	c_3	16	13	99	48	24	200
	c_4	4	2	123	61	10	200
	c_5	5	8	151	20	16	200
x_{10}	c_1	13	10	82	42	53	200
	c_2	2	6	100	55	34	200
	c_3	20	11	96	45	28	200
	c_4	4	1	114	58	23	200
	c_5	7	2	172	10	9	200

Calculate the probability of evaluation terms $p^{(\alpha)}$ for attribute words in each hotel, and express the evaluation terms and probabilities with a probabilistic linguistic term set. For example, the probability of each evaluation term in hotel x_1 attribute c_1 is:

$$p^{(-2)} = \frac{0}{200} = 0, p^{(-1)} = \frac{1}{200} = 0.005, p^{(0)} = \frac{7}{200} = 0.035, p^{(1)} = \frac{60}{200} = 0.3, p^{(2)} = \frac{132}{200} = 0.66.$$

Therefore, the probabilistic linguistic term set is expressed as:

$$L_{11} = \{s_{-1}(0.005), s_0(0.035), s_1(0.300), s_2(0.660)\}.$$

According to the maximum deviation degree method, the weight of each attribute is calculated.

According to the probabilistic linguistic term set of each hotel under different attributes, construct the hotel evaluation matrix P:

$$P = \begin{bmatrix} \{s_{-1}(0.005), s_0(0.035), s_1(0.300), s_2(0.660)\} & \cdots & \{s_{-1}(0.005), s_0(0.165), s_1(0.255), s_2(0.125)\} \\ \{s_{-2}(0.005), s_{-1}(0.015), s_0(0.425), s_1(0.290), s_2(0.265)\} & \cdots & \{s_{-2}(0.010), s_{-1}(0.015), s_0(0.925), s_1(0.045), s_2(0.005)\} \\ \vdots & \ddots & \vdots \\ \{s_{-2}(0.065), s_{-1}(0.050), s_0(0.410), s_1(0.210), s_2(0.265)\} & \cdots & \{s_{-2}(0.035), s_{-1}(0.010), s_0(0.860), s_1(0.050), s_2(0.045)\} \end{bmatrix}_{10\times 5}.$$

take attribute c_1 as an example to calculate the attribute deviation degree. First, standardize $L_{11}(p)$ and $L_{21}(p)$ as follows:

$$L_{11} = \{s_2(0.660), s_1(0.300), s_0(0.035), s_{-1}(0), s_{-1}(0.005)\},$$
$$L_{21} = \{s_2(0.265), s_1(0.290), s_0(0.425), s_{-2}(0.005), s_{-1}(0.015)\}.$$

According to equation (3), calculate the distance between $L_{11}(p)$ and $L_{21}(p)$ as:

$$d(L_{11}(p), L_{21}(p)) = 0.3534.$$

In the same way, the distance between schemes in attribute c_1 can be calculated, as well as the distance between schemes under other attributes.

According to equation (4) and equation (5), the deviation degree of each attribute can be obtained as follows:

$$d_1(w) = 17.4324, d_2(w) = 13.1345, d_3(w) = 15.1358, d_4(w) = 9.8489, d_5(w) = 4.3322.$$

According to equation (5), the total deviation degree of the property is:

$$d_P(w) = 17.3424w_1 + 13.1345w_2 + 15.1358w_3 + 9.8489w_4 + 4.3322w_5.$$

The optimal model for constructing the maximum deviation degree is:

$$\begin{cases} \max d_P(w) = 17.3424w_1 + 13.1345w_2 + 15.1358w_3 + 9.8489w_4 + 4.3322w_5 \\ (w_1)^2 + (w_2)^2 + (w_3)^2 + (w_4)^2 + (w_5)^2 = 1. \end{cases}$$

The Lagrange function is used to solve the model, and the Lagrange function $L(w, \lambda)$ is constructed:

$$\begin{aligned} L(w,\lambda) \quad = \quad & 17.3424w_1 + 13.1345w_2 + 15.1358w_3 + 9.8489w_4 + 4.3322w_5 \\ & +\frac{\lambda}{2}[(w_1)^2 + (w_2)^2 + (w_3)^2 + (w_4)^2 + (w_5)^2 - 1]. \end{aligned}$$

where λ is the Lagrange parameter. To obtain the partial derivation,

$$\begin{cases} 17.3424 + \lambda w_1 = 0, \\ 13.1345 + \lambda w_2 = 0, \\ 15.1358 + \lambda w_3 = 0, \\ 9.8489 + \lambda w_4 = 0, \\ 4.3322 + \lambda w_5 = 0, \\ 17.3424 + \lambda w_1 = 0, \\ \frac{1}{2}[(w_1)^2 + (w_2)^2 + (w_3)^2 + (w_4)^2 + (w_5)^2 - 1] = 0. \end{cases}$$

After solving the Lagrange function and using equation (8) to obtain the standardized attribute weight is:

$$w_1 = 0.2911, w_2 = 0.2193, w_3 = 0.2528, w_4 = 0.1645, w_5 = 0.0723.$$

The probabilistic linguistic aggregation average operator is used to aggregate the probabilistic linguistic term set to obtain its comprehensive evaluation number, and the attribute weight and comprehensive evaluation number are brought into the reverse cloud generator to obtain the evaluation parameter matrix, as shown in Table 3.

Table 3 Evaluation parameter matrix

Attributes	c_1	c_2	c_3	c_4	c_5
Weights	(0.2911, 0.3648, 0.1907)	(0.2193, 0.2749, 0.1437)	(0.2528, 0.3168, 0.1656)	(0.1645, 0.2062, 0.1078)	(0.0723, 0.0906, 0.0474)
x_1	(0.8075, 0.0121, 0.5290)	(0.3400, 0.4261, 0.2227)	(0.6775, 0.8491, 0.4439)	(0.2350, 0.2945, 0.1540)	(0.2500, 0.3133, 0.1638)
x_2	(0.3975, 0.4982, 0.2604)	(0.5325, 0.6674, 0.3489)	(0.4800, 0.6016, 0.3145)	(0.1550, 0.1943, 0.1015)	(0.0100, 0.0125, 0.0066)
x_3	(0.6550, 0.8209, 0.4291)	(0.2750, 0.3447, 0.1802)	(0.5325, 0.6674, 0.3489)	(0.1550, 0.1943, 0.1015)	(0.1050, 0.1316, 0.0688)
x_4	(0.1800, 0.2256, 0.1179)	(0.6350, 0.7959, 0.4160)	(−0.0800, 0.1003, 0.0524)	(0.5025, 0.6298, 0.3292)	(0.0575, 0.0721, 0.0377)
x_5	(0.2550, 0.3196, 0.1671)	(0.4575, 0.5734, 0.2997)	(0.0275, 0.0345, 0.0180)	(0.2450, 0.3071, 0.1605)	(0.1150, 0.1441, 0.0753)
x_6	(0.6625, 0.8303, 0.4340)	(0.5675, 0.7113, 0.3718)	(0.1250, 0.1567, 0.0819)	(0.1850, 0.2319, 0.1212)	(0.1200, 0.1504, 0.0786)
x_7	(0.5300, 0.6643, 0.3472)	(0.1400, 0.1755, 0.0917)	(0.1050, 0.1316, 0.0688)	(0.1275, 0.1598, 0.0835)	(0.0500, 0.0627, 0.0328)
x_8	(0.4575, 0.5734, 0.2997)	(0.4300, 0.5389, 0.2817)	(0.3100, 0.3885, 0.2031)	(0.4550, 0.5703, 0.2981)	(0.1050, 0.1316, 0.0688)
x_9	(0.1475, 0.1849, 0.0966)	(0.3125, 0.3917, 0.2047)	(0.1275, 0.1598, 0.0835)	(0.1775, 0.2225, 0.1163)	(0.0850, 0.1065, 0.0557)
x_{10}	(0.2800, 0.3509, 0.1834)	(0.2675, 0.3353, 0.1752)	(0.1250, 0.1567, 0.0819)	(0.2375, 0.2977, 0.1556)	(0.0300, 0.0376, 0.0197)

According to Table 3, the weight matrix AW can be obtained:

$$AW = [a_{w1}, a_{w2}, a_{w3}, a_{w4}, a_{w5}] = \begin{bmatrix} (0.2911, 0.3648, 0.1907) \\ (0.2193, 0.2749, 0.1437) \\ (0.2528, 0.3168, 0.1656) \\ (0.1645, 0.2062, 0.1078) \\ (0.0723, 0.0906, 0.0474) \end{bmatrix}^{\mathrm{T}}.$$

The evaluation matrix Rx of 10 hotels is shown as follows:

$$Rx_1 = \begin{bmatrix} (0.8075, 0.3648, 0.5290) \\ (0.3400, 0.4261, 0.2227) \\ (0.6775, 0.8491, 0.4439) \\ (0.2350, 0.2945, 0.1540) \\ (0.2500, 0.3133, 0.1638) \end{bmatrix}, Rx_2 = \begin{bmatrix} (0.3975, 0.4982, 0.2604) \\ (0.5325, 0.6674, 0.3489) \\ (0.4800, 0.6016, 0.3145) \\ (0.1550, 0.1943, 0.1015) \\ (0.0100, 0.0125, 0.0066) \end{bmatrix},$$

$$Rx_3 = \begin{bmatrix} (0.6550, 0.8209, 0.4291) \\ (0.2750, 0.3447, 0.1802) \\ (0.5325, 0.6674, 0.3489) \\ (0.1550, 0.1943, 0.1015) \\ (0.1050, 0.1316, 0.0688) \end{bmatrix}, Rx_4 = \begin{bmatrix} (0.1800, 0.2256, 0.1179) \\ (0.6350, 0.7959, 0.4160) \\ (-0.0800, 0.1003, 0.0524) \\ (0.5025, 0.6298, 0.3292) \\ (0.0575, 0.0721, 0.0377) \end{bmatrix},$$

$$Rx_5 = \begin{bmatrix} (0.2550, 0.3196, 0.1671) \\ (0.4575, 0.5734, 0.2997) \\ (0.0275, 0.0345, 0.0180) \\ (0.2450, 0.3071, 0.1605) \\ (0.1150, 0.1441, 0.0753) \end{bmatrix}, Rx_6 = \begin{bmatrix} (0.6625, 0.8303, 0.4340) \\ (0.5675, 0.7113, 0.3718) \\ (0.1250, 0.1567, 0.0819) \\ (0.1850, 0.2319, 0.1212) \\ (0.1200, 0.1504, 0.0786) \end{bmatrix},$$

$$Rx_7 = \begin{bmatrix} (0.5300, 0.6643, 0.3472) \\ (0.1400, 0.1755, 0.0917) \\ (0.1050, 0.1316, 0.0688) \\ (0.1275, 0.1598, 0.0835) \\ (0.0500, 0.0627, 0.0328) \end{bmatrix}, Rx_8 = \begin{bmatrix} (0.4575, 0.5734, 0.2997) \\ (0.4300, 0.5389, 0.2817) \\ (0.3100, 0.3885, 0.2031) \\ (0.4550, 0.5703, 0.2981) \\ (0.1050, 0.1316, 0.0688) \end{bmatrix},$$

$$Rx_9 = \begin{bmatrix} (0.1475, 0.1849, 0.0966) \\ (0.3125, 0.3917, 0.2047) \\ (0.1275, 0.1598, 0.0835) \\ (0.1775, 0.2225, 0.1163) \\ (0.0850, 0.1065, 0.0557) \end{bmatrix}, Rx_{10} = \begin{bmatrix} (0.2800, 0.3509, 0.1834) \\ (0.2675, 0.3353, 0.1752) \\ (0.1250, 0.1567, 0.0819) \\ (0.2375, 0.2977, 0.1556) \\ (0.0300, 0.0376, 0.0197) \end{bmatrix}.$$

Calculate the comprehensive evaluation results for hotel x_1 according to equation (11) and equation (12):

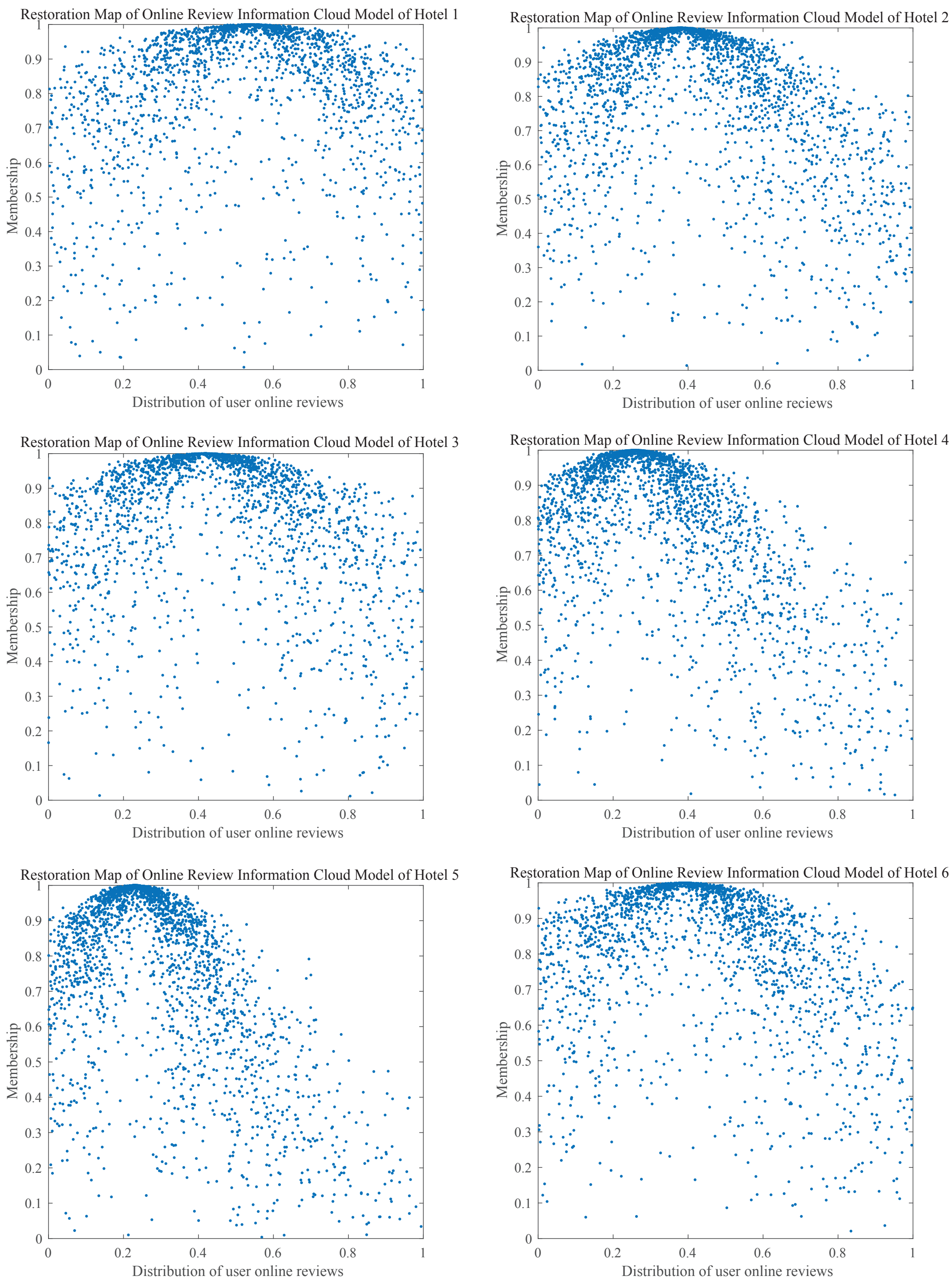
$$\begin{aligned} B_1 &= AW \circ Rx_1 \\ &= \begin{bmatrix} (0.2911, 0.3648, 0.1907) \\ (0.2193, 0.2749, 0.1437) \\ (0.2528, 0.3168, 0.1656) \\ (0.1645, 0.2062, 0.1078) \\ (0.0723, 0.0906, 0.0474) \end{bmatrix}^{\mathrm{T}} \circ \begin{bmatrix} (0.8075, 0.3648, 0.5290) \\ (0.3400, 0.4261, 0.2227) \\ (0.6775, 0.8491, 0.4439) \\ (0.2350, 0.2945, 0.1540) \\ (0.2500, 0.3133, 0.1638) \end{bmatrix} \\ &= (0.5376, 0.4628, 0.2810). \end{aligned}$$

Therefore, the comprehensive evaluation results of the selected 10 hotels x_1 to x_{10} can be obtained as shown in Table 4.

Table 4 Comprehensive evaluation results of hotel review information

Comprehensive evaluation Results	(Ex, En, He)
B_1	$(0.5376, 0.4628, 0.2810)$
B_2	$(0.3801, 0.3650, 0.1908)$
B_3	$(0.4187, 0.4299, 0.2247)$
B_4	$(0.2582, 0.3039, 0.1589)$
B_5	$(0.2301, 0.2333, 0.1219)$
B_6	$(0.3880, 0.4145, 0.2167)$
B_7	$(0.2361, 0.2853, 0.1491)$
B_8	$(0.3883, 0.3475, 0.1816)$
B_9	$(0.1790, 0.1631, 0.0853)$
B_{10}	$(0.2130, 0.1991, 0.1041)$

According to the comprehensive evaluation results listed in Table 4, the normal cloud model of the evaluation information of 10 hotels in Figure 1 can be drawn.


Restoration Map of Online Review Information Cloud Model of Hotel 1
Membership
Distribution of user online reviews
Restoration Map of Online Review Information Cloud Model of Hotel 2
Membership
Distribution of user online reciews
Restoration Map of Online Review Information Cloud Model of Hotel 3
Membership
Distribution of user online reviews
Restoration Map of Online Review Information Cloud Model of Hotel 4
Membership
Distribution of user online reviews
Restoration Map of Online Review Information Cloud Model of Hotel 5
Membership
Distribution of user online reviews
Restoration Map of Online Review Information Cloud Model of Hotel 6
Membership
Distribution of user online reviews

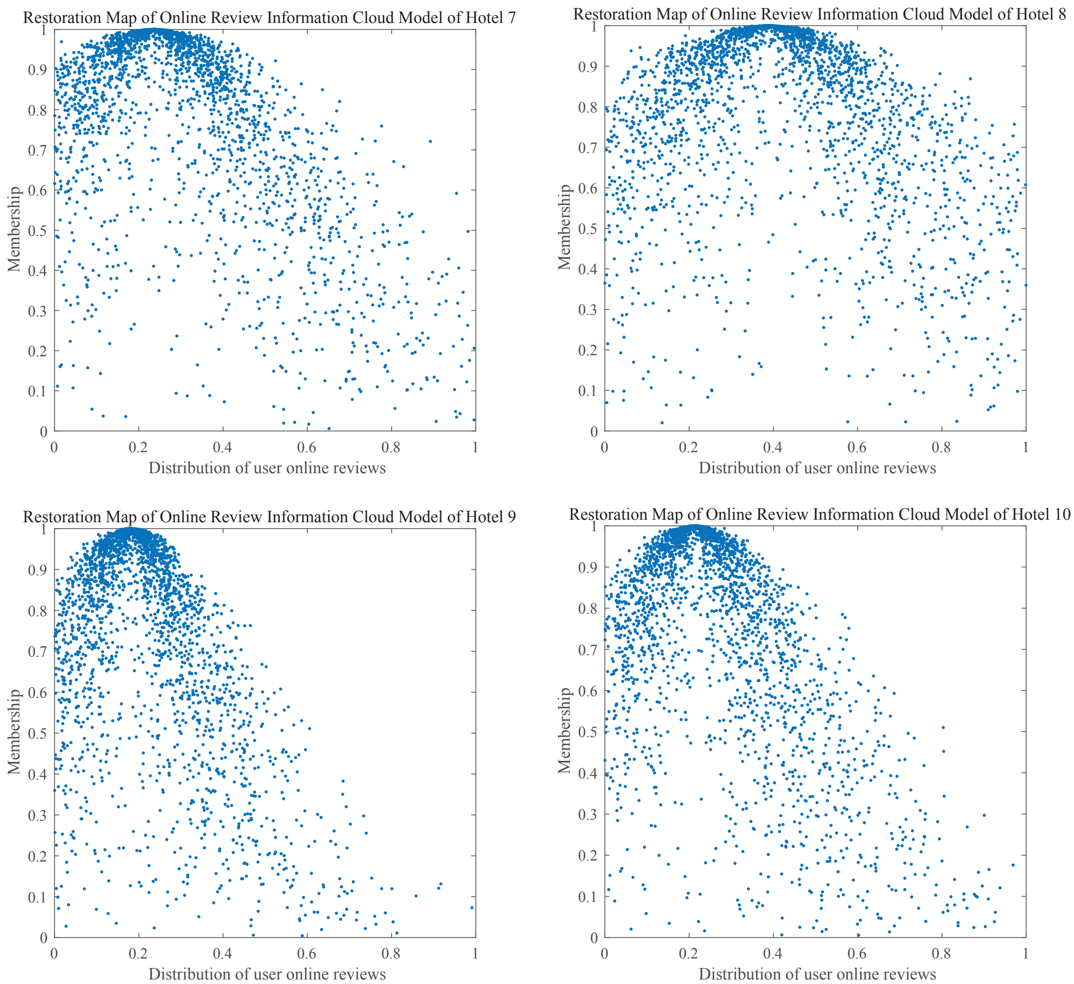


Figure 1 Normal cloud of evaluation results

When sorting the priority relationship, sort according to the size of Ex first. The smaller the Ex, the better the ranking; if the Ex of the two is the same, the smaller the En ranking is better; if both Ex and En are the same, the smaller He is, the better the ranking is. It can be understood that the expected score is sorted first, and if the expectation is the same, the stability of the two is compared, and the ranking of the better stability is preferred. If the expectation and stability are the same, the randomness is compared, and the less randomness is ranked first. For example, the Ex value in the comprehensive evaluation result B_9 of hotel x_9 is 0.1790, which is smaller than the Ex value 0.2130 in the corresponding comprehensive evaluation result B_{10} of hotel x_{10}. Therefore, we have $x_9 \succ x_{10}$, and the comparison results of other hotels can be obtained in the same way. In summary, according to the normal cloud map of the obtained evaluation results combined with the above rules, the ranking results of 10 hotels can be obtained as follows:

$$x_9 \succ x_{10} \succ x_5 \succ x_7 \succ x_4 \succ x_2 \succ x_6 \succ x_8 \succ x_3 \succ x_1.$$

5 Conclusions

Based on the theory of cloud model and the advantage of qualitative information description of probabilistic linguistic term sets, this paper proposes a hotel recommendation algorithm which can comprehensively consider the weight relationship between the attributes of comment sentences. When sorting the priority relationship, first sort according to the *Ex* value. The smaller the *Ex*, the better the ranking; If both *Ex* are the same, the smaller *En*, the better the ranking; If both *Ex* and *En* are the same, the smaller *He*, the better the ranking. It can be understood that the expected scores are sorted first, and if the expectations are the same, the stability of the two is compared, and the one with good stability takes priority; If the expectation and stability are the same, compare the randomness, and the one with less randomness will be ranked first. For example, the *Ex* value in the comprehensive evaluation result B_9 of hotel x_9 is 0.1790, which is less than the *Ex* value 0.2130 in the comprehensive evaluation result B_{10} corresponding to hotel x_{10}. Therefore the *Ex* values of the 10 hotels selected in this paper are 0.5376, 0.3801, 0.4187, 0.2582, 0.2301, 0.3880, 0.2361, 0.3883, 0.1790, 0.2130 respectively. The 10 values are sorted according to the principle of the smaller of the *Ex* value, and the sorting results of the 10 Hotels based on the *Ex* value can be obtained. To sum up, according to the normal cloud chart of the evaluation results and the above rules, the ranking results of 10 hotels are as follows: $x_9 \succ x_{10} \succ x_5 \succ x_7 \succ x_4 \succ x_2 \succ x_6 \succ x_8 \succ x_3 \succ x_1$. If there are two or more schemes with the same *Ex* value, you need to compare their *En* values to get the ranking between schemes in turn. In some special cases, the decision-making based on a single value can also be abandoned and the results can be sorted by combining *Ex* value, *En* value and *He* value. Finally, the research results have practical significance on the premise that the extracted comments are true, and the reliability of how to distinguish the extracted data needs to be improved. Because of the limited space of the data processed in this paper, text data processing in a wider range is the future research direction.

References

[1] Cui C, Wei M, Che L, et al. Hotel recommendation algorithms based on online reviews and probabilistic linguistic term sets. Expert Systems with Applications, 2022, 210: 118503.

[2] Li Z, Amagata D, Zhang Y, et al. HML4Rec: Hierarchical meta-learning for cold-start recommendation in flash sale e-commerce. Knowledge-Based Systems, 2022, 255: 109674.

[3] Lu X, Fan Y, Qin Z, et al. Evolutionary game study of choice of hotel cancellation policies based on cancellation rate. Operations Research and Management Science, 2022, 31(5): 37–42.

[4] Karn A L, Karna R K, Kondamudi B R, et al. Customer centric hybrid recommendation system for E-Commerce applications by integrating hybrid sentiment analysis. Electronic Commerce Research, 2023, 23(1): 279–314.

[5] Pahari S, Ghosh D, Pal A. An online review-based hotel selection process using intuitionistic fuzzy TOPSIS method. Progress in Computing, Analytics and Networking: Proceedings of ICCAN 2017, 2018: 203–214.

[6] Zhou X, Xiong H, Xiao B. A physician recommendation algorithm based on the fusion of label and patient consultation text. Information Sciences, 2023, 41(3): 145–154.

[7] Chen L, Fan G, Bi X, et al. Recommendation method combining interactive features of review and rating matrix. Computer Engineering and Design, 2023, 44(2): 393–399.

[8] Cui X, Wang X, Li W. Research on the algorithm of tourist attraction recommendation based on user's online comments. Journal of Systems Science and Mathematical Sciences, 2020, 40(6): 1103–1116.

[9] Deng H, Chen Z. Food recommendation system based on online reviews. Media Science and Technology of China, 2022(3): 39–41.

[10] Di P, Ni Z, Yin D. A multi-attribute decision making optimization algorithm based on cloud model and evidence theory. Systems Engineering — Theory & Practice, 2021, 41(4): 1061–1070.

[11] Wang R, Wei L, Duan Y, et al. Collaborative filtering recommendation algorithm for Web services based on user-space location score cloud model. Application Research of Computers, 2021, 38(10): 2981–2987.

[12] Gao Y, Dong X, Cheng J. Optimization method of railway route scheme in wind-sand areas based on cloud model and improved TOPSIS. Journal of Railway Science and Engineering, 2023, 20(2): 526–536.

[13] Xiao Y, Sun H, Dai T, et al. A rating prediction method based on cloud model in social recommendation system. Acta Electronica Sinica, 2018, 46(7): 1762–1767.

[14] Liu M, Man J, Peng C, et al. A collaborative filtering recommendation algorithm based on double-dimensional cloud model. Cyber Security and Data Governance, 2018, 8(1): 92–96.

[15] Xiao H, Yang P, Ma X, et al. A hesitant fuzzy multiple-attribute decision-making method considering rank relation. Journal of Intelligent & Fuzzy Systems, 2023 45(2): 3109–3121.

[16] Dong F, Wu L, Shen H. Dynamic multi-attribute decision analysis on three-parameter interval grey number of upper (lower) deviation quasi-normal distribution. Journal of Systems Science and Information, 2022, 10(3): 280–296.

[17] Zhu J, Ma X, Kou G, et al. A three-way consensus model with regret theory under the framework of probabilistic linguistic term sets. Information Fusion, 2023, 95: 250–274.

[18] Wang Q, Zhang X, Lü Z. Hybrid conditional entropy and multi-attribute decision making of incomplete decision information system. Systems Engineering — Theory & Practice, 2022, 42(12): 3401–3411.

[19] Zadeh L A. Fuzzy sets. Information and Control, 1965, 8(3): 338–353.

[20] Rodriguez R M, Martinez L, Herrera F. Hesitant fuzzy linguistic term sets for decision making. IEEE Transactions on Fuzzy Systems, 2011, 20(1): 109–119.

[21] Guo Z, Yang Y, He Z. Evaluation of emergency logistics suppliers based on the improved todim method based on probabilistic linguistic term sets. Operations Research and Management Science, 2022, 31(6): 196–203.

[22] Pang Q, Wang H, Xu Z. Probabilistic linguistic term sets in multi-attribute group decision making. Information Sciences, 2016, 369: 128–143.

[23] Li D, Meng H, Shi X. Membership clouds and membership cloud generators. Journal of Computer Research and Development, 1995, 6: 15–20.

[24] Zhang K, Zhang M. Evaluation of technological innovation and sustainable development capability based on cloud model and evidence theory. Operations Research and Management Science, 2022, 31(4): 109–115.

[25] Bao H, Li S, Wang G. Expressing unit consepts with extreme kurtosis based on cloud model. Fuzzy Systems and Mathematics, 2021, 35(6): 134–142.