Journal of Global Information Management
Volume 32 • Issue 1

A Large Group Emergency Decision-Making Approach on HFLTS With Public Preference Data Mining

Mengke Zhao, School of Economics and Management, Shanghai Maritime University, China
Ji Guo, School of Economics and Management, Shanghai Maritime University, China
Xianhua Wu, School of Economics and Management, Shanghai Maritime University, China*

ABSTRACT

Aiming at the emergency decision-making problem of major emergencies, this article proposes a large group emergency decision-making (LGEDM) approach with public opinions mining on hesitation fuzzy language term set (HFLTS). First, extract keywords that represent general preferences on events from the Weibo platform, classify keywords using the word similarity-based keyword clustering algorithm and identify decision attributes and their weights. Next, define the similarity measure and hesitation fuzzy entropy measure of HFLTS, quantify the decision risk of experts using the risk measurement model, and cluster all experts into several subgroups using the risk metric-based group clustering algorithm. Subsequently, assign clusters' weights on their risk value and size and obtain each cluster's preference matrix by the HIOWA operator. Finally, derive the ranking results of alternatives using the sorting process, and an example of "COVID-19" is presented to verify the rationality and effectiveness of the proposed method.

KEYWORDS

Data Mining, Hesitant Fuzzy Linguistic Term Sets, K-means, Large Group Emergency Decision Making, Public Preference, Risk, Term Frequency-Inverse Word Frequency, Word Similarity

INTRODUCTION

In recent years, major emergencies occurred more and more frequently in China, the types and frequency of disasters have increased obviously, and the scope of involvement has expanded markedly, such as "the explosion accident in Tianjin Binhai New Area on August 12, 2015," "the heavy rainstorm in Zhengzhou on July 20, 2021," and "the Corona Virus Disease on December, 2019," etc. The high complexity and destructive nature of major emergencies have a huge negative impact on the social order, economic development, and people's lives and properties in China (Guo et al., 2020; Tan et al., 2021; Wu et al., 2021b). The primary task after a major emergency is to organize experts analyze and judge the current state of affairs, strive to make scientific and efficient decisions quickly in a short period, control the development of events as possible and reduce the loss or consumption of resources

DOI: 10.4018/JGIM.337610
*Corresponding Author

caused or likely to be caused due to the emergency. Therefore, there is an urgent need for the support of group decision-making methods (Cao et al., 2022; Wu et al., 2020; Wu et al., 2021a). The complexity and variability of mega-emergency events determine that emergency decision-making requires the participation of multiple experts from different fields, which makes the emergency decision-making characterized by complex large-group decision-making, and the traditional group decision-making methods are no longer applicable to such decision-making problems. Various LGEDM methods have been proposed successively (Xu et al., 2015; Xu et al., 2017). Compared with traditional group decision problems, the LGEDM usually involve more than 20 decision makers with the characteristics of time constraint, decision environment uncertainty, decision information limitation, and the possibility of catastrophic loss due to decision errors (Tang et al., 2020; Xing et al., 2022; Wu et al., 2022).

With the development of information technology, the Internet has become a huge platform for the expression of public opinion, and events are fermented very fast on the Internet. Once the emergency breaks out, hundreds of millions of netizens gather and participate in it rapidly, posting related comments on social media platforms, thus generating massive text data. This textual information not only reflects the public's concern about the emergencies but also provides important references for emergency decision-makers. Therefore, in the process of emergency decision-making, how to effectively utilize the public views in social media and express and integrate the decision information of large groups efficiently under urgent time pressure, is one of the pressing problems in the field of emergency decision-making. At present, the research of the LGEDM method under a big data environment has shown initial results in the field of decision making (Xu et al., 2019b; Xu et al., 2019c). In addition, the complexity, dynamics, and urgency of LGEDM problems also imply that more uncertainties are bound to appear in the decision-making process, which we call risk. If these risks cannot be controlled within the effective range, they will become new risk sources, which will further deteriorate the situation and lead to low quality or failure of decision-making. Therefore, introducing decision risk into LGEDM is one of the main methods for ensuring decision reliability, which has attracted a worldwide interest (Yin et al., 2021; Zhong et al., 2020). But in general, the related research is still in the initial stage.

Clustering is a fundamental tool for analyzing and managing large-scale comments, and it is also a vital tool for solving LGEDM problems. Clustering refers to the dimensionality reduction of large-scale decision-makers according to some clustering principles or algorithms for analysis easily (Tang et al., 2019; Xu et al., 2019a, Xiong et al., 2022). Many different clustering algorithms have been proposed in previous literature to achieve clustering for group decision-making (Yu et al., 2020; Zhong et al., 2021; Chen et al., 2018). K-means is a classical distance-based unsupervised clustering algorithm that uses distance as the evaluation index of similarity, that is, the closer the distance between two objects is considered, the greater their similarity. The algorithm has a number of advantages: 1. simple algorithm idea and fast convergence; 2. low time complexity and high efficiency; 3. better clustering effect when clustering large-scale data. The algorithm has been widely used in large group decision problems. (Chao et al., 2020; Liu et al., 2021). Therefore, the k-means algorithm is chosen as the clustering method for this paper. The clustering in the large group decision-making process is usually implemented based on the evaluation information given by the decision-makers (Gou et al., 2018; Mandal et al., 2021). Therefore, choosing a suitable way to express the decision maker's evaluation information is a key issue for the LGEDM problem. With the inspiration of hesitant fuzzy sets (Torra, 2010) and linguistic fuzzy sets (Zadeh, 1975), the Hesitant Fuzzy Linguistic Term Sets (HFLTS) is proposed. HFLTS allows experts to use several linguistic terms evaluating a linguistic variable and has many advantages in describing experts' perceptions and preferences. Since its proposal, the HFLTS has been concerned by many scholars (Farhadinia & Xu, 2018; Wu & Zhang, 2021). With the increasing maturity of hesitant fuzzy language theory and methods and their special advantages in describing decision-maker preferences in complex decision problems, more and more scholars have applied them to large group decision problems in recent years (Rodríguez et al., 2021; Zhong & Xu, 2020).

This paper focuses on the complex LGEDM problem in the social media big data environment, combines data science, risk theory, and group decision theory, and propose an LGEDM approach with public preference attribute mining based on HFLTS, the method is verified by the case of "novel coronavirus pneumonia (COVID-19)".

METHODOLOGY

Hesitant Fuzzy Linguistic Term Sets

The concept of HFLTS was introduced to model experts' hesitation in qualitative contexts (Rodríguez et al., 2012). In the following, some basic concepts and operational laws related to HFLTS are described.

Definition 1. (Xu, 2004) Let $S=\left\{s_\mu \mid \mu=0,1,\ldots,\tau,\tau\in N\right\}$ be a LTS with an odd granularity $\tau+1$, A HFLTS H_S is an order finite subset of the consecutive linguistic terms of S, noted as:

$$H_S=\left\{s_i,s_{i+1},\ldots,s_j \mid s_\mu\in S,k=i,i+1,\ldots,j\right\} \tag{1}$$

To preserve all the given information, the discrete term sets should be extended to a continuous LTS $\bar{S}=\left\{s_\mu \mid \mu\in\left\{0,q\right\}\right\},q(q>\tau)$ S, where $q(q>\tau)$ is a sufficiently large positive integer. If $s_\mu\in S$, then s_μ is termed an original linguistic term, otherwise, s_μ is termed a virtual linguistic term. In general, the expert uses the original linguistic term to express opinions, and the virtual linguistic terms can only appear in calculation.

One significant issue in the process of the GDM is to integrate of individual opinions. Here, an aggregation operator of HFLTS and some basic properties are presented.

Definition 2: (Yager & Filev, 1999) Let $\bar{S}=\left\{s_\mu \mid \mu\in\left\{0,q\right\}\right\}$ be a set of extended LTSs related to $S=\left\{S_\mu \mid \mu=0,1,2,\ldots,\tau\right\}$, and $H_l=\left\{s_{\mu_1},s_{\mu_2},\ldots,s_{\mu_{g^l}}\right\},\left(l=1,2,\ldots,\mathrm{n}\right)$ be n HFLTSs on $\bar{S}$. Then, the hesitant induced ordered weighted averaging (HIOWA) operator is a function that is expressed as follows:

$$HIOWA:\left(H_1\times H_2\times\ldots\times H_n\right)\rightarrow H_C$$

$$HIOWA:\left(\lambda_1,H_1,\lambda_2,H_2,\ldots,\lambda_n,H_n\right)=\Lambda_\tau\left(\sum_{l=1}^{n}\omega_l\cdot\Lambda_\tau^{-1}\left(H_{o(l)}\right)\right) \tag{2}$$

Where, $\Lambda_\tau^{-1}\left(s_\mu\right)=\mu,\Lambda_\tau\left(\mu\right)=s_\mu,H_{o(l)}$ stands for the l^{th} largest in $H_l,\left(l=1,2,\ldots,n\right)$, the ordered inducing variable related to $H_l,\left(l=1,2,\ldots,n\right)$ is denoted as ω_l, $\left(l=1,2,\ldots,n\right)$.

Term Frequency-Inverse Word Frequency (TF-IWF)

TF-IWF is an improved algorithm proposed on basis of the traditional TF-IDF (Term Frequency - Inverse Document Frequency). TF-IDF is a classical algorithm to calculate document feature weights (Sandiwarno, 2020; Tian & Wu, 2018). As known, IDF is unable to reflect the importance and distribution of feature words for its simple structure, which makes it difficult to adjust the TF weights better. Therefore, an improved weighting algorithm TF-IWF is proposed by researchers (Huang et al., 2011; Lu & Li, 2013).

On the one hand, let $n_{i,j}$ denote the frequency of word t_i in text j and $\sum_k n_{k,j}$ denote the sum of all words in document j. Then, the TF value of word t_i in text j is:

$$tf_{ij} = \frac{n_{i,j}}{\sum_k n_{k,j}} \tag{3}$$

On the other hand, let $\sum_{i=1}^{m} nt_i$ denote a sum of the frequencies of all words in the corpus, and nt_i denote the total frequency of word t_i in the corpus. Then, the IWF value of word t_i with respect to the document set or corpus is:

$$iwf_i = log \frac{\sum_{i=1}^{m} nt_i}{nt_i} \tag{4}$$

Then, the TF-IWF value of word t_i corresponding to document j is:

$$TF - IWF_{i,j} \rightarrow tf_{ij} \times iwf_i = \frac{n_{i,j}}{\sum_k n_{k,j}} \times log \frac{\sum_{i=1}^{m} nt_i}{nt_i} \tag{5}$$

The TF-IWF algorithm reduces the impact on word weights of similar texts from the document set/corpus and more accurately expresses the importance of words in documents under investigation. In addition, the weights obtained by the traditional TF-IDF are generally small, even close to 0, and not very accurate. However, the TF-IWF can solve this problem precisely.

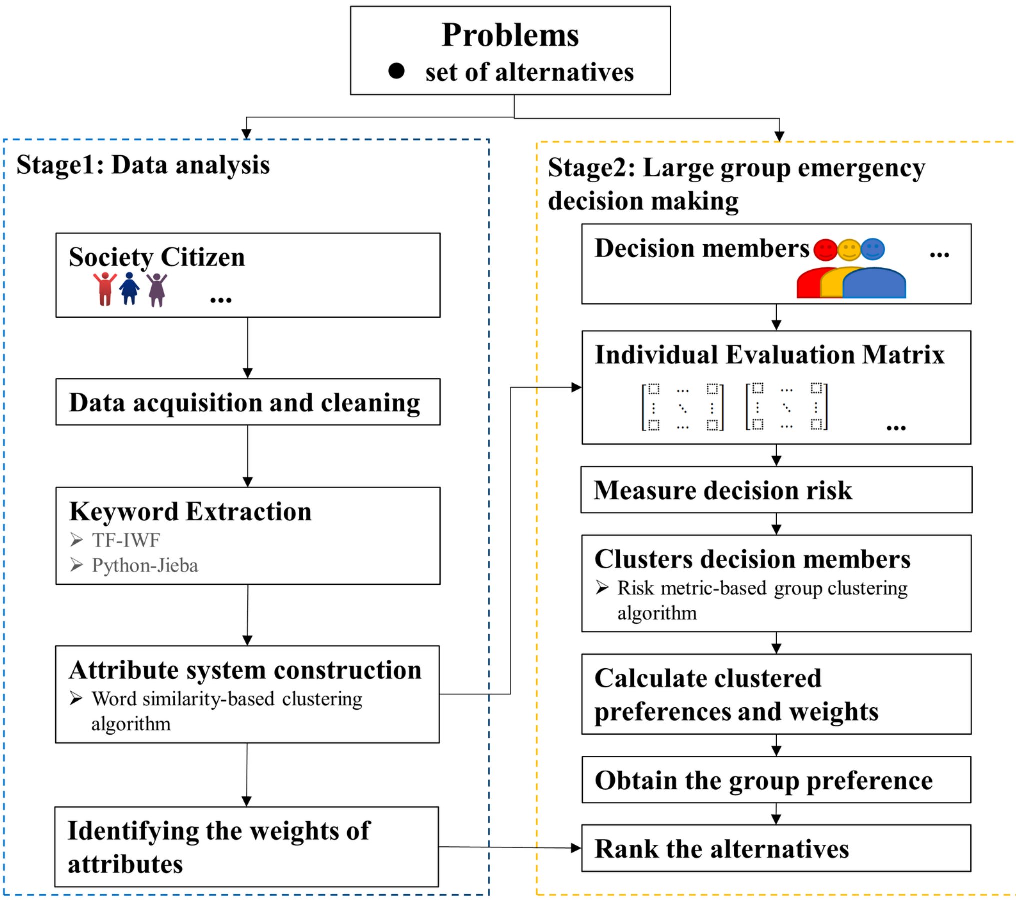
An LGEDM Approach with Public Preference Data Mining on HFLTS

Let $X = \left\{x_{mn} \mid m = 1,2,\ldots,M\right\}$ be the considered alternatives and $E = \left\{e_d \mid d = 1,2,\ldots,D\right\}$ be a set of decision makers. $C = \left\{c_n \mid n = 1,2,\ldots,N\right\}$ is a set of attributes and the weight vector of these attributes is denoted as $U = \left\{u_1, u_2, \ldots, u_N\right\}$ where $\sum_{n=1}^{N} u_n$ and $u_n \geq 0$. $R^d = \left\{r_{mn}^d\right\}_{MN}$ is the decision matrix provided by e^d, where r_{mn}^d is an HFLTS representing the evaluations of x_m with respect to attribute c_n. In this section, a novel LGEDM approach with public preference data mining on HFLTS is introduced. The detailed method flow is shown in Figure 1.

Public Preference-Based Attribute Mining

In this section, a method for mining public preferences and constructing the decision attribute system based on social media data is proposed, and the Sina Weibo is chosen as the analysis base and data collection channel (Chen et al., 2014). The data processing in this part is implemented by Python 3.9, including data collection, data cleaning, word separation, and keyword extraction. The detailed steps are as follows.

Figure 1. The framework of an LGEDM approach with public preference data mining on HFLTS



Step1: Data acquisition and cleaning. The relevant comments of Sina Weibo in a specific period are seized by Web Scraper. And the captured text content is cleaned through Python-based natural language processing, which included removing useless content such as subject tags, @, abnormal values, time fields, URL, etc.

Step2: Keyword extraction. The cleaned text content is cut using the Python-jieba word splitting package to extract keywords, exclude invalid keywords, such as deactivated words, etc. Then the TF-IWF algorithm is used to sort the keywords after text splitting, and the keywords with greater weight are selected as valid keywords.

Step3: Attribute system construction. Normally, the number of valid keywords is set to hundreds or thousands, depending on the event context, data size, and expert experience. Therefore, it is important to integrate these huge numbers of keywords quickly and effectively under the urgent time pressure of LGEDM.

The semantic similarity calculation of words is a method to numerically measure the degree of semantic similarity between two words based on certain calculation methods and various linguistic resources by computer. K-means algorithm is a distance-based clustering algorithm that combines simplicity and classicality, using distance as an evaluation index of similarity. Therefore, it is feasible and meaningful to use the semantic similarity of words as the index of distance measurement between words and combine it with the traditional k-means algorithm to realize keyword clustering.

In summary, a new keyword clustering algorithm is proposed in this section by combining the traditional k-means algorithm (Hartigan & Wong, 1979) and the word similarity algorithm (Chen et al., 2021), which is implemented by using Python3.9 programming. The detailed steps are described as follows.

Phase One: Identify the initial clusters. Assume the number of keywords is L, denoted as: $W^y, y = 1,2,\ldots,L$. Let t=0. Firstly, the K words is selected randomly as the initial clustering centers, denoted as $G_k^t, k = 1,2,\ldots,K$. The number of K values is determined by the management experience of experts and the number of data points. Subsequently, the similarity between the rest of the keywords and each clustering center is calculated separately, denoted as $SD_{y,k}^t, k = 1,2,\ldots,K$; finally, the keyword W^y is assigned to the category k represented by the clustering center with the highest similarity $SD_{y,k}^t$ with it, until all the keywords are finished being classified.

Phase two: Update clusters. Firstly, the average similarity between each word and every cluster separately is recalculated based on the initial clustering results, expressed as:

$$SD_{y,k}^{t+1} = \sum_{n_k^t=1}^{N_k^t} SD\left(W^y, W_{n_k^t}^t\right), y = 1,2,\ldots,L; k = 1,2,\ldots,K. \tag{6}$$

where N_k^t denotes the number of words contained in the kth cluster at stage t.

Similarly, the maximum similarity $SD_{y,k}^{t+1}$ is selected and the words W^y is reassigned to the corresponding category k. The k-means algorithm is used to update the clustering in this process until the stopping condition is reached.

Phase three: stopping condition. The maximum number of iterations Q and the threshold $\sigma \to 0\left(\sigma > 0\right)$ are set. Let $SD_{t,t+1}$ presents the difference of the average similarity of all keywords to each cluster at stage t and stage t+1, denoted as:

$$SD_{t,t+1} = \frac{1}{LK}\sum_{y=1}^{L}\sum_{k=1}^{K}\left|SD_{y,k}^{t+1} - SD_{y,k}^{t}\right| \tag{7}$$

When $SD_{t,t+1} < \sigma$ or $t+1 > Q$, the clustering iteration stopped; otherwise, returned to step2. Details of the clustering algorithm were provided in Table 1.

Output results, each cluster represents an attribute, and the corresponding attribute set is represented as: $C = \left\{c_n \mid n = 1,2,\ldots,K\right\}$.

Step4: Identifying the weights of attributes. The TF-IWF values corresponding to all keywords under each attribute are accumulated, expressed as:

$$TF - IWF^n = \sum_{h^n=1}^{H^n} tf - iwf_{h^n}, n = 1,2,\ldots,N. \tag{8}$$

where, H^n denotes the number of keywords under the nth attribute.

Table 1. Word similarity-based keyword clustering algorithm (Algorithm 1)

Input: $W^{y}, y = 1,2,\ldots,L$ (Keywords)
K (The number of categories)
HowNet, CiLin (knowledge base)
σ (The threshold value)

Output: Clusters. $\widetilde{G_1}, \widetilde{G_2}, \ldots, \widetilde{G_K}$

Step1: Let t=0, choose K initial clustering centers from $W^{y}, y = 1,2,\ldots,L$, noted as $G_k^t, k = 1,2,\ldots,K.$

Step2: For every $W^{y}, y = 1,2,\ldots,L, y \notin G_k^t$, apply word semantic similarity in [2] to compute $SD_{y,k}^t, k = 1,2,\ldots,K$.

Classify W^{y} into the category of G_k^t with the highest similarity degree $SD_{y,k}^t$ among them.

Step3: Let t=t+1, recomputing the mean similarity between words and each category separately by Eq. (6), $SD_{y,k}^{t+1}, y = 1,2,\ldots,L; k = 1,2,\ldots,K.$

Re-classify W^{y} into the category of G_k^{t+1} with the highest similarity degree $SD_{y,k}^{t+1}$ among them.

Step4: Using Eq. (7) to compute the difference of the mean similarity for all keywords to each category between stage t and stage t+1, $SD_{t,t+1}$.

Step5: If $SD_{t,t+1} < \sigma$ or $t + 1 > Q$, proceed to the next step;
Otherwise, let t=t+1, and return to step 2.

Step6: Output the final clusters, $\widetilde{G_1}, \widetilde{G_2}, \ldots, \widetilde{G_K}$.

Based on this, the weight value of each attribute is calculated, denoted as:

$$u_n = \frac{TF - IWF^{n}}{\sum_{n=1}^{N} TF - IWF^{n}} \tag{9}$$

Risk Metric Model With Uncertainty and Conflicts

Decision risk is the possibility of decision errors in the decision-making process due to uncertainty generated by decision individual or group factors, and uncertainty of decision is a manifestation of decision risk. Therefore, it is necessary to use specific methods to quantify the connection between uncertainty and risk. In this paper, the authors adopt a risk measure model for the degree indicator from Xu et al. (2019c) to quantify the risk level of decision maker as follows:

$$H\left(X\right) = -\sum_{x \in X} P\left(x\right) log_{\frac{1}{\alpha}} \alpha P\left(x\right) \tag{10}$$

where $P\left(x\right)$ denote the degree indicator of x. α is the risk adjustment factor, which indicates the probability of decision maker's error, that is the error level, $\alpha \in \left[0,1\right]$. The smaller the value of α, the lower the probability of decision maker's error, the higher the decision-making ability, the better the control of decision risk, and the higher the accuracy of the decision.

In the process of LGEDM, the differences in psychological cognition, background, and knowledge of decision-makers, as well as the complexity and uncertainty of the decision problem, lead to inevitable preference conflicts among experts when evaluating alternatives. In addition, constrained by many practical factors, experts often have some uncertainties in expressing their evaluation information and cannot give a definite evaluation value. Usually, the greater the uncertainty of experts in expressing their information indicates the lower the reference value of the information given by the experts. Therefore, to ensure the validity of decision results, it is crucial to consider both factors of group conflict and hesitant ambiguity of assessment information in the LGEDM.

In this paper, the decision risk is considered caused by two aspects: hesitation ambiguity of preference information and group conflict. Thus, the hesitation fuzzy measures and conflict measures of HFLTS are introduced.

Definition 3: (Wei et al., 2018) Let $H_S = \left\{s_{\alpha_1}, s_{\alpha_2}, \ldots, s_{\alpha_l}\right\}$ be the hesitant fuzzy linguistic terms denoted on the set of linguistic terms $S = \left\{s_\mu \mid \mu = 0,1,\ldots,\tau, \tau \in N\right\}$, then the hesitation entropy of H_S can be expressed as:

$$E_h\left(\mathrm{H_S}\right) = \frac{l+3}{\tau+2},\ \ l \geq 2. \tag{11}$$

The hesitation entropy $E_h\left(\mathrm{H_S}\right)$ satisfies the following requirements:

- $E_h\left(\mathrm{H_S}\right) = 0$, if and only if $H_S = \left\{s_\mu\right\}, \mu = 0,1,\ldots,\tau$;
- $E_h\left(\mathrm{H_S}\right) = 1$, if and only if $H_S = \left\{s_0, s_1, \ldots, s_g\right\}$.

Definition 4: (Wei et al., 2018) Let $H_S = \left\{s_{\alpha_1}, s_{\alpha_2}, \ldots, s_{\alpha_l}\right\}$ be the hesitant fuzzy linguistic terms denoted on the set of linguistic terms $S = \left\{s_\mu \mid \mu = 0,1,\ldots,\tau, \tau \in N\right\}$, then the fuzzy entropy of H_S can be expressed as:

$$E_f\left(\mathrm{H_S}\right) = \frac{1}{l}\sum_{i=1}^{l}\frac{4\alpha_i}{\tau}\left(1 - \frac{\alpha_i}{\tau}\right) \tag{12}$$

The fuzzy entropy $E_f\left(\mathrm{H_S}\right)$ satisfies the following requirements.

- $E_f\left(\mathrm{H_S}\right) = 0$, if and only if $H_S = \left\{s_0\right\}$ or $H_S = \left\{s_\tau\right\}$;
- $E_f\left(\mathrm{H_S}\right) = 1$, if and only if $H_S = \left\{s_{\frac{\tau}{2}}\right\}$.

Then the comprehensive entropy that considers hesitancy and ambiguity is denoted as:

$$E_c\left(\mathrm{H_S}\right) = \frac{E_f + \beta E_h}{1 + \beta E_h} \tag{13}$$

The projection has the advantage that it can reflect the difference between two elements due to its ability for including the angle between two angles as well as describe their distances (Xu & Liu, 2013; Zhang & Wang, 2017). To compute the conflict degree between decision-makers, this paper defines the relative projection models by extending the method of Zhang et al. (2018) as follows.

Definition 5: Let $R_l = \left[H_{mn}^l\right]_{MN} \left(l = 1,2\right)$ be two matrices denoted in the linguistic term set $S = \left\{s_\mu \mid \mu = 0,1,\ldots,\tau,\tau \in N\right\}$, two matrices, where $H_{mn}^l = \left\{s_{\mu_{mn}^1}, s_{\mu_{mn}^2}, \ldots, s_{\mu_{mn}^{g_{mn}^l}}\right\}, \left(l = 1,2\right)$, the relative projection degree from R_1 to R_2 is denoted as:

$$RP_{R_2}\left(R_1\right) = \frac{\sum_{m=1}^{M}\sum_{n=1}^{N}\sum_{P_{mn}^1=1}^{g_{mn}^1}\sum_{P_{mn}^2=1}^{g_{mn}^2}\mu_{mn}^{p_{mn}^1}\cdot\mu_{mn}^{p_{mn}^2}}{\sum_{m=1}^{M}\sum_{n=1}^{N}\sum_{P_{mn}^2=1}^{g_{mn}^2}g_{mn}^1\cdot\left(\mu_{mn}^{p_{mn}^2}\right)^2} \tag{14}$$

Then, the similarity degree between R_1 and R_2 is expressed as:

$$SD\left(R_1, R_2\right) = \frac{RP_{R_2}\left(R_1\right) + RP_{R_1}\left(R_2\right)}{2} \tag{15}$$

The conflict degree is: $\mathrm{D}\left(R_1, R_2\right) = 1 - SD\left(R_1, R_2\right)$

Thus, based on the risk measurement model of degree indicators, the risk value of decision -maker can be calculated as follows.

First, the hesitation fuzzy risk of decision-maker d is calculated:

$$I_U^{\mathrm{d}} = \frac{1}{MN}\sum_{m=1}^{M}\sum_{n=1}^{N}I_U^{r_{mn}^d} \tag{16}$$

where, $I_U^{r_{mn}^d}$ is the hesitancy fuzzy entropy of decision-maker d about the nth attribute preference information for the mth alternative, as follows:

$$I_U^{r_{mn}^d} = -E_c\left(r_{mn}^d\right)log_{\frac{1}{\alpha}}\alpha E_c\left(r_{mn}^d\right) \tag{17}$$

and then, the conflict risk between decision maker d and the group decision makers is calculated:

$$I_C^{\mathrm{d}} = \frac{1}{D-1}\left[-\sum_{d'=1,d'\neq d}^{D}\frac{\mathrm{D}\left(R_d, R_{d'}\right)}{\sum_{d'=1,d'\neq d}^{D}\mathrm{D}\left(R_d, R_{d'}\right)}\times log_{\frac{1}{\alpha}}\frac{\alpha\mathrm{D}\left(R_d, R_{d'}\right)}{\sum_{d'=1,d'\neq d}^{D}\mathrm{D}\left(R_d, R_{d'}\right)}\right] \tag{18}$$

Finally, the group decision makers' risk value vector $I = \left\{I^1, I^2, \ldots, I^D\right\}$ for the group decision is obtained, and which is given by:

$$I^{d} = min\left[\beta I_{U}^{\mathrm{d}} + \left(1-\beta\right) I_{C}^{\mathrm{d}},1\right], d = 1,2,\ldots,D. \tag{19}$$

Risk Metric-Based Group Clustering Algorithm

In this section, a decision risk-based large group clustering algorithm is designed based on the k-means algorithm, which divide the decision members into several clusters with equal decision risk levels and can help control the overall decision risk. The clustering algorithm is implemented by using Python3.9 programming, and the specific steps are as follows.

Step1: Identify the initial clusters. Let t=0. First, the K values randomly is chosen as the initial clustering centers, denoted as: $I_{k}^{t}, k = 1,2,\ldots,K$; here the number of K values is determined by the management experience of experts and the number of data points. Subsequently, the risk deviation values of all the rest experts from each clustering center are calculated separately, denoted as:

$$I_{d,k}^{t} = \left|I_{k}^{t} - I^{d}\right|, k = 1,2,\ldots,K; d = 1,2,\ldots,D. \tag{20}$$

Finally, the decision-maker e^{d} is assigned to the category k with the smallest value of risk deviation Journals.eqp $I_{d,k}^{t}$, until all experts are assigned to the corresponding category.

Step2: Update the clusters. First, the risk deviation value of each expert from each cluster is recalculated based on the initial clustering result, denoted as:

$$I_{d,k}^{t+1} = \frac{1}{N_{k}^{t}} \sum_{n_{k}^{t}=1}^{N_{k}^{t}} \left|I_{n_{k}^{t}}^{t} - I^{d}\right|, k = 1,2,\ldots,K, d = 1,2,\ldots,D. \tag{21}$$

where N_{k}^{t} denotes the number of decision-makers included in the kth category at stage t.

Similarly, the smallest risk deviation value $I_{d,k}^{t+1}$ is chosen to reassign the decision-maker d to the corresponding category. The k-means algorithm is used to update the clusters in this process until the stopping condition has been reached.

Step3: Stopping condition. Set the maximum number of iterations Q and the threshold $\sigma \to 0\left(\sigma > 0\right)$.

Let $I_{t,t+1}$ denotes the difference of the average risk deviation values from all experts to each category at stage t and stage t+1, denoted as:

$$I_{t,t+1} = \frac{1}{DK} \sum_{d=1}^{D} \sum_{k=1}^{K} \left|I_{d,k}^{t+1} - I_{d,k}^{t}\right| \tag{22}$$

The clustering iteration stops when $I_{t,t+1} < \gamma$ or $t+1 > Q$; otherwise, it returns to the Step2. Details of the clustering algorithm are provided in Table 2.

In this paper, the authors assume that the experts of each cluster have equal importance, then the preference matrix $R^{G^{k}}$ of each cluster is obtained using the HIOWA operator, and the

Table 2. Risk metric-based group clustering algorithm (Algorithm 2)

Input: $I^d, d = 1,2,\ldots,D$ (Individual risk value) K (The number of categories) γ (The threshold value) **Output:** Clusters. $G_1, G_2, \ldots, G_k$ **Step1:** Let t=0, choose K initial clustering centers from $I^d, d = 1,2,\ldots,D$, noted as $I_k^t, k = 1,2,\ldots,K.$ **Step2:** For every $I^d, d = 1,2,\ldots,D, d \notin I_k^t$., apply Eq. (20) to compute the risk deviation value with each clustering center, $I_{d,k}^{t+1}, k = 1,2,\ldots,K, d = 1,2,\ldots,D$. Classify I^d int ohe category of G_k^t with the lowest risk deviation value among them. **Step3:** Let t=t+1, recomputing the risk deviation value $I_{d,k}^{t+1}$ between $I^d, d = 1,2,\ldots,D$ and $G_k^t, k = 1,2,\ldots,K$. Re-classify I^d into the category of G_k^{t+1} with the lowest risk deviation value among them. **Step4:** Using Eq. (22) to compute the difference of the mean risk deviation value for all $I^d, d = 1,2,\ldots,D$ to each category between stage t and stage t+1, noted as $I_{t,t+1}$. **Step5:** If $I_{t,t+1} < \gamma$ or $t+1 > Q,$ proceed to the next step; Otherwise, let t=t+1, and return to step 3. **Step6:** Output the clusters, $G_1, G_2, \ldots, G_k$.

cluster weights is assigned by combining the cluster size and the cluster risk value, the calculation functions are as follows:

The weights are assigned based on risk values.

$$\omega_I^{G_k} = \frac{1 - I^{G_k}}{K - \sum_{k=1}^{K} I^{G_k}}, \; I^{G_k} = \frac{1}{D^k} \sum_{d^k=1}^{D^k} I^{d^k} \tag{23}$$

where D^k represents the number of decision makers in the kth cluster and I^{G_k} represents the risk level of cluster k.

The weights are assigned based on the size of the clusters.

$$\omega_S^{G_k} = \frac{m_{G_k}}{\sum_{k=1}^{K} m_{G_k}} \tag{24}$$

where m_{G_k} denotes the number of decision makers in cluster k.

The comprehensive weight of clusters combining cluster size and cluster risk value is expressed as:

$$\omega_{G_k} = \alpha * \omega_I^{G_k} + \left(1-\alpha\right) * \omega_S^{G_k} \tag{25}$$

The Sorting Process

After the large-scale decision-makers are divided into several subgroups, the optimal alternative is determined by the following sorting process:

Step 1: The collective decision matrix is obtained. By the HIOWA operator, the collective decision matrix $R_C = \left\{r_{mn}^c\right\}_{MN}$ is obtained, where $r_{mn}^c = HIOWA\left(r_{mn}^1, r_{mn}^2, \ldots, r_{mn}^K\right)$.

The numerical matrix is given. Each component in the element r_{ij}^c is transformed equivalently into a crisp value by Eq. (26), noted as $R^V = \left\{r_{mn}^V\right\}_{MN}$.

$$r_{mn}^V = \frac{1}{g_{mn}^V} \sum_{p_{mn}^V=1}^{g_{mn}^V} \mu_{p_{mn}^V} \tag{26}$$

where, g_{mn}^V denotes the number of linguistic terms included in element r_{mn}^c from the matrix, and p_{mn}^V denotes the p_{mn}^V th linguistic term included in element r_{mn}^c from the matrix. r_{mn}^V and r_{mn}^c are considered to be equivalent here except their different expressions.

The overall score $c\left(x_i\right)$ of each alternative is calculated according to Eq. (27) to rank the alternatives.

$$c\left(x_i\right) = \sum\nolimits_{n=1}^{N} u_n \times r_{mn}^V \tag{27}$$

In summary, the procedures of an LGEDM approach with public preference data mining on HFLTS:

Step1: Several response alternatives are formulated quickly by decision members based on their own experience.

Step2: The public preference information is seized from Sina Weibo by Web Scraper, and the captured text content is cleaned through Python-based natural language processing. Then, the text content is cut up by using the Python - jieba package, and several effective keywords are extracted via TF-IWF methods, which is described in section 2.2.

Step3: The extracted valid keywords are classified into k categories that represent k attributes by using Algorithm 1. Subsequently, the TF-IWF values of the corresponding keywords in each attribute are accumulated to determine the attribute weights.

Step4: The evaluative information of alternatives is provided by decision makers separately based on the formulated attribute system, which is represented by the HFLTS.

Step5: The hesitation fuzzy entropy and conflict degree of decision-makers are computed using Def. (3), (4), (5); on the basis, the risk values of decision members are obtained by using Eq. (16), (18), (19).

Step6: The large group decision-makers are classified into several clusters by utilizing Algorithms 2. Then the subgroup weights are computed by Eq. (23), (24), (25) and the preference matrix of each cluster is obtained by the HIOWA operator.

Step7: The collective decision matrix is obtained by the HIOWA operator, and the collective decision matrix is transformed equivalently into a numerical matrix by Eq. (26). Then, the ranking results of the alternatives are obtained by calculating the overall score of each alternative according to Eq. (27).

CASE STUDY AND DISCUSSION

In this section, a case study of "Corona Virus Disease 2019, COVID-19" is provided to demonstrate the proposed approach. Early in 2020, pneumonia caused by the novel Coronavirus (Corona Virus Disease 2019, COVID-19) spread from its origin, Wuhan, to the whole country within days (Sandiwarno et al., 2020; Yi et al., 2021). The COVID-19 is a new strain of coronavirus that has never been found in humans before and its powerful transmission exceeds our imagination (Wang et al., 2020; Xu et al., 2020). On January 31, 2020, the World Health Organization declared the outbreak as an international public health emergency, showing its seriousness (Cho et al., 2022; Mishra et al., 2022). The outbreak drew strong public attention at the beginning of the outbreak; simultaneously, numerous Internet users published their opinions and insights via social networks after it occurred. we can mine valuable information from it to drive the decision-making process. According to "Chinese actions against the Corona Virus Disease 2019 ", the epidemic in China is mainly divided into five stages, this paper aim at the prevention and control of the epidemic after the initial control stage (Xu & Yu, 2022), suppose the National Health Commission (NHC) quickly convened 20 experts from the relevant fields $E = \left\{e_d \mid d = 1,2,\ldots,20\right\}$ to formulate 3 alternative plans $X = \left\{x_1, x_2, x_3\right\}$ in conjunction with the prevailing situation. The main contents of alternatives are as follows:

x_1 : Normal inter-city traffic, set up epidemic checkpoints, and immediately isolate those with abnormal health conditions; publicize and honor heroic deeds in the fight against the epidemic, and strengthen national social security.

x_2 : Ban traffic from key epidemic cities, and require nucleic acid test results and 14-day quarantine for travel to and from cities, broadcast news about the epidemic's progress; introduce policies to relieve pressure on industries heavily affected by the epidemic, and set up free psychological counseling sites online.

x_3 : Closed the epidemic key areas, other normal passage, set up the epidemic checkpoint, travel to and from different cities need to home isolation, found abnormal signs immediately isolated and do nucleic acid testing, send the epidemic prevention and control news through microblogging, strengthen public mental health education.

In this case, we suppose that the linguistic label set for the evaluation of alternatives is

$$S = \left\{s_0 : very\ poor, s_1 : poor, s_2 : slightly\ poor, s_3 : medium, s_4 : slightly\ good, s_5 : good, s_6 : very\ good\right\}.$$

Illustration of the Proposed Method

Below, the authors utilize the proposed model to select the best alternative.

Step1: Sina-Weibo is selected as the data collection channel, and the relevant text comment information during January 19, 2020, 0:00 to January 20, 2020, 0:00 related to the COVID-19 is captured for the data analysis. Then, we carry out data cleaning on the captured comments, including eliminating wrong fields, invalid fields, and abnormal values, URL, etc. At last, 64535 pieces of textual information are retained in total. The data obtained in this paper are also feasible and representative for decision making problems related to epidemic prevention and control.

Step2: First, cut the text content and extracted the keywords by the Python - jieba package. Then, the keywords' TF-IWF weight are calculated by using Eq. (5) and the 500 words with the highest TF-IWF values are selected as candidate keywords. At last, the valid keywords extracted are divided into 5 categories by Algorithm 1, which represents 5 attributes $C = \left\{c_n \mid n = 1,2,\ldots,5\right\}$. Thus, an attribute system for decision making about COVID-19 can be formed, as shown in Table 3.

Table 3. An attribute system for decision making about COVID-19

Attribute	c_1	c_2	c_3	c_4	c_5
Keywords	Medical staff, Hurry-up, Spare no effort, control, detection, treatment, etc.	Epidemic situation, panic, wish, cheer, believe, pray, salute, etc.	Wuhan, virus, wild game, pneumonia, source, news, government, etc.	safeness, confirmed diagnosis, infection, symptoms, cure, etc.	masks, Spring Festival, student, prevention, work, solation, etc.
u^n	0.08	0.22	0.16	0.28	0.26

Subsequently, the TF-IWF values of the corresponding keywords under each attribute are accumulated to determine the attribute weights, and the results are as follows.

$$\omega_1 = 0.08,\ \omega_2 = 0.22, \omega_3 = 0.16, \omega_4 = 0.28, \omega_5 = 0.26.$$

Step3: The decision matrix R^d provided by decision makers independently is obtained based on LTS S.

$$R^1 = \begin{bmatrix} \{s_5, s_6\} & \{s_5\} & \{s_4, s_5\} & \{s_6\} & \{s_4\} \\ \{s_4\} & \{s_3\} & \{s_4\} & \{s_3, s_4\} & \{s_2\} \\ \{s_3, s_4\} & \{s_3\} & \{s_4, s_5\} & \{s_2\} & \{s_4\} \end{bmatrix} R^2 = \begin{bmatrix} \{s_2\} & \{s_2, s_3\} & \{s_2, s_3\} & \{s_2\} & \{s_2, s_3\} \\ \{s_5, s_6\} & \{s_5\} & \{s_4, s_5\} & \{s_5\} & \{s_2\} \\ \{s_3\} & \{s_3, s_4\} & \{s_2\} & \{s_1\} & \{s_1\} \end{bmatrix}$$

$$R^3 = \begin{bmatrix} \{s_3, s_4\} & \{s_4\} & \{s_2, s_3\} & \{s_2\} & \{s_6\} \\ \{s_2\} & \{s_4\} & \{s_3\} & \{s_2, s_3\} & \{s_5, s_6\} \\ \{s_2, s_3\} & \{s_2\} & \{s_3\} & \{s_3, s_4\} & \{s_4, s_5\} \end{bmatrix} R^4 = \begin{bmatrix} \{s_6\} & \{s_5, s_6\} & \{s_5, s_6\} & \{s_6\} & \{s_4\} \\ \{s_4\} & \{s_3, s_4\} & \{s_4\} & \{s_3, s_4\} & \{s_2\} \\ \{s_3, s_4\} & \{s_3\} & \{s_4, s_5\} & \{s_2\} & \{s_4\} \end{bmatrix}$$

$$R^5 = \begin{bmatrix} \{s_2, s_3\} & \{s_2\} & \{s_3\} & \{s_2, s_3\} & \{s_2, s_3\} \\ \{s_6\} & \{s_5, s_6\} & \{s_4, s_5\} & \{s_5\} & \{s_1\} \\ \{s_3\} & \{s_3, s_4\} & \{s_3\} & \{s_1\} & \{s_1\} \end{bmatrix} R^6 = \begin{bmatrix} \{s_3\} & \{s_4, s_5\} & \{s_2\} & \{s_2\} & \{s_6\} \\ \{s_4, s_5\} & \{s_4\} & \{s_3\} & \{s_2, s_3\} & \{s_5, s_6\} \\ \{s_2, s_3\} & \{s_2\} & \{s_3\} & \{s_3, s_4\} & \{s_4, s_5\} \end{bmatrix}$$

$$R^7 = \begin{bmatrix} \{s_5, s_6\} & \{s_6\} & \{s_5\} & \{s_6\} & \{s_4\} \\ \{s_4\} & \{s_3, s_4\} & \{s_4\} & \{s_3, s_4\} & \{s_3\} \\ \{s_3, s_4\} & \{s_3\} & \{s_4, s_5\} & \{s_2\} & \{s_4\} \end{bmatrix} R^8 = \begin{bmatrix} \{s_1\} & \{s_2\} & \{s_2, s_3\} & \{s_2\} & \{s_2, s_3\} \\ \{s_6\} & \{s_6\} & \{s_4\} & \{s_4, s_5\} & \{s_2\} \\ \{s_2, s_3\} & \{s_3\} & \{s_2\} & \{s_1, s_2\} & \{s_1\} \end{bmatrix}$$

$$R^9 = \begin{bmatrix} \{s_3, s_4\} & \{s_5\} & \{s_5\} & \{s_5, s_6\} & \{s_3\} \\ \{s_4\} & \{s_4\} & \{s_3, s_4\} & \{s_3, s_4\} & \{s_2, s_3\} \\ \{s_5, s_6\} & \{s_5\} & \{s_4, s_5\} & \{s_5, s_6\} & \{s_3\} \end{bmatrix} R^{10} = \begin{bmatrix} \{s_2\} & \{s_2, s_3\} & \{s_2, s_3\} & \{s_2\} & \{s_2, s_3\} \\ \{s_6\} & \{s_5\} & \{s_4, s_5\} & \{s_5, s_6\} & \{s_1\} \\ \{s_2, s_3\} & \{s_3, s_4\} & \{s_2\} & \{s_1\} & \{s_0, s_1\} \end{bmatrix}$$

$$R^{11} = \begin{bmatrix} \{s_1\} & \{s_2, s_3\} & \{s_2, s_3\} & \{s_2\} & \{s_1, s_2\} \\ \{s_5, s_6\} & \{s_4\} & \{s_4, s_5\} & \{s_5\} & \{s_1\} \\ \{s_3\} & \{s_3, s_4\} & \{s_2, s_3\} & \{s_1\} & \{s_1\} \end{bmatrix} R^{12} = \begin{bmatrix} \{s_3, s_4\} & \{s_6\} & \{s_4, s_5\} & \{s_5\} & \{s_2\} \\ \{s_4\} & \{s_3, s_4\} & \{s_4\} & \{s_3, s_4\} & \{s_3\} \\ \{s_5, s_6\} & \{s_6\} & \{s_4, s_5\} & \{s_6\} & \{s_3\} \end{bmatrix}$$

$$R^{13}=\begin{bmatrix}\{s_1\} & \{s_2\} & \{s_2,s_3\} & \{s_2\} & \{s_3\}\\ \{s_5\} & \{s_6\} & \{s_5\} & \{s_4,s_5\} & \{s_2\}\\ \{s_3\} & \{s_2,s_3\} & \{s_2\} & \{s_1\} & \{s_1,s_2\}\end{bmatrix}\quad R^{14}=\begin{bmatrix}\{s_3\} & \{s_5\} & \{s_4,s_5\} & \{s_4\} & \{s_2\}\\ \{s_4\} & \{s_3,s_4\} & \{s_4\} & \{s_3\} & \{s_2\}\\ \{s_5,s_6\} & \{s_6\} & \{s_5\} & \{s_6\} & \{s_2,s_3\}\end{bmatrix}$$

$$R^{15}=\begin{bmatrix}\{s_3\} & \{s_3,s_4\} & \{s_2\} & \{s_2\} & \{s_6\}\\ \{s_4,s_5\} & \{s_4\} & \{s_3\} & \{s_2,s_3\} & \{s_5,s_6\}\\ \{s_2,s_3\} & \{s_2\} & \{s_2\} & \{s_3,s_4\} & \{s_5\}\end{bmatrix}\quad R^{16}=\begin{bmatrix}\{s_3\} & \{s_4,s_5\} & \{s_4\} & \{s_3,s_4\} & \{s_1,s_2\}\\ \{s_3,s_4\} & \{s_4\} & \{s_3\} & \{s_2,s_3\} & \{s_3\}\\ \{s_5\} & \{s_6\} & \{s_5,s_6\} & \{s_6\} & \{s_3\}\end{bmatrix}$$

$$R^{17}=\begin{bmatrix}\{s_3\} & \{s_4,s_5\} & \{s_2\} & \{s_2\} & \{s_6\}\\ \{s_4\} & \{s_4,s_5\} & \{s_3\} & \{s_2,s_3\} & \{s_5\}\\ \{s_2,s_3\} & \{s_3\} & \{s_2\} & \{s_3,s_4\} & \{s_4,s_5\}\end{bmatrix}\quad R^{18}=\begin{bmatrix}\{s_1,s_2\} & \{s_2\} & \{s_2,s_3\} & \{s_2\} & \{s_3\}\\ \{s_5,s_6\} & \{s_6\} & \{s_4\} & \{s_4,s_5\} & \{s_2\}\\ \{s_3,s_4\} & \{s_2,s_3\} & \{s_2\} & \{s_1,s_2\} & \{s_1\}\end{bmatrix}$$

$$R^{19}=\begin{bmatrix}\{s_1\} & \{s_2,s_3\} & \{s_2,s_3\} & \{s_2\} & \{s_2,s_3\}\\ \{s_5,s_6\} & \{s_6\} & \{s_4,s_5\} & \{s_5\} & \{s_2\}\\ \{s_3\} & \{s_2,s_3\} & \{s_2\} & \{s_1\} & \{s_1\}\end{bmatrix}\quad R^{20}=\begin{bmatrix}\{s_5,s_6\} & \{s_5\} & \{s_5\} & \{s_6\} & \{s_4\}\\ \{s_3\} & \{s_4\} & \{s_4\} & \{s_3,s_4\} & \{s_2,s_3\}\\ \{s_3,s_4\} & \{s_3\} & \{s_4,s_5\} & \{s_2\} & \{s_4\}\end{bmatrix}$$

Step4: Compute the risk value of all decision-makers using Eqs. (16), (18) and (19). On the basis, classify the large group decision-makers into 4 categories by utilizing the Algorithm 2. Then, obtain the preference matrix of each cluster using the HIOWA operator, and the results are shown in Table 4.

Table 4. Clustering results

	Members	$\boldsymbol{R^{G^k}}$	$\boldsymbol{I^{G_k}}$
$\boldsymbol{G^1}$	e_1,e_4,e_7,e_{20}	$\begin{matrix}\{s_3,s_{3.25}\} & \{s_{3.75},s_{4.25}\} & \{s_2,s_{2.25}\} & \{s_2\} & \{s_6\}\\ \{s_{3.5},s_4\} & \{s_4,s_{4.25}\} & \{s_3\} & \{s_2,s_3\} & \{s_5,s_{5.75}\}\\ \{s_2,s_3\} & \{s_{2.25}\} & \{s_{2.5}\} & \{s_3,s_4\} & \{s_{4.25},s_5\}\end{matrix}$	0.4362
$\boldsymbol{G^2}$	$e_6,e_9,e_{12},e_{14},e_{16}$	$\begin{matrix}\{s_3,s_{3.5}\} & \{s_5,s_{5.25}\} & \{s_{4.25},s_{4.75}\} & \{s_{4.25},s_{4.75}\} & \{s_2,s_{2.25}\}\\ \{s_{3.75},s_4\} & \{s_{3.5},s_4\} & \{s_{3.5},s_{3.75}\} & \{s_{2.75},s_{3.5}\} & \{s_{2.5},s_{2.75}\}\\ \{s_5,s_{5.75}\} & \{s_{5.75}\} & \{s_{4.5},s_{5.25}\} & \{s_{5.75},s_6\} & \{s_{2.75},s_3\}\end{matrix}$	0.3987
$\boldsymbol{G^3}$	$e_2,e_3,e_{10},e_{15},e_{17}$	$\begin{matrix}\{s_{5.25},s_6\} & \{s_{5.25},s_{5.5}\} & \{s_{4.75},s_{5.25}\} & \{s_6\} & \{s_4\}\\ \{s_{3.75}\} & \{s_{3.25},s_{3.75}\} & \{s_4\} & \{s_3,s_4\} & \{s_{2.25},s_{2.5}\}\\ \{s_3,s_4\} & \{s_3\} & \{s_4,s_5\} & \{s_2\} & \{s_4\}\end{matrix}$	0.3816
$\boldsymbol{G^4}$	e_5， $e_8,e_{11},e_{13},e_{18},e_{19}$	$\begin{matrix}\{s_{1.375},s_{1.63}\} & \{s_2,s_{2.5}\} & \{s_{2.13},s_3\} & \{s_2,s_{2.13}\} & \{s_{2.13},s_{2.88}\}\\ \{s_{5.38},s_{5.88}\} & \{s_{5.38},s_{5.5}\} & \{s_{4.13},s_{4.75}\} & \{s_{4.63},s_{5.13}\} & \{s_{1.625}\}\\ \{s_{2.75},s_{3.13}\} & \{s_{2.63},s_{3.5}\} & \{s_{2.13},s_{2.25}\} & \{s_1,s_{1.25}\} & \{s_{0.88},s_{1.13}\}\end{matrix}$	0.2015

Step5: Compute the subgroup weights ω_{G^k} by Eq. (23), (24), (25).

$\omega^{G_1} = 0.2148,\ \omega^{G_2} = 0.2329,\ \omega^{G_3} = 0.2395,\ \omega^{G_4} = 0.3092$

Then, the collective decision matrix is obtained using the HIOWA operator.

$$R^C = \begin{bmatrix} \left\{s_{3.00}, s_{3.42}\right\} & \left\{s_{3.82}, s_{4.21}\right\} & \left\{s_{3.19}, s_{3.76}\right\} & \left\{s_{3.44}, s_{3.59}\right\} & \left\{s_{3.38}, s_{3.68}\right\} \\ \left\{s_{4.22}, s_{4.54}\right\} & \left\{s_{4.15}, s_{4.48}\right\} & \left\{s_{3.71}, s_{3.96}\right\} & \left\{s_{3.24}, s_{4.03}\right\} & \left\{s_{2.71}, s_{2.99}\right\} \\ \left\{s_{3.16}, s_{3.90}\right\} & \left\{s_{3.35}, s_{3.63}\right\} & \left\{s_{3.19}, s_{3.63}\right\} & \left\{s_{2.76}, s_{3.12}\right\} & \left\{s_{2.77}, s_{3.07}\right\} \end{bmatrix}$$

Step6: Transform the collective decision matrix into a numerical matrix equivalently by Eq. (14).

$$R^V = \begin{bmatrix} 3.21 & 4.02 & 3.48 & 3.52 & 3.53 \\ 4.38 & 4.32 & 3.84 & 3.64 & 2.85 \\ 3.53 & 3.49 & 3.41 & 2.98 & 2.92 \end{bmatrix}$$

And then, the synthetical values of alternatives are calculated based on the results of example by Eq. (15).

$c\left(x_1\right)=3.60,\ c\left(x_2\right) = 3.66, c\left(x_3\right) = 3.19$.

Finally, the ranking of the alternatives can be shaped as $x_2 > x_1 > x_3$. Therefore, the best alternative for this problem is x_2 .

Comparative Analysis

For further validation, a comparative analysis is performed between the method from this paper and the method from Zhang et al. (2019) to illustrate the rationality of the proposed method. To avoid the interference of any other factors, only the clustering and ranking process are modified accordingly, and the number of categories k is unified to 4. Note that the data used in the comparative analyses are from the case study in Section 4.1. The detailed results are listed in Table 5.

Given the results in Table 5, the following conclusions can be drawn as follows:

1. The ranking results obtained by the two methods are equal, that was $x_2 > x_1 > x_3$, and x_2 is the optimal solution, which demonstrates the rationality of the method in this paper.
2. The clustering results from the two methods are slightly different. This is because the classification standards of the two methods are different. This paper's method divides the group based on the risk level of decision-makers, each clustered expert has a similar risk level by this method, while the method from Zhang et al. (2019) is clustered on the distance between decision-makers.
3. The risk level of the results obtained in this paper is significantly lower than the results obtained by the method from Zhang et al. (2019). The main reason is that the method from Zhang et al. (2019) mainly consider fuzzy clustering and weight calculation methods, but the method in this paper not only consider these two aspects, but also consider decision risk factors from both group conflict and individual decision information, and incorporate them into the clustering and weight assignment process.

Table 5. Comparison results of methods

Methods	Method from this paper	Method from Zhang et al. (2019)
Clustering result	e_1, e_4, e_7, e_{20} $e_6, e_9, e_{12}, e_{14}, e_{16}$ $e_2, e_3, e_{10}, e_{15}, e_{17}$ $e_5, e_8, , e_{11}, e_{13}, e_{18}, e_{19}$	e_1, e_4, e_7, e_{20} $e_9, e_{12}, e_{14}, e_{16}$ e_3, e_6, e_{15}, e_{17} $e_2, e_5, e_8, e_{10}, e_{11}, e_{13}, e_{18}, e_{19}$
Ranking results	$x_2 > x_1 > x_3$	$x_2 > x_1 > x_3$
Risk value	0.3418	0.4215

By comparison, the method in this paper is reasonable, effective, and superior. In addition, the group decision-making method based on risk measures proposed in this paper can effectively reduce the risk level of group decision-making and enable the decision result to be more scientific.

CONCLUSION

In this paper, a new LGEDM approach based on public preference attribute mining is proposed for the emergency decision problem by combining data science, risk theory and group decision theory, which establishes a certain theoretical foundation for the LGEDM method of the big data environment. In addition to the "COVID-19" mentioned in the example section, this study is also applicable to the emergency decision-making problem in the context of other major pubic-related emergencies, such as major natural disasters, unexpected accidental disasters, other major public health events, etc. The main innovations and contributions in this paper are as follows:

1. The construction of the decision attribute system in this paper is based on public preference information mining, which can effectively avoid the subjective interference of traditional experts when constructing the decision attribute system and could improve public satisfaction for decision making.
2. In this paper, a new clustering algorithm is assigned to cluster keywords based on the word semantic similarity algorithm and k-means clustering algorithm. It can quickly and effectively merge massive keywords extracted from public preference information and accelerate the decision analysis process.
3. The decision risk caused by individual uncertainty and group conflict under hesitant fuzzy language environment was quantified using the degree index risk measure model. On the basis, group clustering and weight assignment process are achieved, which effectively reduces decision risk and improves decision quality.
4. The linguistic information processing process is completely implemented by the Python program, which does not involve complex semantic analysis of massive grammar and sentences, and has high text processing efficiency, thus can meet the timeliness requirements of large group emergency decision-making.

Furthermore, there are still some limitations to be overcome in the future: (1) in the early stage of a major emergency outbreak, the relevant data from the network often exists more noise and information over-dispersion, which will affect the decision results. (2) in this paper, only the risk factors caused by individual information hesitation fuzziness and group conflict are considered, while other risk factors, such as non-cooperative behavior, trust risk, and other factors are not considered. Therefore, these aspects can be considered in future studies to further improve the decision quality.

ACKNOWLEDGMENT

This research was supported by the Major Social Science Foundation of China [grant number 18ZDA052]. All authors of this article declare there are no competing interest.

REFERENCES

Cao, M. S., Liu, Y. J., Gai, T. T., Zhou, M., Fujita, H., & Wu, J. (2022). A Comprehensive Star Rating Approach for Cruise Ships Based on Interactive Group Decision Making with Personalized Individual Semantics. *Journal of Marine Science and Engineering*, *10*(5), 1–20. doi:10.3390/jmse10050638

Chao, X. G., Kou, G., Peng, Y., & Herrera-Viedma, E. (2020). Large-scale group decision-making with non-cooperative behaviors and heterogeneous preferences: An application in financial inclusion. *European Journal of Operational Research*, *288*(1), 271–293. doi:10.1016/j.ejor.2020.05.047

Chen, H. J., Zhen, X., & Chen, X. (2014). Survey on information diffusion in microblog. *Application Research of Computers, 31*(2), 333-338.

Chen, J. Y., Zhou, K., & Lu, J. C. (2021). IC-based Approach for Calculating Word Semantic Similarity in HowNet and Cilin. *Journal of Chinese Computer Systems,* 1-7. 10.20009/j.cnki.21-1106/TP.2020-1078

Chen, Z. S., Martinez, L., Chin, K. S., & Tsui, K. L. (2018). Two-stage aggregation paradigm for HFLTS possibility distributions: A hierarchical clustering perspective. *Expert Systems with Applications*, *104*, 43–66. doi:10.1016/j.eswa.2018.03.013

Cho, V., Roll, L. C., Wu, C. H., & Tang, V. (2022). Changing Digital Age in the Wake of COVID-19: How Does Humility Impact on Virtual Leaderless Teams? *Journal of Global Information Management*, *30*(4), 1–23. doi:10.4018/JGIM.290831

Farhadinia, B., & Xu, Z. (2018). Novel hesitant fuzzy linguistic entropy and cross-entropy measures in multiple criteria decision making. *Applied Intelligence*, *48*(11), 1–13. doi:10.1007/s10489-018-1186-2

Gou, X., Xu, Z., & Herrera, F. (2018). Consensus Reaching Process for Large-scale Group Decision Making with Double Hierarchy Hesitant Fuzzy Linguistic Preference Relations. *Knowledge-Based Systems*, *157*(1), 20–33. doi:10.1016/j.knosys.2018.05.008

Guo, J., Wu, X. H., & Wei, G. (2020). A new economic loss assessment system for urban severe rainfall and flooding disasters based on big data fusion. *Environmental Research*, *188*, 109822. doi:10.1016/j.envres.2020.109822 PMID:32615356

Hartigan, J. A., & Wong, M. A. (1979). Algorithm as 136: A k-means clustering algorithm. *Applied Statistics*, *28*(1), 100–108. doi:10.2307/2346830

Huang, C. H., Yin, J., & Hou, F. (2011). A text similarity measurement combining word semantic information with tf-idf method. *Jisuanji Xuebao*, *34*(5), 856–864. doi:10.3724/SP.J.1016.2011.00856

Liu, Q., Wu, H. Y., & Xu, Z. S. (2021). Consensus model based on probability k-means clustering algorithm for large scale group decision making. *International Journal of Machine Learning and Cybernetics*, *12*(6), 1609–1626. doi:10.1007/s13042-020-01258-5

Lu, Y. H., & Li, Y. F. (2013). Improvement of Text Feature Weighting Method Based on TF-IDF Algorithm. *Library and Information Service, 57*(3), 90-95.

Mandal, P., Samanta, S., & Pal, M. (2021). Large-scale group decision-making based on pythagorean linguistic preference relations using experts clustering and consensus measure with non-cooperative behavior analysis of clusters. *Complex & Intelligent Systems*, 1–15. doi:10.1007/s40747-021-00369-y

Mishra, D., & Satapathy, S. (2022). The Pandemic COVID-19 and Its Impact on Indian Agricultural Sectors: An Assessment of Farmers. *Journal of Global Information Management*, *30*(4), 1–27. doi:10.4018/JGIM.297906

Rodríguez, R. M., Labella, L., Sesma-Sara, M., Bustince, H., & Martínez, L. (2021). A cohesion-driven consensus reaching process for large scale group decision making under a hesitant fuzzy linguistic term sets environment. *Computers & Industrial Engineering*, *155*(17), 107158. doi:10.1016/j.cie.2021.107158

Rodríguez, R. M., Martínez, L., & Herrera, F. (2011). *Hesitant Fuzzy Linguistic Term Sets*. IEEE Press. doi:10.1007/978-3-642-25664-6_34

Rodriguez, R. M., Martinez, L., & Herrera, F. (2012). Hesitant fuzzy linguistic term sets for decision making. *IEEE Transactions on Fuzzy Systems*, *20*(1), 109–119. doi:10.1109/TFUZZ.2011.2170076

Sandiwarno, S. (2020). Users' Emotions Analysis based on Hybrid Feature Extraction Techniques. *International Journal of Scientific Research in Computer Science Engineering and Information Technology*, 291-296. .10.32628/CSEIT206658

Sanjivani, J., Malhar, J., & Trupti, W. (2020). An overview of corona virus disease of 2019 (COVID-19). *Journal of Critical Reviews*, *7*(10). doi:10.31838/jcr.07.10.12

Tan, L., Wu, X., Guo, J., & Santibanez-Gonzalez, E. D. R. (2021). Assessing the impacts of covid-19 on the industrial sectors and economy of China. *Risk Analysis*. doi:10.1111/risa.13805 PMID:34448216

Tang, M., Liao, H. C., Xu, J. P., Streimikiene, D., & Zheng, X. S. (2020). Adaptive consensus reaching process with hybrid strategies for large-scale group decision making. *European Journal of Operational Research*, *282*(3), 957–971. doi:10.1016/j.ejor.2019.10.006

Tang, M., Zhou, X. Y., Liao, H. C., Xu, J. P., Fujita, H., & Herrera, F. (2019). Ordinal consensus measure with objective threshold for heterogeneous large-scale group decision making. *Knowledge-Based Systems*, *180*, 62–74. doi:10.1016/j.knosys.2019.05.019

Tian, H., & Wu, L. (2018). Microblog Emotional Analysis Based on TF-IWF Weighted Word2vec Model. *2018 IEEE 9th International Conference on Software Engineering and Service Science (ICSESS). IEEE*. IEEE. doi:10.1109/ICSESS.2018.8663837

Torra, V. (2010). Hesitant fuzzy sets. *International Journal of Intelligent Systems*, *25*(6), 529–539. doi:10.1002/int.20418

Wang, Y., Shen, H., Gao, W. W., Zheng, J. Q., Wu, Y. W., Liu, C. Z., & Chen, H. (2020). Advances in Genomics Research of Novel Coronavirus. *Chinese Journal of Modern Applied Pharmacy*, *37*(5), 525–529. doi:10.13748/j.cnki.issn1007-7693.2020.05.002

Wei, C. P., Rodriguez, R. M., & Martinez, L. (2018). Uncertainty measures of extended hesitant fuzzy linguistic term sets. *IEEE Transactions on Fuzzy Systems*, *26*(3), 1–1. doi:10.1109/TFUZZ.2017.2724023

Wu, J., Cao, M., Chiclana, F., Dong, Y., & Herrera-Viedma, E. (2020). An optimal feedback model to prevent manipulation behaviors in consensus under social network group decision making. *IEEE Transactions on Fuzzy Systems*, *99*, 1–1. doi:10.1109/TFUZZ.2020.2985331

Wu, J., Hong, Q., Cao, M. S., Liu, Y. J., & Fujita, H. (2022). A group consensus-based travel destination evaluation method with online reviews. *Applied Intelligence*, *52*(2), 1306–1324. doi:10.1007/s10489-021-02410-6

Wu, J., Wang, S., Chiclana, F., & Herrera-Viedma, E. F., & Ellow. (. (2021a). Two-fold personalized feedback mechanism for social network consensus by uninorm interval trust propagation. *IEEE Transactions on Cybernetics*, *99*, 1. doi:10.1109/TCYB.2021.3076420 PMID:34003760

Wu, S., & Zhang, Z. (2021). New distance measures of hesitant fuzzy linguistic term sets. *Physica Scripta, 96*(1), 015002. .10.1088/1402-4896/abc385

Wu, X., Zhao, J., Yun, K., Guo, J., & Gao, G. (2021b). Construction and verification of a rainstorm death risk index based on grid data fusion: A case study of the beijing rainstorm on july 21, 2012. *Natural Hazards*, *107*(3), 1–26. doi:10.1007/s11069-021-04507-0

Xing, Y. M., Cao, M. S., Liu, Y. J., Zhou, M., & Wu, J. (2022). A Choquet integral based interval Type-2 trapezoidal fuzzy multiple attribute group decision making for Sustainable Supplier Selection. *Computers & Industrial Engineering*, *165*, 107935. doi:10.1016/j.cie.2022.107935

Xiong, K., Dong, Y. C., & Zhao, S. H. (2022). A Clustering Method with Historical Data to Support Large-Scale Consensus-Reaching Process in Group Decision-Making. *International Journal of Computational Intelligence Systems*, *15*(1), 17. doi:10.1007/s44196-022-00072-x

Xu, B. L., Guan, J. L., Shu, C., Yu, T., Luo, F., Cao, X. L., Wang, Y. B., & Zhou, C. Y. (2020). Advances in related research on novel coronavirus (COVID-19). *Zhongguo Yiyuan Ganranxue Zazhi*, *30*(06), 839–844. doi:10.11816/cn.ni.2020-206034

Xu, G. L., & Liu, F. (2013). An approach to group decision making based on interval multiplicative and fuzzy preference relations by using projection. *Applied Mathematical Modelling*, *37*(6), 3929–3943. doi:10.1016/j.apm.2012.08.007

Xu, X., Du, Z., Chen, X. H., & Cai, C. (2019a). Confidence consensus-based model for large-scale group decision making: A novel approach to managing non-cooperative behaviors. *Information Sciences*, *477*, 410–427. doi:10.1016/j.ins.2018.10.058

Xu, X. H., Cai, C. G., & Chen, X. H. (2015). A Multi-stage Conflict Style Large Group Decision" making Emergency Method Based on Interval Fuzzy Number. *Operations Research and Management Science*, *24*(004), 9–15. doi:10.3969/j.issn.1007-3221.2015.04.002

Xu, X. H., Du, Z. J., Chen, X. H., & Zhou, Y. J. (2017). Conflict large-group emergency decision-making method while protecting minority opinions. *Journal of Management Sciences in China, 20*(11), 10-23.

Xu, X. H., Wang, L. L., Chen, X. H., & Liu, B. S. (2019b). Large group emergency decision-making method with linguistic risk appetites based on criteria mining. *Knowledge-Based Systems*, *182*, 104849. doi:10.1016/j.knosys.2019.07.020

Xu, X. H., Yin, X. P., & Chen, X. H. (2019c). A large-group emergency risk decision method based on data mining of public attribute preferences. *Knowledge-Based Systems*, *2019*(163), 495–509. doi:10.1016/j.knosys.2018.09.010

Xu, X. H., & Yu, Z. X. (2022). A large group emergency decision making method and application based on attribute mining of public behavior big data in social network environment. *Control and Decision*, *37*(01), 175–184. doi:10.13195/j.kzyjc.2020.1789

Xu, Z. (2004). Uncertain linguistic aggregation operators-based approach to multiple attribute group decision making under uncertain linguistic environment. *Information Sciences*, *168*(1), 171–184. doi:10.1016/j.ins.2004.02.003

Yager, R. R., & Filev, D. P. (1999). Induced ordered weighted averaging operators. *IEEE Transactions on Systems Man & Cybernetics Part B Cybernetics A Publication of the IEEE Systems Man & Cybernetics Society, 29*(2), 141-150. .10.1109/3477.752789

Yi, K., Rong, Y., Wang, C., Huang, L., & Wang, F. (2021). COVID-19: Advance in laboratory diagnostic strategy and technology. *Molecular and Cellular Biochemistry*, *4*(3), 1–18. doi:10.1007/s11010-020-04004-1 PMID:33389499

Yin, X. P., Xu, X. H., & Pan, B. (2021). Selection of strategy for large group emergency decision-making based on risk measurement. *Reliability Engineering & System Safety*, *208*, 107325. doi:10.1016/j.ress.2020.107325

Yu, S. M., Du, Z. J., & Xu, X. H. (2020). Hierarchical Punishment-Driven Consensus Model for Probabilistic Linguistic Large-Group Decision Making with Application to Global Supplier Selection. *Group Decision and Negotiation*, *30*(6), 1343–1372. doi:10.1007/s10726-020-09681-3

Zadeh, L. A. (1975). The concept of a linguistic variable and its application to approximate reasoning—i. *Information Sciences*, *8*(3), 199–249. doi:10.1016/0020-0255(75)90036-5

Zhang, X. Y., & Wang, J. Q. (2017). Consensus-based framework to MCGDM under multi-granular uncertain linguistic environment. *Journal of Intelligent & Fuzzy Systems*, *33*(2), 1–12. doi:10.3233/IFS-162184

Zhang, X. Y., Wang, J. Q., & Hu, J. H. (2018). A consensus approach to multi-granular linguistic MCGDM with hesitant fuzzy linguistic information by using projection. *Journal of Intelligent & Fuzzy Systems*, *34*(3), 1959–1974. doi:10.3233/JIFS-171629

Zhong, X. Y., & Xu, X. H. (2020). Clustering-based method for large group decision making with hesitant fuzzy linguistic information: Integrating correlation and consensus. *Applied Soft Computing*, *87*, 105973. doi:10.1016/j.asoc.2019.105973

Zhong, X. Y., Xu, X. H., Chen, X. H., & Goh, M. (2020). Large group decision-making incorporating decision risk and risk attitude: A statistical approach. *Information Sciences*, *533*, 120–137. doi:10.1016/j.ins.2020.04.003

Zhong, X. Y., Xu, X. H., & Yin, X. P. (2021). A multi-stage hybrid consensus reaching model for multi-attribute large group decision-making: Integrating cardinal consensus and ordinal consensus. *Computers & Industrial Engineering*, *158*, 107443. doi:10.1016/j.cie.2021.107443

Mengke Zhao is currently a 2nd year Ph.D. candidate at the School of Economics and Management, Shanghai Maritime University, Shanghai, China. She is supervised by Prof. Ji Guo. Her research interests are mainly on big data analysis and management, and the theory and method of large group emergency decision making. By now, she has published 2 papers in international journals. She is involved in three research projects as "Research on Optimal Control of Air Pollutant Emission in Big Data Era (18ZDA052), Research on Meteorological Disaster Emergency Management Based on Big Data Fusion, and Big data fusion method to study meteorological disaster emergency decision support (91546117)".

Ji Guo is a Professor at the School of Economics and Management, Shanghai Maritime University, China. Her areas of specialty include environmental economics and emergency management. She has published 4 books and more than 60 academic articles in journals like Annals of Operational Research, Environmental Hazards, Journal of International Trade & Economic Development, Journal of Cleaner Production, Weather, Climate and Society, and others. As a Principal Investigator, she has 2 national research projects as "Research Plan of National Social Science Foundation: Research on Optimal Control of Haze Emission Based on Data Envelopment Analysis (17BGL142) and Innovative Evaluation Methods of Meteorological Public Service Benefit (11CGL100)".

Xianhua Wu is a Distinguished Professor of Shanghai Maritime University, the Party Committee Deputy Secretary, and the president of the School of Economics and Management. His areas of specialty include disaster risk analysis, big data analysis, applied statistics, and economics of the shipping industry. He has published 8 books such as Economic Impacts and Emergency Management of Disasters in China (Published by Springer Nature) and more than 130 academic articles in journals like Risk Analysis, Ecological Economics, Annals of Operations Research, Transportation Research Part D, Science of the Total Environment, Journal of International Trade & Economic Development, Weather, Climate and Society, and others. As a Principal Investigator, He has 7 national research projects, he is also a member of the Steering Committee of Economics Education (2018-2022), MOE, China.