Frontiers in Business, Economics and Management
ISSN: 2766-824X | Vol. 7, No. 2, 2023

Research on Information Quality Evaluation Method of Q&A Community from the Perspective of Fuzzy Evaluation

-- Taking CSDN as an Example

Dan Li*

School of Management, Shanghai University, Shanghai, China

Abstract: [Purpose/significance] Facing the current situation of information overload in the Q&A community, this paper discusses how to effectively evaluate the quality of information released in the community, so as to facilitate the staff of Q&A community to screen and recommend valuable and practical answers for platform users, which will enhance user stickiness. [Method/process] Firstly, the hesitant fuzzy information evaluation matrix is established for the qualitative evaluation information. Secondly, the entropy weight method is used to initially weight the indicators in consideration of the subjectivity of information matrix. Then, considering the overlap of indicator information, the initial weights are revised based on the correlation between the indicators to improve the accuracy of answer sorting. Finally, a comprehensive ranking of answer quality is obtained based on the weighted grey relational analysis method, and the optimal answer is selected. [Result/conclusion] An example of technical Q&A community from CSDN is selected to analyze. The results show that the hesitant fuzzy sets as evaluation information can effectively solve the problem of evaluation personnel's judgment level difference, and the revised index weights are used to make the answer ranking more in line with the actual situation.

Keywords: Q&A community, Information quality management, Hesitant fuzzy set, Grey relational analysis.

1. Introduction

According to the 49th Statistical Report on the Development of the Internet in China, as of December 2021, the number of Internet users in China has reached 1.032 billion, with 42.96 million new Internet users added in December 2020 [1]. More and more people rely on the Internet to find the information they need to deal with difficult problems and doubts they encounter in their studies, life or work. Social Q&A systems, or Q&A communities, are social platforms that focus on user knowledge sharing, whereby users post questions and exchange and interact with the community's answerers to share knowledge. As this think-tank that brings together the wisdom of the masses is generated by users, unlike the answers of authoritative experts, the problem of uneven quality of information inevitably arises. Therefore, how to solve the problem of information quality in Q&A communities and realise effective interaction in Q&A communities has become a focus of attention for scholars at home and abroad.

In the evaluation of online information resources, scholars have proposed different evaluation methods according to the characteristics of online information, for example, Wang Xianya et al. (2022) [2] used questionnaire survey method to study the quality of seller reputation information on e-commerce platforms; Cheng Quan et al. (2020) [3] studied medical and health information based on the evaluation theory and method of evidence-based reasoning; Li Jun et al. (2021) [4] used fuzzy comprehensive evaluation method to evaluate provincial government information resource sharing platforms; Chen Yue and Deng Zhaohua (2019) [5] conducted a comprehensive evaluation of the influence of health care websites based on grey correlation analysis. As grey correlation analysis is more stable than the common comprehensive evaluation methods, some scholars later used it for information quality evaluation and continuously supplemented and optimized it. Scholars have chosen different assignment methods in the evaluation process. In order that the relative importance of indicators does not violate the common sense of decision makers, some scholars have used subjective assignment methods. Guo Shunli et al. (2017) [6] conducted an expert questionnaire to determine the indicator weights and assessed the credibility of C2C platform information based on a weighted grey correlation. Some scholars have also considered that subjective weighting is more arbitrary and less reliable for decision making, and adopted objective weighting method for indicators. Ma Zhonggui et al. (2006) [7] used grey correlation analysis to determine the weights of evaluation factors to study the problem of extracting useful information from Internet information. Cui Jie et al. (2008) [8] improved the subjective and objective integration method based on grey correlation to solve the weights of indicators, which made the importance of each indicator differ significantly. Yi Ming and Zhang Tingting (2019) [9] used rough sets to approximate and assign weights to indicators, and then solved the weighted grey correlation to quantify the similarity between alternative answers and ideal answers. This shows that grey correlation analysis can be used not only to determine the weight of indicators, but also for comprehensive evaluation. The use of grey correlation analysis for comprehensive evaluation better handles the information incompleteness in measuring the merits of solutions, and provides a powerful means of parsing out the ranking of solutions [10]. If the method is used to assign indicators to systems containing qualitative indicators, it is not easy to reduce the error and influence caused by subjective factors in the judging process. Therefore, this paper proposes to choose the grey correlation analysis method to conduct a comprehensive assessment of alternative answers.

For the characteristics of answer quality assessment in Q&A communities, the qualitative index values of answers need to be scientifically and rationally transformed into quantitative ones. Xu Yong et al. (2016) [11] proposed to establish a fuzzy comprehensive evaluation matrix to transform qualitative evaluation into quantitative evaluation, assigning weighted indicators based on the number of times the relevant indicator set was mentioned, and then conducting a comprehensive evaluation of Taobao product text reviews. The hesitant fuzzy information evaluation matrix provides an effective means to quantify qualitative indicator values, which can better reflect the judgmental disagreement existing among evaluators; however, ignoring the homogeneity among indicators in the assignment process may lead to inaccurate ranking results. Therefore, in this paper, we propose to use the hesitant fuzzy information evaluation matrix to quantify the qualitative indicators, adopt the entropy weighting method to objectively assign weights to the indicators, and then revise the initial weights based on the correlation measure among the indicators, based on which we use the weighted grey correlation analysis to rank the alternative answers.

The rest of our paper is structured as follows: section 2 reviews the related work; section 3 focuses on the model presented in this paper; section 4 presents the description and analysis of the experimental results and also the comparative experiments performed; section 5 discusses the advantages and limitations of the model summarizes and looks ahead.

2. Traditional Grey Correlation Analysis

Deng Julong (1987) [12] proposed a grey correlation analysis based on the geometric relationship between sequences to measure the strength of the relationship between things. After that, scholars at home and abroad have continuously improved the grey system doctrine and applied it to audit and evaluation, multi-attribute decision making and inference prediction. Its basic calculation process is as follows [12].

Let the set of scenarios be $P=\{P_1,P_2,\cdots,P_m\}$ and the data series reflecting the changing characteristics of each scenario be $\{p_1(k)\},\{p_2(k)\},\cdots,\{p_m(k)\}$, $k=1,2,\cdots,z$ respectively. The correlation coefficient of scenario P_i to the reference scenario P_o at moment k is defined as

$$\xi[p_o(k),p_i(k)]=\frac{\Delta_{\min}+\rho\Delta_{\max}}{|p_o(k)-p_i(k)|+\rho\Delta_{\max}} \tag{1}$$

where $\Delta_{\max}=\max_i\max_k|p_o(k)-p_i(k)|$, $\Delta_{\min}=\min_i\min_k|p_o(k)-p_i(k)|$, $p_o(k)$ are the ideal values at moment k, ρ is the resolution factor, and $\rho\in[0,1]$.

The grey correlation between scenario P_i and the reference scenario P_o is

$$r(P_o,P_i)=\frac{1}{z}\sum_{k=1}^{z}\xi[p_o(k),p_i(k)] \tag{2}$$

Traditional grey correlation analysis can measure the closeness of the relationship between indicators and can also rank the solutions. Eq. (2) is the arithmetic mean of the grey correlation coefficients to obtain the grey correlation degree, which is slightly inadequate when dealing with decision problems containing multiple indicators with different weights. Therefore, the diversity between the qualitative assessment indicators of Q&A community answers will be addressed in this study, and differential weights will be assigned.

3. Assessing the Quality of Quizzes Based on Grey Fuzzy Evaluation Method

3.1. Establishing a Hesitant Fuzzy Information Evaluation Matrix

Traditional information evaluation matrices are only suitable for quantitative indicators that are easy to use with deterministic numerical measures, and are flawed when measuring the quality of questions and answers that contain multiple qualitative indicators. Fuzzy evaluation matrices are particularly advantageous when faced with problems where the value of qualitative indicators is heavily influenced by subjective perceptions, as they can translate individual subjective thinking into data form. In addition, since evaluators have different modes of thinking and it is not easy to evaluate each indicator of an alternative answer with exact values, Hesitant Fuzzy Set (HFS) is chosen as evaluation information in this paper.

Torra (2010) [13] proposed a set of possible values to quantitatively describe the hesitation problem encountered when making decisions, and the HFS was born.

3.1.1. Hesitation Fuzzy Set

Definition 1 [13]: let X be the given non-empty set that will

$$A=\{\langle x,h(x)\rangle|x\in X\}$$

is called the Hesitant Fuzzy Set and is a mapping of $X\to[0,1]$, where $h(x)$ is the Hesitant Fuzzy Element (HFE), denoting the possible affiliation of $x\in X$. The elements in $h(x)$ are α^k, $\alpha^k\in[0,1]$, $k=1,2,\cdots,l_h$, and l_h denotes the number of elements contained in the HFE h.

In general, the elements of a given hesitant fuzzy element are arranged in an unordered manner. In order to facilitate the decision calculation, the elements of the hesitant fuzzy element $h(x)$ shall be arranged in ascending order.

In addition, the elements of a given hesitant fuzzy element are arranged in an unordered manner. To facilitate the decision calculation, the elements α^k in the HFE $h(x)$ shall be arranged in ascending order.

3.1.2. Hesitation Fuzzy Element-Completion Strategy

Typically, the HFEs in Hesitant Fuzzy Sets E and F are of unequal length, i.e. $l[h_E(x_k)]\neq l[h_F(x_k)]$, such that $l_{x_k}=\max\{l[h_E(x_k)],l[h_F(x_k)]\}$. In order to enable the two HFSs to be characterized and compared, the set of fuzzy numbers with fewer fuzzy elements is filled so that it is the same length as the set with more fuzzy elements. The strategy of filling is influenced by the individual risk perception of the decision maker: risk averse people add the elemental minimum of the fuzzy element to the HFE with fewer elements, while risk preferences add the elemental maximum.

In recent years the theory has been extended and applied to a variety of fields such as multi-attribute decision making, risk assessment, and semantic computing.

3.2. Initial Configuration of Indicator Weights for HFSs Based on Fuzzy Cross-Entropy

There are various methods for assigning indicators to multi-indicator evaluation problems, which can be subjective, objective or combined to determine the indicator weights, but each method is applicable to different contexts. The HFS is to quantify the qualitative indicators, which is already subjective, so this paper mainly considers objective weighting. Liu Li et al. (2009) [14] determined the weight coefficients based on the entropy weight method, and assessed the water quality of the samples by an improved fuzzy comprehensive evaluation method. In the traditional fuzzy comprehensive evaluation method, most of the commonly used weight methods only consider the contribution made by a single indicator, without involving the connection between multiple things to be evaluated, and are unable to describe the impact of varying degrees of difference in the value of indicators on the weight distribution among indicators [14]. Since the entropy weight method can reflect the degree of effective information provided by each indicator in the problem, this paper will objectively assign weights to each indicator of the original evaluation value (i.e. Hesitant Fuzzy Set) of question and answer based on fuzzy cross-entropy, in order to preserve the large amount of information in the original indicators and avoid the loss of information caused by the conversion of indicators. The entropy measure is used to establish the initial weight of the indications based on the fuzzy properties of the original evaluation information.

Theorem 1 [15]: Let h be a HFE, then the fuzzy cross-entropy of h is

$$E(h) = 1 - C(h, h^c) \tag{3}$$

$$= 1 - \frac{2}{l_h T} \sum_{k=1}^{l_h} \left\{ \frac{(1+qh_{\sigma(k)}) \ln(1+qh_{\sigma(k)})}{2} + \frac{\left[1+q\left(1-h_{\sigma(l_h-k+1)}\right)\right] \ln\left[1+q\left(1-h_{\sigma(l_h-k+1)}\right)\right]}{2} - \frac{2+qh_{\sigma(k)}+q\left(1-h_{\sigma(l_h-k+1)}\right)}{2} \ln \frac{2+qh_{\sigma(k)}+q\left(1-h_{\sigma(l_h-k+1)}\right)}{2} \right\}$$

where $T = (1+q)\ln(1+q) - (2+q)[\ln(2+q) - \ln 2]$, $q > 0$. $h_{\sigma(k)}$ is the kth smallest number in h and $C(h, h^c)$ denotes the self-cross entropy of h. In general, indicators with small values of $E(h)$ should be assigned large weights; indicators with large values of $E(h)$ should be assigned small weights [15].

If there are m alternative answers $P_i (i = 1,2,\cdots,m)$, n indicators $G_j (j = 1,2,\cdots,n)$ with indicator weight vectors $\omega = (\omega_1, \omega_2, \cdots, \omega_n)^T$, $\omega_j \in [0,1]$, $j = 1,2,\cdots,n$ and $\sum_{j=1}^{n} \omega_j = 1$. Denote by h_{ij} the set of possible degrees h_{ij}^k of alternative answer P_i satisfying indicator G_j, and use fuzzy cross-entropy to determine the indicator weights, as shown in Equation (4).

$$\omega_j = \frac{1-E_j}{n-\sum_{j=1}^{n} E_j}, j = 1,2,\cdots,n \tag{4}$$

where $E_j = \frac{1}{m} \sum_{i=1}^{m} E(h_{ij})$. The fuzzy cross-entropy $E(h_{ij})$ of the HFE h_{ij} is calculated by Equation (3).

3.3. Correction of Indicator Weights Based on Grey Correlation Analysis

The previous step determines the indicator weights by fuzzy cross entropy, which effectively measures the importance of each indicator in the evaluation information with fuzzy characteristics. However, there may be overlapping associations between indicators in the assessment system, and the more similar the change trend of two indicator value series, the stronger the correlation between two indicators [16]. Fuzzy cross-entropy assignment does not take into account the overlapping information between indicators, resulting in less accurate assignment, while grey correlation analysis can measure the degree of correlation between indicators, so grey correlation analysis can be used to calculate the correlation coefficient between indicators and correct the initial weights to improve the accuracy of the weights. The specific steps are as follows.

3.3.1. Integration of Elements in HFEs Based on Geometric Score Functions

Definition 2: Let h_{ij} be a HFE, then the geometric score function of h_{ij} is

$$s(h_{ij}) = \sqrt[l_h]{\prod_{k=1}^{l_h} \alpha_{ij}^k} \tag{5}$$

where α_{ij}^k denotes the elements of the HFE h_{ij}.

The geometric score of each HFE was calculated according to Equation (5), which in turn yielded the geometric score matrix S, as shown in Equation (6).

$$S = \begin{bmatrix} s(h_{11}) & \dots & s(h_{1n}) \\ \vdots & \ddots & \vdots \\ s(h_{m1}) & \dots & s(h_{mn}) \end{bmatrix} \tag{6}$$

3.3.2. Measuring the Grey Correlation between Two Indicators

Based on the geometric score matrix S, the grey correlation r_{ij} between two indicators was calculated using Equations (1) and (2), which in turn led to the grey correlation matrix C_1 between indicators, as shown in Equation (7).

$$C_1 = (r_{ij})_{n \times n} = \begin{bmatrix} r_{11} & \dots & r_{1n} \\ & \ddots & \vdots \\ & & r_{nn} \end{bmatrix} \tag{7}$$

The greater the r_{ij}, the closer the two indicators' trends are, and the more comparable they are.

3.3.3. Standardisation of C_1 Based on Minimum Correlation

To increase the effective grey correlation, the grey correlation between the original indicators is normalised using the minimal correlation between indicators $\gamma = \min_{i,j}(r_{ij})$.

Definition 3: Let $C_1 = (r_{ij})_{n \times n}$ be the grey correlation matrix between indicators in hesitant fuzzy multi-attribute decision making, then call

$$C_2 = (d_{ij})_{n \times n} = \left(\gamma/r_{ij}\right)_{n \times n} = \begin{bmatrix} \gamma/r_{11} & \dots & \gamma/r_{1n} \\ & \ddots & \vdots \\ & & \gamma/r_{nn} \end{bmatrix} \tag{8}$$

is the effective grey correlation matrix between indicators, where γ/r_{ij} denotes the degree of independence between indicators. If $\gamma/r_{ij} = 1$, it means there is no correlation between two indicators; if $\gamma/r_{ij} < 1$, it means there is a correlation between two indicators, and the smaller γ/r_{ij}, the higher the correlation degree.

3.3.4. Adjusting the Weights Obtained Based on Fuzzy Cross-Entropy ω_j

The weights of the indicators obtained from the entropy measure are corrected using the effective grey correlation matrix C_2, as shown in Equation (9).

$$\omega_j{}' = \omega_j \times \prod_{i=1}^{j-1} d_{ij} \times \prod_{i=j+1}^{n} d_{ji} \tag{9}$$

The adjusted indicator weights are $\omega_j{}' = (\omega_1{}', \omega_2{}', \cdots, \omega_n{}')^T$. Since $\sum_{j=1}^{n} \omega_j{}' \neq 1$, they are readjusted as shown in Equation (10).

$$\omega_j{}'' = {\omega_j{}'}\Big/{\sum_{j=1}^{n} \omega_j{}'} \tag{10}$$

So the final adjusted vector of indicator weights is $\omega_j{}'' = (\omega_1{}'', \omega_2{}'', \cdots, \omega_n{}'')^T$.

Compared to the initial weight ω_j, the modified indicator weight $\omega_j{}''$ eliminates the correlation information between two indicators and effectively improves the accuracy of assignment.

3.4. Answer Quality Ranking Based on Improved Weighted Grey Correlation Expressed in HFSs

Due to the ambiguity and uncertainty between the quality of Q&A community answers and their evaluation factors in this study, as well as the incomplete and unbalanced evaluation of information by evaluators with varied information reserves and information demands, the grey correlation analysis method is chosen to evaluate the similarity between the answers to be evaluated and the best answers in the Q&A community. The method quantifies factors with unclear boundaries and not easy to quantify, and overcomes the shortcomings of traditional multiple correlation analysis and multiple regression analysis [17]. The specific steps are as follows.

3.4.1. Positive Processing of Decision Information

Each indication has a different polarity; some are benefit indicators, while others are cost indicators. For the original hesitation fuzzy information matrix $A = (\alpha_{ij})_{m\times n}$, the expanded hesitation fuzzy information evaluation matrix $B = (\beta_{ij})_{m\times n}$ is obtained after completing the HFEs. The standard unification of B is performed by the normalization algorithm [18] to obtain the candidate answer evaluation square $H = (h_{ij})_{m\times n}$ with consistent vector directionality, i.e. the forward processing information set.

When the indicator is of the benefit type, it is processed according to Equation (11).

$$h_{ij}^k = \frac{\beta_{ij}^k - \min_i \beta_{ij}^k}{\max_i \beta_{ij}^k - \min_i \beta_{ij}^k} \tag{11}$$

When the indicator is of the cost type, it is treated according to Equation (12).

$$h_{ij}^k = \frac{\max_i \beta_{ij}^k - \beta_{ij}^k}{\max_i \beta_{ij}^k - \min_i \beta_{ij}^k} \tag{12}$$

where h_{ij}^k denotes the elements of the normalised HFE h_{ij} of the answer P_i under the indicator G_j.

3.4.2. Identifying the Ideal Answer P_o

The ideal solution $P_o = \{h_{o1}, h_{o2}, \cdots, h_{oz}\}$ is derived for the hesitant fuzzy information evaluation matrix $H = (h_{ij})_{m\times n}$ following complementation and normalisation, where $z = \sum_{j=1}^{n} l_{h.j}$, $h_{ot} = \max_i h_{ij}^k$, $k = 1,2,\cdots,l_{h.j}$, $t = 1,2\cdots,z$ and $l_{h.j}$ denotes the length of the HFE $h_{.j}$ under the indicator G_j.

3.4.3. Building a Sequence of Differences between the Ideal Answer P_o and the Alternative Answer P_i Based on the HFSs of H

Definition 4: Let $P_o = (h_{o1}, h_{o2}, \cdots, h_{on})$ be the ideal answer, and $P_i = (h_{i1}, h_{i2}, \cdots, h_{in})$ be the alternative answer, and $h = (\alpha_1, \alpha_2, \cdots, \alpha_l)$ be the inscription of a HFE containing l possible affiliations. To subtract the HFE h_{o1} from h_{i1}.

$$\Delta_{oi1} = h_{o1} - h_{i1} = \left(\alpha_{o1}^1 - \alpha_{i1}^1, \alpha_{o1}^2 - \alpha_{i1}^2, \cdots, \alpha_{o1}^{l_1} - \alpha_{i1}^{l_1}\right) \tag{13}$$

where l_1 denotes the length of h_{o1} (or h_{i1}). In this way $\Delta_{oi2}, \Delta_{oi3}, \cdots, \Delta_{oin}$ is computed and the final result of the subtraction operation of the two HFEs is obtained.

3.4.4. Measuring the Grey Correlation between Alternative Answers and the Ideal Answer and Ranking the Answers

First, the correlation coefficient of elements of HFEs in $H = (h_{ij})_{m\times n}$ is calculated using Equation (1), where z denotes the total number of elements of all HFEs in the same answer sequence in H, i.e. $z = \sum_{j=1}^{n} l_{h.j}$.

Secondly, the correlation coefficient of all elements in each HFE in $H = (h_{ij})_{m\times n}$ is averaged to obtain the correlation coefficient of each HFE η_{ij}.

$$\eta_{ij} = \sum_{k=\lambda}^{\left(\lambda + l_{h_{ij}} - 1\right)} \frac{\xi[p_O(k), p_i(k)]}{l_{h_{ij}}} \tag{14}$$

where $\lambda = 1 + \sum_{\alpha=2}^{j} l_{h_{i(\alpha-1)}}$, $i = 1,2,\cdots,m$, $j = 1,2,\cdots,n$. When $j = 1$, $\lambda = 1$.

Next, the grey correlation coefficient matrix N is formed from the correlation coefficients η_{ij} between each alternative answer and the ideal answer, as shown in Equation (15).

$$N = (\eta_{ij})_{m\times n} = \begin{bmatrix} \eta_{11} & \cdots & \eta_{1n} \\ \vdots & \ddots & \vdots \\ \eta_{m1} & \cdots & \eta_{mn} \end{bmatrix} \tag{15}$$

Then the grey correlation R_i between the alternative answer and the ideal answer is

$$R_i = \sum_{j=1}^{n} \eta_{ij} \omega_j{}'' \quad i = 1,2,\cdots,m \tag{16}$$

where $\omega_j{}''$ represents the final indicator weights adjusted by Equations (9) and (10).

Finally, the R_i values are ranked, with the greater the grey correlation, the closer the answer is to the ideal answer.

4. Empirical Analysis

4.1. Case Sources and Selection of Indicators

4.1.1. Case Study Source

This study selects CSDN, the most representative technical Q&A community in China, which was founded in 1999 and is currently the largest domestic UGC (User Generated Content) content platform in the IT field, with 31 million registered members, of which the Q&A community is an important part [19].

A question and its answer data are randomly selected from the CSDN Q&A channel, and the retrieved data contains the question's full formulation, as well as the number of responses received and the content of those responses, as shown in Table 1.

Table 1. CSDN questions and their answers

Question	Question description	Answer symbol	Answer description
What is the best software to use to draw the structure of a neural network in a thesis? (Original Text: 论文中神经网络的结构图用什么软件绘制比较好？)	I have to make a structural diagram of GAN for my thesis, and I don't know what software or method I can use to make a structural diagram that meets the standards of the thesis. (Original Text:论文中要做一个 GAN 的结构示意图，不知道使用什么软件或是方法做出的结构图能够达到论文的标准。)	P_1	https://www.zhihu.com/question/20542936
		P_2	Try PowerPoint, it works well. If you're on a Mac, keynote is free. Google docs is also free if you have access to an extranet. The above three are basically popular drawing tools. Remember to export to vector PDF, then you can use it everywhere, including pdfLaTeX. If you want to niche a bit, GoogleNet that I am using pydot & Graphviz drawing, but these things need more hand tuning place. (Original Text: 试一下 PowerPoint，很好使的。如果你用 Mac 的话，keynote 免费。如果你能翻墙，Google docs 也免费。以上三个基本上是大众画图神器。记得导出成矢量 PDF，然后就到处可以用了，包括 pdfLaTeX。如果你想小众一点，GoogleNet 那个图我是用 pydot+Graphviz 画的，但是这些东西需要手调的地方多一些。)
		P_3	Visio is good, and of course the flowcharts in WPS are fine. (Original Text: Visio 很香，当然 WPS 里面的流程图也可以。)
		P_4	Visio is available and there is also software for online drawing. (Original Text: Visio 可以，也有在线绘图的软件吧。)

4.1.2. Selection of Indicators

Sussman and Siegal (2003) [20] concluded from their research on the information acceptance model that the quality of the information itself and the reliability of the information source are two factors affecting users' perceived value. Based on this, this paper considers the factors affecting the level of quality of Q&A information in two dimensions: answer information entropy (G_1), answer practicality (G_2), answer sentiment value (G_3), source reliability (G_4) and originality (G_5), as shown in Table 2. Note: In order to better reflect the subjective nature of user needs and the "fuzzy" nature of question and answer evaluation data, only five qualitative indicators are selected here, without considering the relatively simple quantitative indicators (e.g. number of answer likes, number of answer comments).

Table 2. Answer quality indicator system

Dimension	Indicator	Indicator symbol	Indicator meaning	Type of indicator
Answers	Answer information entropy	G_1	The amount of information contained in the text.	Benefit-based
	Answer practicality	G_2	The answers are implementable and useful.	Benefit-based
	Answer sentiment value	G_3	Messages with positive emotions help to strengthen the user's identification [21].	Benefit-based
Respondents	Source reliability	G_4	The honesty, competence, enthusiasm and objectivity of the respondent.	Benefit-based
	Originality	G_5	Answers are the product of independent thinking on the part of the respondent.	Benefit-based

4.2. Establishing a Hesitant Fuzzy Information Evaluation Matrix

4.2.1. Establishing the Original Hesitation Fuzzy Information Evaluation Matrix

For the qualitative evaluation indicators mentioned above, different evaluators have different focuses, and thus we choose to build the hesitation fuzzy number evaluation information matrix, as shown in Table 3.

Table 3. Hesitant fuzzy information evaluation matrix $A = (\alpha_{ij})_{4\times 5}$

	G_1	G_2	G_3	G_4	G_5
P_1	$\{0.5, 0.7, 0.8, 0.9\}$	$\{0.2, 0.5, 0.6\}$	$\{0.4, 0.5, 0.6\}$	$\{0.3, 0.5, 0.6, 0.8\}$	$\{0.3, 0.4, 0.6\}$
P_2	$\{0.2, 0.4, 0.6\}$	$\{0.2, 0.3, 0.5\}$	$\{0.3, 0.6, 0.7, 0.8, 0.9\}$	$\{0.4, 0.6, 0.7, 0.9\}$	$\{0.2, 0.5\}$
P_3	$\{0.1, 0.4, 0.5, 0.8\}$	$\{0.1, 0.4, 0.6, 0.7\}$	$\{0.4, 0.5, 0.7, 0.9\}$	$\{0.3, 0.5, 0.6, 0.9\}$	$\{0.2, 0.4, 0.5\}$
P_4	$\{0.3, 0.5, 0.6, 0.7\}$	$\{0.4, 0.6\}$	$\{0.3, 0.5, 0.6\}$	$\{0.2, 0.5, 0.6, 0.7, 0.9\}$	$\{0.4, 0.7\}$

Note: The elements of the HFE have been sorted in ascending order.

Answer P_1 has good answer information entropy, but its answer sentiment value is average. Answer P_2 has good source reliability, but assessors differ more on its answer sentiment value, and its answer utility and originality are average. Assessors disagree more on the answer information entropy and answer utility ratings for answer P_3, which has a good answer sentiment value. Answer P_4 has better originality and answer utility, but assessors differ more on its source reliability rating value.

4.2.2. Complementing and Normalising the HFEs in A

Platform decision makers tend to have a risk appetite, so the HFEs of varying length in the HFS are supplemented using fuzzy element maxima (Table 4). Considering that these indicators are benefit-based, there is no need to deal with the expanded evaluation matrix $B = (\beta_{ij})_{4\times 5}$, i.e. $H = (h_{ij})_{4\times 5} = B = (\beta_{ij})_{4\times 5}$.

Table 4. Extended hesitant fuzzy answer evaluation information matrix $B = (\beta_{ij})_{4\times 5}$

	G_1	G_2	G_3	G_4	G_5
P_1	$\{0.5, 0.7, 0.8, 0.9\}$	$\{0.2, 0.5, 0.6, 0.6\}$	$\{0.4, 0.5, 0.6, 0.6, 0.6\}$	$\{0.3, 0.5, 0.6, 0.8, 0.8\}$	$\{0.3, 0.4, 0.6\}$
P_2	$\{0.2, 0.4, 0.6, 0.6\}$	$\{0.2, 0.3, 0.5, 0.5\}$	$\{0.3, 0.6, 0.7, 0.8, 0.9\}$	$\{0.4, 0.6, 0.7, 0.9, 0.9\}$	$\{0.2, 0.5, 0.5\}$
P_3	$\{0.1, 0.4, 0.5, 0.8\}$	$\{0.1, 0.4, 0.6, 0.7\}$	$\{0.4, 0.5, 0.7, 0.9, 0.9\}$	$\{0.3, 0.5, 0.6, 0.9, 0.9\}$	$\{0.2, 0.4, 0.5\}$
P_4	$\{0.3, 0.5, 0.6, 0.7\}$	$\{0.4, 0.6, 0.6, 0.6\}$	$\{0.3, 0.5, 0.6, 0.6, 0.6\}$	$\{0.2, 0.5, 0.6, 0.7, 0.9\}$	$\{0.4, 0.7, 0.7\}$

4.3. Indicator Assignment

4.3.1. Assigning Indicators with Fuzzy Cross-Entropy

The fuzzy cross-entropy array Q for the fuzzy information set in the case of $q = 2$ is calculated and built using Equation (3), as shown in Table 5.

Table 5. Fuzzy cross-entropy matrix of the fuzzy information set $Q = (E_{h_{ij}})_{4\times 5}$

	G_1	G_2	G_3	G_4	G_5
P_1	0.798 7	0.975 3	0.988 5	0.952 9	0.980 7
P_2	0.980 1	0.936 5	0.890 2	0.833 4	0.941 1
P_3	0.989 8	0.978 9	0.868 1	0.911 3	0.928 3
P_4	0.995 2	0.980 9	0.984 6	0.972 5	0.942 2

The indicator weights based on the entropy measure are calculated according to Equation (4): $\omega = (0.202, 0.110, 0.230, 0.282, 0.177)^T$.

4.3.2. Adjusting Weights

(1) Integrating HFEs

For the information evaluation matrix A (Table 3), the geometric score $s(h_{ij})$ of h_{ij} is calculated using Equation (5), which in turn yields the geometric score matrix S.

$$S = [s(h_{ij})]_{4\times 5} = \begin{array}{c} \\ P_1 \\ P_2 \\ P_3 \\ P_4 \end{array} \begin{array}{c} \begin{array}{ccccc} G_1 & G_2 & G_3 & G_4 & G_5 \end{array} \\ \begin{bmatrix} 0.708\,5 & 0.391\,5 & 0.493\,2 & 0.518\,0 & 0.416\,0 \\ 0.363\,4 & 0.310\,7 & 0.618\,8 & 0.623\,6 & 0.316\,2 \\ 0.355\,7 & 0.360\,0 & 0.595\,8 & 0.533\,5 & 0.342\,0 \\ 0.501\,0 & 0.489\,9 & 0.448\,1 & 0.519\,4 & 0.529\,2 \end{bmatrix} \end{array}$$

(2) Calculating the Grey Correlation between Indicators

According to the research by Fan Kai and Wu Haoying (2002) [22], the corresponding $\rho = \lceil \sigma \rceil = \lceil 0.579 \rceil$ for the information evaluation matrix is calculated, so the discrimination coefficient $\rho = 0.6$ is used in this work. Using Equations (1) and (2) to calculate the correlation degree between two indicators, the grey correlation matrix between indicators $C_1 = (r_{ij})_{5\times 5}$ is obtained, as shown in Figure 1.

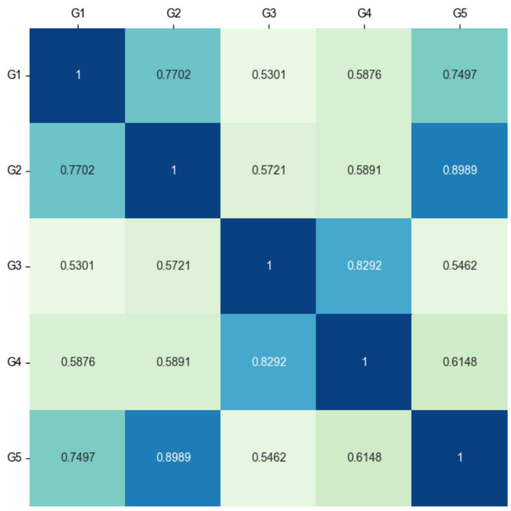

Figure 1. Grey correlation matrix between indicators C_1

(3) Standardising C_1

Since the minimum indicator correlation $\gamma = \min(r_{ij}) = r_{13} = 0.530\,1$, the normalised effective grey correlation matrix $C_2 = (d_{ij})_{5\times5} = \left({}^{\gamma}/{r_{ij}}\right)_{5\times5}$ is obtained, as shown in Figure 2.

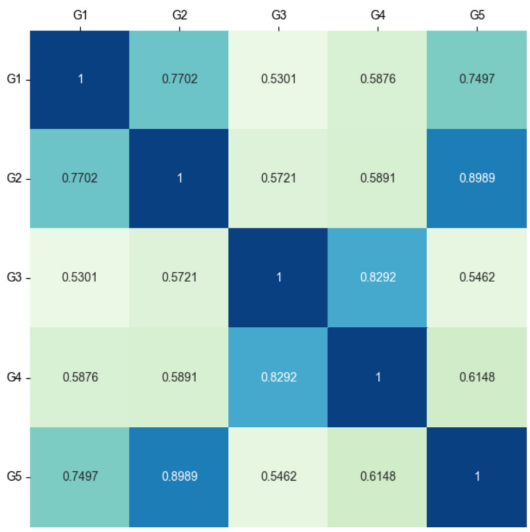

Figure 2. Effective grey correlation matrix C_2

(4) Adjusting the weights ω_j

Based on C_2, the corrected indicator weights ω_j'' are calculated using Equations (9) and (10).

$$\omega_j'' = (0.198\,8, 0.083\,4, 0.296\,4, 0.282\,9, 0.138\,5)^T$$

4.4. Answer Ranking Based on Grey Correlation Analysis

For $H = (h_{ij})_{4\times5} = B = (\beta_{ij})_{4\times5}$, the correlation coefficien η_{ij} for each h_{ij} is calculated using Equation (14) to obtain the grey correlation coefficient matrix N.

$$N = (\eta_{ij})_{4\times5} = \begin{array}{c} \\ P_1 \\ P_2 \\ P_3 \\ P_4 \end{array} \begin{array}{c} \begin{array}{ccccc} G_1 & G_2 & G_3 & G_4 & G_5 \end{array} \\ \begin{bmatrix} 1.000\,0 & 0.739\,3 & 0.660\,1 & 0.705\,9 & 0.618\,7 \\ 0.469\,7 & 0.560\,3 & 0.882\,4 & 1.000\,0 & 0.545\,5 \\ 0.492\,4 & 0.747\,5 & 0.941\,2 & 0.823\,5 & 0.511\,8 \\ 0.545\,5 & 0.926\,5 & 0.601\,3 & 0.700\,5 & 1.000\,0 \end{bmatrix} \end{array}$$

Based on N and ω_j'', R_i is calculated using Equation (16).

$$R_i = (0.741\,5, 0.760\,1, 0.743\,1, 0.700\,6)^T$$

Therefore, according to the value of R_i, the comprehensive ranking of the alternative answers is $P_2 > P_3 > P_1 > P_4$, and the best filtered answer is P_2.

4.5. Analysis of Answer Ranking Results

Firstly, we analyze the answer P_2, which is better than the other three alternatives in terms of textual information (G_1 answer information entropy), answer implementability (G_2 answer practicality). It is informative and shows the enthusiasm and honesty of the respondent (G_4 source reliability), and such an answer can also help to enhance the audience's sense of identity (G_3 answer sentiment value). P_2 is ranked highly and filtered as the relatively best answer.

Secondly, answer P_1, which presented only one external link, was analysed; formally, the external link contains the content of the answer and to some extent enhances the credibility of the answer. However, the website link requires the user to jump out of the content of that answer, which to some extent reduces the efficiency of users in accessing information. Therefore, with the revised weight ω_j'', P_1 does not have a dominant ranking.

Again, the analysis of the lower-ranked answer P_4 shows that the information entropy of P_4 is relatively low compared to that of P_1. Although the text of P_1 is not long, its text is a web link, and P_1 includes more information than P_4. Compared to P_3, answer P_4 provides a less clear solution to the question (recommendation of software for drawing structural diagrams) than P_3, mentioning only "software for online drawing" in general terms, and is therefore the least recommended answer P_4 in the overall assessment.

4.6. Sensitivity Analysis

4.6.1. Analysis of the Impact before and after the Correction of Indicator Weights

The grey correlation of each answer P_i with the ideal answer P_o and their ranking results before and after the adjustment of indicator weights are shown in Table 6.

Table 6. Grey correlation between each answer and the ideal answer before and after the weighting adjustment and the ranking of answers

Indicator weights	$R_{(P_1)}$	$R_{(P_2)}$	$R_{(P_3)}$	$R_{(P_4)}$	Ranking
Before adjustment	0.743 7	0.738 0	0.721 0	0.724 9	$P_1 > P_2 > P_4 > P_3$
After adjustment	0.741 5	0.760 1	0.743 1	0.700 6	$P_2 > P_3 > P_1 > P_4$

As can be seen from Table 6, answer P_2 takes the place of P_1 after the adjustment of weights and jumps to the first position, while answer P_4 is at the bottom. The correlation matrix C_1 between the indicators shows that answer practicality G_2 has a strong correlation with originality G_5 (0.898 9), and answer sentiment value G_3 has a low correlation with the rest of the indicators (all less than 0.580 0) except for a high correlation with source reliability G_4

(0.829 2), thus adjusting the weights to weaken the correlation between the indicators. Therefore, the final answer sentiment value G_3 weighting (0.296 4) is relatively high, and the answer practicality G_2 (0.083 4) and originality G_5 weightings (0.138 5) are relatively low, allowing the answer ranking to be adjusted and the adjusted ranking results to be more reasonable and accurate.

4.6.2. Analysis of the Effect of the HFE-Completion Strategy on the Results

When platform decision makers tend to be risk averse, HFEs of unequal length in the fuzzy set are complemented using the elemental minimum of the HFE, as shown in Table 7.

Table 7. Extended hesitant fuzzy answer evaluation information matrix $B' = \left(\beta_{ij}'\right)_{4\times 5}$

	G_1	G_2	G_3	G_4	G_5
P_1	$\{0.5, 0.7, 0.8, 0.9\}$	$\{0.2, 0.2, 0.5, 0.6\}$	$\{0.4, 0.4, 0.4, 0.5, 0.6\}$	$\{0.3, 0.3, 0.5, 0.6, 0.8\}$	$\{0.3, 0.4, 0.6\}$
P_2	$\{0.2, 0.2, 0.4, 0.6\}$	$\{0.2, 0.2, 0.3, 0.5\}$	$\{0.3, 0.6, 0.7, 0.8, 0.9\}$	$\{0.4, 0.4, 0.6, 0.7, 0.9\}$	$\{0.2, 0.2, 0.5\}$
P_3	$\{0.1, 0.4, 0.5, 0.8\}$	$\{0.1, 0.4, 0.6, 0.7\}$	$\{0.4, 0.4, 0.5, 0.7, 0.9\}$	$\{0.3, 0.3, 0.5, 0.6, 0.9\}$	$\{0.2, 0.4, 0.5\}$
P_4	$\{0.3, 0.5, 0.6, 0.7\}$	$\{0.4, 0.4, 0.4, 0.6\}$	$\{0.3, 0.3, 0.3, 0.5, 0.6\}$	$\{0.2, 0.5, 0.6, 0.7, 0.9\}$	$\{0.4, 0.4, 0.7\}$

Based on B', indicator weighting and answer ranking are performed and the results are shown in the third row of Table 8.

Table 8. Gray correlation of each answer with the ideal answer and answer ranking under different HFE completion strategies

Types of risk for platform decision makers	$R_{(P_1)}$	$R_{(P_2)}$	$R_{(P_3)}$	$R_{(P_4)}$	Ranking
Risk-loving	0.741 5	0.760 1	0.743 1	0.700 6	$P_2 > P_3 > P_1 > P_4$
Risk-averse	0.787 6	0.672 8	0.724 3	0.737 9	$P_1 > P_4 > P_3 > P_2$

As can be seen from Table 8, the ranking of each alternative answer varies considerably when different HFE-completion strategies are chosen. When the platform decision maker is risk-loving, P_2 is the relatively best answer. When the platform decision maker is risk-averse, answer P_1 is recommended, but P_2 is least recommended.

The impact of different HFE-completion strategies on decision making is summarised for alternative answers (i.e. comments) containing links to websites (two scenarios, i.e. unopened URL and opened URL). Firstly, changes in the completion strategies affect changes in the weights. Secondly, different types of information affect users' overall perceptions, and in the process of judging information quality, users consider textual information more important than hyperlink information [23]. Again, for decision makers using a risk-loving completion strategy, it is recommended that decision makers choose answers with detailed textual content rather than open URLs. Because URL skipping reduces the user's efficiency in accessing information to a certain extent, risk-loving decision makers may not click on the link to save time costs (risk not understanding the content behind the URL) and choose P_2 with richer text content and moderate text length (which can be regarded as having high sentiment value G_3 and reliability G_4). For decision makers with a risk-averse strategy, it is recommended that decision makers choose answers that contain external links. Although it is more time consuming to jump links, the risk-averse decision maker clicks on the link to be on the safe side, i.e. chooses P_1 with the possibility of multiple solutions due to the page jump (which can be considered as a higher information entropy G_1).

In summary, the use of modified weights allows platform decision-makers with different risk types to select the answer they want (if the decision-maker tends to be risk-loving, it is recommended that they select the answer that is text-rich and does not require opening external links; if the decision-maker tends to be risk-averse, it is recommended that they select the answer that contains external links), which also demonstrates the reasonableness of the weighting adjustment and the effectiveness of the evaluation method.

5. Conclusion

This study addresses the problem of varying quality of information in Q&A communities. Considering the "fuzziness" in the quantification of the answers' qualitative indicators, a hesitant fuzzy information evaluation matrix was established to reasonably quantify the qualitative indicators and effectively measure the differences in assessment among evaluators. As the evaluation matrix was already subjective, fuzzy cross-entropy was used to objectively assign weight to the indicators, and the grey correlation between indicators was measured to correct the weights, taking into account the possible overlapping information between indicators. Finally, based on the grey fuzzy evaluation method, the weighted grey correlation of the answers was obtained and the answers were ranked accordingly. Selected examples of questions and answers from CSDN were analysed, and the results show that the grey fuzzy evaluation method quantifies the qualitative indicators and can process and analyse indicators that reflect the subjective nature of user needs. By modifying the indicator weights, the homogeneous information between indicators was effectively weakened, making the final answer evaluation ranking more realistic. A sensitivity analysis was also conducted for the changes in the HFE-completion strategy to explore the rationality of the recommended answers. However, this paper only considers qualitative indicators when ranking answers, and will next consider how to combine qualitative indicators with quantitative indicators to evaluate the quality of Q&A community answers.

References

[1] CNNIC (2022). The 49th statistical report on the development of the Internet in China. Retrieved from http://www.cnnic.net.cn/hlwfzyj/hlwxzbg/.

[2] Wang X Y, Wang C Y, Mu J (2022). Research on the effectiveness of seller reputation in platform—mediated network market: the effect of information quality，information source credibility and assurance seal sensitivity. Journal of Dalian University of Technology (Social Sciences), 43(1), 28-37.

[3] Cheng Q, Wang H X, Pian W J (2020). Research on information quality comprehensive evaluation of medical health websites based on evidence reasoning. Digital Library Forum, (4), 53-59.

[4] Li J, Zheng A H, Li G Y, et al. (2021). Performance evaluation method of provincial administrative information resource sharing platforms based on big data. Journal of University of Jinan (Science and Technology), 35(4), 381-386.

[5] Chen Y, Deng Z H (2019). Study on the assessment of comprehensive influence of health care websites in China. Journal of Medical Informatics, 40(6), 49-53+72.

[6] Guo S L, Zhang X X, Li Z M (2017). A study of C2C platform product information credibility based on the grey relation grade. Information Science, 35(7), 142-148.

[7] Ma Z G, Ye B, Zeng G P, et al. (2006). Study on individual preference model in mutual-move intelligent communication. Application Research of Computers, (8), 55-57.

[8] Cui J, Dang Y G, Liu S F (2008). An improved approach for determining weights of attributes in decision making based on grey incidence. Chinese Journal of Management Science, 16(5), 141-145.

[9] Yi M, Zhang T T (2019). Ranking answer quality of popular Q&A community Data Analysis and Knowledge Discovery, 3(6), 12-20.

[10] Hao Q, Zhu M L, Li B (1996). The ranking method and applications of the fuzzy and grey degree of correlation of schemes. Systems Engineering — Theory & Practice, (6), 44-49.

[11] Xu Y, Zhang H, Chen L (2016). A fuzzy comprehensive evaluation method for UGC based on sentiment analysis: taking Taobao product text review UGC as an example. Information Studies:Theory & Application, 39(6), 64-69.

[12] Deng J L (1987). The primary methods of grey system theory. Wuhan, China: Huazhong University of Science and Technology Press.

[13] Vicene T (2010). Hesitant fuzzy sets. International Journal of Intelligent Systems, 25(6), 529-539. doi: 10.1002/int.20418.

[14] Liu L, Zhou J Z, Yang J J, et al. (2009). Improved fuzzy comprehensive assessment method based on information entropy. Computer Engineering, 35(18), 4-6.

[15] Xu Z S, Xia M M (2012). Hesitant fuzzy entropy and cross-entropy and their use in multiattribute decision-making. International Journal of Intelligent Systems, 27(9), 799-822.

[16] Wang H Y, Meng X L, Yu R, et al. (2017). Research on order parameter identification of the food quality chain coordination system. Systems Engineering — Theory & Practice, 37(7), 1741-1751.

[17] Yu T, Huang T, Pan Y X, et al. (2013). Comprehensive urban waste disposal forecasting mode based on the BP artificial neural network and grey relational degree. Journal of Safety and Environment, 13(4), 94-97.

[18] Yu L, Fang Z G, Wu L F, et al. (2014). Maximum entropy configuration model of objective index weight based on grey category characteristics difference. Systems Engineering — Theory & Practice, 3(8), 2065-2070.

[19] CSDN (2021). 2021 CSDN network overview. Retrieved from https://www.csdn.net/company/index.html#about/.

[20] Sussman S W, Siegal W S (2003). Informational influence in organizations: an integrated approach to knowledge adoption. Information Systems Research, 14(1), 47-65.

[21] Joyce E, Kraut R E (2017). Predicting continued participation in newsgroups. Journal of Computer-Mediated Communication, 11(3), 723-747.

[22] Fan K, Wu H Y (2002). A new method on identification conficient of relational grade for gray system. Journal of Wuhan University of Technology, 24(7), 86-88.

[23] Hu Y, Wei X Y, Wang C (2017). Research on the construction of micro-blog information quality evaluation indicator system. Information Science, 35(6), 44-50.