

RESEARCH

Open Access



Novel distance measures of hesitant fuzzy sets and their applications in clustering analysis

Fuping Liao, Wu Li*, Xiaoqiang Zhou and Gang Liu

*Correspondence:

liwu0817@163.com

School of Information Science And
Engineering, Hunan Institute of
Science and Technology, Yueyang,
Hunan, People's Republic of China

Abstract

Distance and similarity measures are very important in clustering, pattern recognition, decision-making and other scientific fields. For the existing hesitant fuzzy distance, most of them do not consider the hesitance degree. Even if the hesitance degree is considered, only the degree of dispersion or the number of hesitant fuzzy values are considered. Aiming at these shortages, a new hesitance degree is defined, which has better accuracy and applicability. Then, some hesitant fuzzy distance measures based on the proposed hesitance degree are proposed, which can overcome some shortcomings of the existing distance measures. Finally, the new hesitant fuzzy distance is applied to the hierarchical hesitant fuzzy k-means clustering algorithm, and an illustration example is given to illustrate the effectiveness of the proposed method.

Keywords: Hesitant fuzzy set, Distance measure, Hesitance degree, Clustering analysis

Introduction

The theory of fuzzy sets proposed by Zadeh [1] has achieved a great success in various fields. Afterwards, many new theories and approaches about uncertainty and imprecision have been proposed by scholars, such as intuitionistic fuzzy sets (IFS) [2], interval-valued intuitionistic fuzzy sets [3], linguistic variables [4], type-2 fuzzy sets [5], fuzzy multiset [6], picture fuzzy sets (PFS) [7], etc. With the growing complexity and uncertainty of the real-life problems, it is hard to establish the degree of membership of fuzzy set. To do this, Torra [8] introduced the concept of hesitant fuzzy set (HFS) which permitted the membership having a set of possible values. As an extended form of the fuzzy set, hesitant fuzzy set can better simulate the hesitation preference of decision makers to deal with the actual situation of people hesitating between several possible values. Since the hesitation fuzzy set came out, it has received extensive attention and obtained rich research results. For example, Zhang [9] proposed the hesitant fuzzy power average operator, it is characterized by the weight of hesitation fuzzy information depends on the degree of support for it with other hesitation fuzzy information. Considering that attributes may be related to each other in realistic decision-making problems, Zhu [10, 11] proposed the hesitant

fuzzy Bonferroni mean operator and hesitant fuzzy Bonferroni geometric operator. Wei [12] considered the priority relationship between attributes and proposed the hesitant fuzzy prioritized operator. Xu et al. [13] introduced a hesitant fuzzy TOPSIS method based on the principle of maximum deviation and applied it to multi-attribute decision-making problems. Liao et al. [14] presented the hesitant fuzzy VIKOR multi-attribute decision-making method considering the psychological preference of decision-makers. Wang et al. [15] introduced the prospect value function of hesitant fuzzy elements based on prospect theory and distance measure, and then proposed a multi-attribute decision-making method according to the TOPSIS method that considers the risk preference of decision maker. Hesitant fuzzy sets also have been applied to the other fields such as cluster analysis [16–19], decision analysis [20–23] and pattern recognition [24–27] and so on.

Distance measure is one of the important direction in the theory of hesitant fuzzy set. So far, many research results on hesitant fuzzy distance have been obtained. For instance, Xu and Xia [28] first proposed a variety of hesitant fuzzy distance measures and discussed their properties. On the basis of hesitant fuzzy distance measure by Xu, Tong [29] introduced a hybrid hesitant fuzzy distance measure considering the preference of decision makers. And Peng [30] presented a generalized hesitant fuzzy cooperative weighted distance measure. Although the above hesitant fuzzy distance measures have many merits, they require that each corresponding hesitant fuzzy element has the same length. When the length of hesitant fuzzy elements is not equal, it is necessary to add elements to meet the requirements. However, this is bound to change the original information of hesitant fuzzy elements. That is to change the real expression of experts. To overcome the shortcoming, Tang et al. [31] proposed a distance measure without considering the length of the hesitant fuzzy element. But except for the length of the hesitant fuzzy element is 1, the distance between two identical hesitant fuzzy elements is not equal to 0, which is contrary to the fact. Later, some researchers further consider the hesitance degree of hesitant fuzzy element in distance measure. Zhang and Xu [18] proposed the concept of hesitation index which determined by the degree of dispersion of hesitant fuzzy values in the hesitant fuzzy element, and proposed a series of distance and similarity measures that consider hesitation index of hesitant fuzzy sets. Li et al. [32] proposed the concept of hesitance degree which determined by the number of hesitant fuzzy values in the hesitant fuzzy element, and proposed a series of hesitant fuzzy distance measures containing hesitance degree. However, it needs to be pointed out that the hesitance degree mentioned above only considers the degree of dispersion or the number of hesitant fuzzy values in the hesitant fuzzy element, which is imperfect and has the defect of insufficient discrimination.

Methods

According to the above analysis, the existing hesitant fuzzy distance measures have different shortcomings. To overcome the shortcomings, we first define a new hesitance degree by considering the degree of dispersion and the number of hesitant fuzzy values in hesitant fuzzy element, and put forward some distance measures based on the proposed hesitance degree. The distance is divided into two cases of equal length and unequal length between two hesitant fuzzy elements, which can solve the problem of

original information distortion caused by supplementary elements in the case of inconsistent lengths. Further, we apply the new hesitant fuzzy distance to the hierarchical hesitant fuzzy K-means clustering.

The paper is organized as follows. In [Methods](#) section, some concepts related to hesitant fuzzy sets are introduced. In [Preliminaries](#) section, a new hesitance degree and some new hesitant fuzzy distance measures are proposed, and their properties are discussed. In [Some New hesitant fuzzy distance measures](#) section, we applied the new distance measure to the hierarchical hesitant fuzzy K-means clustering algorithm. The fifth section is the conclusion of this paper.

Preliminaries

Definition 1 [8] *Given a fixed set X , then a hesitant fuzzy set (HFS) on X is in terms of a function that when applied to X returns a subset of $[0, 1]$.*

For convenience, Xia and Xu [33] usually express HFS simply as a mathematical symbol:

$$E = \{ \langle x, h_E(x) \rangle \mid x \in X \} \quad (1)$$

where $h_E(x)$ is a set of some different values in $[0, 1]$, representing the possible membership degrees of the element $x \in X$ to E . For convenience, we call $h = h_E(x)$ a hesitant fuzzy element (HFE) and H the set of all HFEs.

For the convenience of comparison, We arrange the elements in $h_E(x_i)$ in increasing order, and let $h_E^{\sigma(j)}(x_i)$ be the j th largest value in $h_E(x_i)$.

Li [32] put forward the axiomatic definition of distance measure for hesitant fuzzy sets (HFSs).

Definition 2 [32]. *Let A , B and C be three HFSs on X . Then, d is called a hesitant fuzzy distance measure for HFSs, which satisfies the following properties:*

- (1) $0 \leq d(A, B) \leq 1$;
- (2) $d(A, B) = 0$ if and only if $A = B$;
- (3) $d(A, B) = d(B, A)$;
- (4) $d(A, B) + d(B, C) \geq d(A, C)$.

It is noted that the number of values in different HFEs may be different, Xu and Xia extend the shorter one by adding the same value until both of them have the same length when we compare them. Let $l(h_E(x_i))$ be the number of values in $h_E(x_i)$, and $l_{x_i} = \max\{l(h_A(x_i)), l(h_B(x_i))\}$. Xu and Xia [28] proposed a series of hesitant fuzzy set distances as follows:

Definition 3 [28]. *Let A and B be two HFSs on $X = \{x_1, x_2, \dots, x_n\}$. Then, the hesitant normalized Hamming distance as follows:*

$$d_{xh}(A, B) = \frac{1}{n} \sum_{i=1}^n \left[\frac{1}{l_{x_i}} \sum_{j=1}^{l_{x_i}} \left| h_A^{\sigma(j)}(x_i) - h_B^{\sigma(j)}(x_i) \right| \right] \quad (2)$$

the hesitant normalized Euclidean distance as follows:

$$d_{xe}(A, B) = \left[\frac{1}{n} \sum_{i=1}^n \left(\frac{1}{l_{x_i}} \sum_{j=1}^{l_{x_i}} \left| h_A^{\sigma(j)}(x_i) - h_B^{\sigma(j)}(x_i) \right|^2 \right) \right]^{\frac{1}{2}} \quad (3)$$

the generalized hesitant normalized distance:

$$d_{xg}(A, B) = \left[\frac{1}{n} \sum_{i=1}^n \left(\frac{1}{l_{x_i}} \sum_{j=1}^{l_{x_i}} \left| h_A^{\sigma(j)}(x_i) - h_B^{\sigma(j)}(x_i) \right|^\lambda \right) \right]^{\frac{1}{\lambda}} \quad (4)$$

where $\lambda > 0$.

In order to measure the deviation of each HFE in each HFS, Zhang and Xu [18] et al. proposed the concept of hesitance degree of HFS.

Definition 4 [18]. Let H be an HFS in a reference set X , denoted by $H = \{ \langle x, h_H(x) \rangle \mid x \in X \}$ and $h_H(x_i) = \{ h_H^{\sigma(j)}(x_i) \mid j = 1, 2, \dots, l_h \}$. Then, the hesitance degree of x in H can be defined as follows:

$$h_Z(h_H(x)) = \begin{cases} \sqrt{\frac{2 \sum_{k>j=1}^{l_h} \left(h_H^{\sigma(k)}(x_i) - h_H^{\sigma(j)}(x_i) \right)^2}{l_h \times (l_h - 1)}}, & l_h > 1 \\ 0, & l_h = 1 \end{cases} \quad (5)$$

where l_h is the number of the elements in $h_H(x_i)$.

In general, the bigger the range among the possible values in each HFE is, the larger the hesitance degree of the HFE is. By considering the impact of the hesitance degree of HFEs, Xu and Zhang proposed a new method for measuring the distance between HFSs:

Definition 5 [18]. Let A and B be two HFSs on X . Then, the hesitant normalized Hamming distance including hesitance degree between A and B is defined as:

$$d_{hzh}(A, B) = \frac{1}{n} \sum_{i=1}^n \left(\frac{\alpha}{l_{x_i}} \sum_{j=1}^{l_{x_i}} \left| h_A^{\sigma(j)}(x_i) - h_B^{\sigma(j)}(x_i) \right| + \beta |h_Z(h_A(x_i)) - h_Z(h_B(x_i))| \right) \quad (6)$$

the hesitant normalized Euclidean distance including hesitance degree is defined as:

$$d_{hze}(A, B) = \left[\frac{1}{n} \sum_{i=1}^n \left(\frac{\alpha}{l_{x_i}} \sum_{j=1}^{l_{x_i}} \left| h_A^{\sigma(j)}(x_i) - h_B^{\sigma(j)}(x_i) \right|^2 + \beta |h_Z(h_A(x_i)) - h_Z(h_B(x_i))|^2 \right) \right]^{\frac{1}{2}} \quad (7)$$

the generalized hesitant normalized distance including hesitance degree is defined as:

$$d_{hzg}(A, B) = \left[\frac{1}{n} \sum_{i=1}^n \left(\frac{\alpha}{l_{x_i}} \sum_{j=1}^{l_{x_i}} \left| h_A^{\sigma(j)}(x_i) - h_B^{\sigma(j)}(x_i) \right|^\lambda + \beta |h_Z(h_A(x_i)) - h_Z(h_B(x_i))|^\lambda \right) \right]^{\frac{1}{\lambda}} \quad (8)$$

where $\lambda > 0$, $\alpha, \beta \in [0, 1]$, $\alpha + \beta = 1$, $h_A^{\sigma(j)}(x_i)$ and $h_B^{\sigma(j)}(x_i)$ are the j th values in $h_A(x_i)$ and $h_B(x_i)$, respectively. $h_Z(h_A(x_i))$ and $h_Z(h_B(x_i))$ are referred to the hesitance degree of two HFEs $h_A(x_i)$ and $h_B(x_i)$, respectively.

Li [32] defined a hesitance degree based on the number of hesitant fuzzy values in hesitant fuzzy elements, and proposed a series of hesitant fuzzy distance measures.

Definition 6 [32]. Let H be an HFS on $X = \{x_1, x_2, \dots, x_n\}$. Then, the hesitance degree of x in H can be defined as follows:

$$h_L(h_H(x_i)) = 1 - \frac{1}{l(h_H(x_i))} \quad (9)$$

where $l(h_H(x_i))$ be the length of $h_H(x_i)$.

Therefore, the hesitance degree of the HFS H is defined as:

$$h_L(H) = \frac{1}{n} \sum_{i=1}^n h_L(h_H(x_i)) \quad (10)$$

Definition 7 Let M_1, M_2, \dots, M_m and B be a set of HFS on $X = \{x_1, x_2, \dots, x_n\}$, then for any M_k and M_t , $k, t = 1, 2, \dots, m$, the normalized Hamming distance including hesitance degree between M_k and M_t is defined as follows:

$$d_{hlh}(M_k, M_t) = \frac{1}{2n} \sum_{i=1}^n \left[|h_L(h_{M_k}(x_i)) - h_L(h_{M_t}(x_i))| + \frac{1}{l(x_i)} \sum_{j=1}^{l(x_i)} |h_{M_k}^{\sigma(j)}(x_i) - h_{M_t}^{\sigma(j)}(x_i)| \right] \quad (11)$$

the normalized Euclidean distance including hesitance degree between M_k and M_t is defined as follows:

$$d_{hle}(M_k, M_t) = \left[\frac{1}{2n} \sum_{i=1}^n \left(|h_L(h_{M_k}(x_i)) - h_L(h_{M_t}(x_i))|^2 + \frac{1}{l(x_i)} \sum_{j=1}^{l(x_i)} |h_{M_k}^{\sigma(j)}(x_i) - h_{M_t}^{\sigma(j)}(x_i)|^2 \right) \right]^{\frac{1}{2}} \quad (12)$$

the normalized generalized distance including hesitance degree between M_k and M_t is defined as follows:

$$d_{hlg}(M_k, M_t) = \left[\frac{1}{2n} \sum_{i=1}^n \left(|h_L(h_{M_k}(x_i)) - h_L(h_{M_t}(x_i))|^\lambda + \frac{1}{l(x_i)} \sum_{j=1}^{l(x_i)} |h_{M_k}^{\sigma(j)}(x_i) - h_{M_t}^{\sigma(j)}(x_i)|^\lambda \right) \right]^{\frac{1}{\lambda}} \quad (13)$$

where $\lambda \geq 1$, $l(x_i) = \max\{l(h_{M_k}(x_i)), l(h_{M_t}(x_i))\}$, $h_{M_k}^{\sigma(j)}(x_i)$ and $h_{M_t}^{\sigma(j)}(x_i)$ are the j th values in $h_{M_k}(x_i)$ and $h_{M_t}(x_i)$, respectively.

In order to relax the limitation that the corresponding hesitant fuzzy elements have the same length. Tang et al. [31] proposed a series of distance measures.

Definition 8 Let A and B be two HFSs on X . Then, the hesitant normalized Hamming distance between A and B is defined as:

$$d_{lth}(A, B) = \frac{1}{n} \sum_{i=1}^n \frac{\sum_{j=1}^{l_A(x_i)} \sum_{k=1}^{l_B(x_i)} |h_A^{\sigma(j)}(x_i) - h_B^{\sigma(k)}(x_i)|}{l_A(x_i) l_B(x_i)} \quad (14)$$

the normalized Euclidean distance between A and B is defined as follows:

$$d_{lte}(A, B) = \left[\frac{1}{n} \sum_{i=1}^n \frac{\sum_{j=1}^{l_A(x_i)} \sum_{k=1}^{l_B(x_i)} (h_A^{\sigma(j)}(x_i) - h_B^{\sigma(k)}(x_i))^2}{l_A(x_i) l_B(x_i)} \right]^{\frac{1}{2}} \quad (15)$$

the normalized generalized distance between A and B is defined as follows:

$$d_{ltg}(A, B) = \left[\frac{1}{n} \sum_{i=1}^n \frac{\sum_{j=1}^{l_A(x_i)} \sum_{k=1}^{l_B(x_i)} |h_A^{\sigma(j)}(x_i) - h_B^{\sigma(k)}(x_i)|^\lambda}{l_A(x_i) l_B(x_i)} \right]^{\frac{1}{\lambda}} \quad (16)$$

where $\lambda > 0$, $h_A^{\sigma(j)}(x_i)$ are the j th values in $h_A(x_i)$ and $h_B^{\sigma(k)}(x_i)$ are the k th values in $h_B(x_i)$, $l_A(x_i)$ and $l_B(x_i)$ are the lengths of $h_A(x_i)$ and $h_B(x_i)$, respectively.

Some New hesitant fuzzy distance measures

According to analysis, the existing method only considers the number or the degree of dispersion, which is obviously one-sided. Therefore, by simultaneously considering them, we propose a new hesitance degree as follows.

Definition 9 Let A be an HFS in a reference set $X = \{x_1, x_2, \dots, x_n\}$, denoted by $A = \{ \langle x_i, h_A(x_i) \rangle \mid x_i \in X \}$ and $h_A(x_i) = \{ h_A^{\sigma(j)}(x_i) \mid j = 1, 2, \dots, l_{h_A} \}$. Then, the hesitance degree of x in A can be defined as follows:

$$h(h_A(x_i)) = \begin{cases} \frac{2\theta}{l_{h_A} \times (l_{h_A} - 1)} \sum_{k>j=1}^{l_{h_A}} |h_A^{\sigma(k)}(x_i) - h_A^{\sigma(j)}(x_i)| + \mu g (l_{h_A} - 1) & l_{h_A} > 1 \\ 0 & l_{h_A} = 1 \end{cases} \quad (17)$$

where l_{h_A} is the length of $h_A(x_i)$, g is the minimum accuracy of values in the hesitant fuzzy element $h_A(x_i)$, $\theta, \mu \in [0, 1]$, $\theta + \mu = 1$.

Therefore, the hesitance degree of the HFS A is defined as:

$$h(A) = \frac{1}{n} \sum_{i=1}^n h(h_A(x_i)) \quad (18)$$

Remark 1 n is the number of digits after the decimal point of the hesitant fuzzy element, then $g = 1/10^n$. For example, let $h = \{0.2, 0.3\}$ be a hesitant fuzzy element, then the minimum accuracy $g = 1/10 = 0.1$. If $h = \{0.25, 0.36\}$, then the minimum accuracy $g = 1/10^2 = 0.01$.

Next, we use a numerical example to illustrate the advantages of the proposed hesitance degree in processing hesitation fuzzy information.

Example 1 Let $h_1 = \{0.3, 0.5\}$, $h_2 = \{0.5, 0.6\}$ and $h_3 = \{0.3, 0.5, 0.6\}$ be three hesitant fuzzy elements, $g = 0.1$, $\theta = \mu = 0.5$. Then, their hesitance degrees are calculated by the different formulas respectively.

the result calculated by formula (5) is as follows:

$$h_Z(h_1) = 0.2, h_Z(h_2) = 0.1, h_Z(h_3) = 0.2$$

the result calculated by formula (9) is as follows:

$$h_L(h_1) = 0.5, h_L(h_2) = 0.5, h_L(h_3) = 0.66$$

the result calculated by formula (17) is as follows:

$$h(h_1) = 0.15, h(h_2) = 0.1, h(h_3) = 0.2$$

From the above results, we can find that $h_Z(\{0.3, 0.5\}) = h_Z(h_1) = h_Z(h_3) = h_Z(\{0.3, 0.5, 0.6\})$ and $h_L(\{0.3, 0.5\}) = h_L(h_1) = h_L(h_2) = h_L(\{0.5, 0.6\})$. Obviously, the results calculated by formula (5) and formula (9) are unreasonable. However, $h(h_1) \neq h(h_2) \neq h(h_3)$. That is to say the proposed hesitance degree can clearly distinguish the hesitance degrees of hesitant fuzzy elements h_1 , h_2 and h_3 , which is consistent with people's intuitive feeling. Therefore, the proposed hesitance degree is more reasonable than the existing hesitance degree mentioned above.

Based on the proposed hesitance degree, we propose some new distance measures, which can compare HFEs of equal or unequal length, so we can avoid destroying the original information by adding elements when the length is unequal.

Definition 10 Let $h_A(x_i)$ and $h_B(x_j)$ be two HFEs. Then, the normalized Hamming distance between $h_A(x_i)$ and $h_B(x_j)$ is defined as:

$$d_{hllh}(h_A, h_B) = \begin{cases} \frac{\alpha}{l_{h_A} l_{h_B}} \sum_{j=1}^{l_{h_B}} \sum_{i=1}^{l_{h_A}} |h_A(x_i) - h_B(x_j)| & l_{h_A} \neq l_{h_B} \\ +\beta |h(h_A) - h(h_B)| \\ \frac{\alpha}{l_{h_A}} \sum_{i=1}^{l_{h_A}} |h_A^{\sigma(j)}(x_i) - h_B^{\sigma(j)}(x_i)| & l_{h_A} = l_{h_B} \\ +\beta |h(h_A) - h(h_B)| \end{cases} \quad (19)$$

The normalized Euclidean distance between $h_A(x_i)$ and $h_B(x_j)$ is defined as:

$$d_{hlll}(h_A, h_B) = \begin{cases} \left[\frac{\alpha}{l_{h_A} l_{h_B}} \sum_{j=1}^{l_{h_B}} \sum_{i=1}^{l_{h_A}} (h_A(x_i) - h_B(x_j))^2 \right. & l_{h_A} \neq l_{h_B} \\ \left. +\beta (h(h_A) - h(h_B))^2 \right]^{1/2} \\ \left[\frac{\alpha}{l_{h_A}} \sum_{i=1}^{l_{h_A}} (h_A^{\sigma(j)}(x_i) - h_B^{\sigma(j)}(x_i))^2 \right. & l_{h_A} = l_{h_B} \\ \left. +\beta (h(h_A) - h(h_B))^2 \right]^{1/2} \end{cases} \quad (20)$$

The Hausdorff metric distance between $h_A(x_i)$ and $h_B(x_j)$ is defined as:

$$d_{hllh}(h_A, h_B) = \begin{cases} \max_{i,j} (\alpha |h_A(x_i) - h_B(x_j)| + \beta |h(h_A) - h(h_B)|) & l_{h_A} \neq l_{h_B} \\ \max_i (\alpha |h_A^{\sigma(j)}(x_i) - h_B^{\sigma(j)}(x_i)| + \beta |h(h_A) - h(h_B)|) & l_{h_A} = l_{h_B} \end{cases} \quad (21)$$

The normalized generalized distance between $h_A(x_i)$ and $h_B(x_j)$ is defined as:

$$d_{hllg}(h_A, h_B) = \begin{cases} \left[\frac{\alpha}{l_{h_A} l_{h_B}} \sum_{j=1}^{l_{h_B}} \sum_{i=1}^{l_{h_A}} |h_A(x_i) - h_B(x_j)|^\lambda + \beta |h(h_A) - h(h_B)|^\lambda \right]^{1/\lambda} & l_{h_A} \neq l_{h_B} \\ \left[\frac{\alpha}{l_{h_A}} \sum_{i=1}^{l_{h_A}} |h_A^{\sigma(j)}(x_i) - h_B^{\sigma(j)}(x_i)|^\lambda + \beta |h(h_A) - h(h_B)|^\lambda \right]^{1/\lambda} & l_{h_A} = l_{h_B} \end{cases} \quad (22)$$

where $\lambda > 0$, $\alpha, \beta \in [0, 1]$, $\alpha + \beta = 1$, l_{h_A} and l_{h_B} are the lengths of HFEs $h_A(x_i)$ and $h_B(x_j)$, respectively.

Especially, if $\lambda = 1$, then formula (22) degenerates to formula (21). If $\lambda = 2$, then formula (22) degenerates to formula (20). If $\lambda \rightarrow \infty$, then formula (22) degenerates to formula (21).

Example 2 Let $h_1 = \{0.3, 0.5\}$, $h_2 = \{0.5, 0.6\}$ and $h_3 = \{0.3, 0.5, 0.6\}$ be three hesitant fuzzy elements, $g = 0.1$, $\theta = \mu = 0.5$. Then, the process of calculating the Hamming distance between HFEs is as follows

$$\begin{aligned} d(h_1, h_2) &= \frac{\alpha}{2} (|0.3 - 0.5| + |0.5 - 0.6|) + \beta |0.15 - 0.1| \\ d(h_2, h_3) &= \frac{\alpha}{2 \times 3} (|0.5 - 0.3| + |0.5 - 0.5| + |0.5 - 0.6| + |0.6 - 0.3| \\ &\quad + |0.6 - 0.5| + |0.6 - 0.6|) + \beta |0.1 - 0.2| \end{aligned}$$

Lemma 1 (Minkowski's inequality [34]). Let $(a_1, a_2, \dots, a_n), (b_1, b_2, \dots, b_n) \in R^n$, and $1 \leq p < \infty$. Then

$$\left(\sum_{k=1}^n |a_k + b_k|^p \right)^{\frac{1}{p}} \leq \left(\sum_{k=1}^n |a_k|^p \right)^{\frac{1}{p}} + \left(\sum_{k=1}^n |b_k|^p \right)^{\frac{1}{p}} \quad (23)$$

Theorem 1 Let $\{A_1, A_2, \dots, A_m\}$ be a set of HFS on $X = \{x_1, x_2, \dots, x_n\}$, $I = \{1, 2, \dots, m\}$, $k, t \in I$. Then, $d_{hllh}(A_k, A_t)$, $d_{hllg}(A_k, A_t)$, and $d_{hllg}(A_k, A_t)$ are hesitant fuzzy distances.

Proof As d_{hllh} , d_{hllg} and d_{hd} are the special cases of d_{hllg} , here we only prove that d_{hllg} is a distance measure. According to Definition 10, it can be obtained easily that Property

(1) and Property (2) in Definition 2 hold. In the following, we prove that Property (3) and Property (4) hold. \square

(1) By Definition 10, We have

$$\begin{aligned}
 & d_{hllg}(h_A, h_B) \\
 &= \begin{cases} \left[\frac{\alpha}{l_{h_A} l_{h_B}} \sum_{j=1}^{l_{h_B}} \sum_{i=1}^{l_{h_A}} |h_A(x_i) - h_B(x_j)|^\lambda + \beta |h(h_A) - h(h_B)|^\lambda \right]^{1/\lambda} & l_{h_A} \neq l_{h_B} \\ \left[\frac{\alpha}{l_{h_A}} \sum_{i=1}^{l_{h_A}} |h_A^{\sigma(j)}(x_i) - h_B^{\sigma(j)}(x_i)|^\lambda + \beta |h(h_A) - h(h_B)|^\lambda \right]^{1/\lambda} & l_{h_A} = l_{h_B} \end{cases} \\
 &= \begin{cases} \left[\frac{\alpha}{l_{h_A} l_{h_B}} \sum_{i=1}^{l_{h_A}} \sum_{j=1}^{l_{h_B}} |h_B(x_j) - h_A(x_i)|^\lambda + \beta |h(h_B) - h(h_A)|^\lambda \right]^{1/\lambda} & l_{h_A} \neq l_{h_B} \\ \left[\frac{\alpha}{l_{h_A}} \sum_{j=1}^{l_{h_A}} |h_B^{\sigma(j)}(x_i) - h_A^{\sigma(j)}(x_i)|^\lambda + \beta |h(h_B) - h(h_A)|^\lambda \right]^{1/\lambda} & l_{h_A} = l_{h_B} \end{cases} \\
 &= d_{hllg}(h_B, h_A)
 \end{aligned}$$

Thus, $d_{hllg}(h_A, h_B) = d_{hllg}(h_B, h_A)$, i.e., Property (3) holds.

(2) Property (4) is $d(A, B) + d(B, C) \geq d(A, C)$, By Definition 10, it can be equivalently transformed into the following inequality:

$$\begin{aligned}
 & \begin{cases} \left[\frac{\alpha}{l_{h_A} l_{h_B}} \sum_{j=1}^{l_{h_B}} \sum_{i=1}^{l_{h_A}} |h_A(x_i) - h_B(x_j)|^\lambda + \beta |h(h_A) - h(h_B)|^\lambda \right]^{1/\lambda} & l_{h_A} \neq l_{h_B} \\ \left[\frac{\alpha}{l_{h_A}} \sum_{i=1}^{l_{h_A}} |h_A^{\sigma(j)}(x_i) - h_B^{\sigma(j)}(x_i)|^\lambda + \beta |h(h_A) - h(h_B)|^\lambda \right]^{1/\lambda} & l_{h_A} = l_{h_B} \end{cases} \\
 &+ \begin{cases} \left[\frac{\alpha}{l_{h_B} l_{h_C}} \sum_{k=1}^{l_{h_C}} \sum_{j=1}^{l_{h_B}} |h_B(x_j) - h_C(x_k)|^\lambda + \beta |h(h_B) - h(h_C)|^\lambda \right]^{1/\lambda} & l_{h_B} \neq l_{h_C} \\ \left[\frac{\alpha}{l_{h_B}} \sum_{j=1}^{l_{h_B}} |h_B^{\sigma(j)}(x_i) - h_C^{\sigma(j)}(x_i)|^\lambda + \beta |h(h_B) - h(h_C)|^\lambda \right]^{1/\lambda} & l_{h_B} = l_{h_C} \end{cases} \\
 &\geq \begin{cases} \left[\frac{\alpha}{l_{h_A} l_{h_C}} \sum_{k=1}^{l_{h_C}} \sum_{i=1}^{l_{h_A}} |h_A(x_i) - h_C(x_k)|^\lambda + \beta |h(h_A) - h(h_C)|^\lambda \right]^{1/\lambda} & l_{h_A} \neq l_{h_C} \\ \left[\frac{\alpha}{l_{h_A}} \sum_{i=1}^{l_{h_A}} |h_A^{\sigma(j)}(x_i) - h_C^{\sigma(j)}(x_i)|^\lambda + \beta |h(h_A) - h(h_C)|^\lambda \right]^{1/\lambda} & l_{h_A} = l_{h_C} \end{cases}
 \end{aligned}$$

which can be further converted into:

$$\begin{aligned}
 & \left\{ \left[\frac{\alpha l_{h_C}}{l_{h_A} l_{h_B} l_{h_C}} \sum_{j=1}^{l_{h_B}} \sum_{i=1}^{l_{h_A}} |h_A(x_i) - h_B(x_j)|^\lambda + \beta |h(h_A) - h(h_B)|^\lambda \right]^{1/\lambda} \right. & l_{h_A} \neq l_{h_B} \\
 & \left. \left[\frac{\alpha}{l_{h_A}} \sum_{i=1}^{l_{h_A}} |h_A^{\sigma(j)}(x_i) - h_B^{\sigma(j)}(x_i)|^\lambda + \beta |h(h_A) - h(h_B)|^\lambda \right]^{1/\lambda} \right. & l_{h_A} = l_{h_B} \\
 & + \left\{ \left[\frac{\alpha l_{h_A}}{l_{h_A} l_{h_B} l_{h_C}} \sum_{k=1}^{l_{h_C}} \sum_{j=1}^{l_{h_B}} |h_B(x_j) - h_C(x_k)|^\lambda + \beta |h(h_B) - h(h_C)|^\lambda \right]^{1/\lambda} \right. & l_{h_B} \neq l_{h_C} \\
 & \left. \left[\frac{\alpha}{l_{h_B}} \sum_{j=1}^{l_{h_B}} |h_B^{\sigma(j)}(x_i) - h_C^{\sigma(j)}(x_i)|^\lambda + \beta |h(h_B) - h(h_C)|^\lambda \right]^{1/\lambda} \right. & l_{h_B} = l_{h_C} \\
 & \geq \left\{ \left[\frac{\alpha l_{h_B}}{l_{h_A} l_{h_B} l_{h_C}} \sum_{k=1}^{l_{h_C}} \sum_{i=1}^{l_{h_A}} |h_A(x_i) - h_C(x_k)|^\lambda + \beta |h(h_A) - h(h_C)|^\lambda \right]^{1/\lambda} \right. & l_{h_A} \neq l_{h_C} \\
 & \left. \left[\frac{\alpha}{l_{h_A}} \sum_{i=1}^{l_{h_A}} |h_A^{\sigma(j)}(x_i) - h_C^{\sigma(j)}(x_i)|^\lambda + \beta |h(h_A) - h(h_C)|^\lambda \right]^{1/\lambda} \right. & l_{h_A} = l_{h_C}
 \end{aligned}$$

i.e.,

$$\begin{aligned}
 & \left\{ \left[\frac{\alpha}{l_{h_A} l_{h_B} l_{h_C}} \sum_{k=1}^{l_{h_C}} \sum_{j=1}^{l_{h_B}} \sum_{i=1}^{l_{h_A}} |h_A(x_i) - h_B(x_j)|^\lambda + \beta |h(h_A) - h(h_B)|^\lambda \right]^{1/\lambda} \right. & l_{h_A} \neq l_{h_B} \\
 & \left. \left[\frac{\alpha}{l_{h_A}} \sum_{i=1}^{l_{h_A}} |h_A^{\sigma(j)}(x_i) - h_B^{\sigma(j)}(x_i)|^\lambda + \beta |h(h_A) - h(h_B)|^\lambda \right]^{1/\lambda} \right. & l_{h_A} = l_{h_B} \\
 & + \left\{ \left[\frac{\alpha}{l_{h_A} l_{h_B} l_{h_C}} \sum_{k=1}^{l_{h_C}} \sum_{j=1}^{l_{h_B}} \sum_{i=1}^{l_{h_A}} |h_B(x_j) - h_C(x_k)|^\lambda + \beta |h(h_B) - h(h_C)|^\lambda \right]^{1/\lambda} \right. & l_{h_B} \neq l_{h_C} \\
 & \left. \left[\frac{\alpha}{l_{h_B}} \sum_{j=1}^{l_{h_B}} |h_B^{\sigma(j)}(x_i) - h_C^{\sigma(j)}(x_i)|^\lambda + \beta |h(h_B) - h(h_C)|^\lambda \right]^{1/\lambda} \right. & l_{h_B} = l_{h_C} \\
 & \geq \left\{ \left[\frac{\alpha}{l_{h_A} l_{h_B} l_{h_C}} \sum_{k=1}^{l_{h_C}} \sum_{j=1}^{l_{h_B}} \sum_{i=1}^{l_{h_A}} |h_A(x_i) - h_C(x_k)|^\lambda + \beta |h(h_A) - h(h_C)|^\lambda \right]^{1/\lambda} \right. & l_{h_A} \neq l_{h_C} \\
 & \left. \left[\frac{\alpha}{l_{h_A}} \sum_{i=1}^{l_{h_A}} |h_A^{\sigma(j)}(x_i) - h_C^{\sigma(j)}(x_i)|^\lambda + \beta |h(h_A) - h(h_C)|^\lambda \right]^{1/\lambda} \right. & l_{h_A} = l_{h_C}
 \end{aligned}$$

Since the following equation holds:

$$|h_A(x_i) - h_C(x_k)| = |h_A(x_i) - h_B(x_j) + h_B(x_j) - h_C(x_k)| \quad (24)$$

On the condition that $1 \leq \lambda \leq +\infty$, we can reason from Lemma 1 that $d_{hllg}(A, C) \leq d_{hllg}(A, B) + d_{hllg}(B, C)$. Therefore, Property (4) is verified.

Thus, we complete the proof of Theorem 1.

Example 3 Let $h_1 = \{0.1, 0.5\}$, $h_2 = \{0.3, 0.8\}$, $h_3 = \{0.5, 0.6\}$, $h_4 = \{0.3, 0.5\}$ and $h_5 = \{0.3, 0.5, 0.6\}$ be five hesitant fuzzy elements, $g = 0.1$, $\theta = \mu = 0.5$, $\alpha = \beta = 0.5$. Then use different formulas to calculate the distance measures. The results are shown in Table 1.

From Table 1, it can be seen that $d_{hllh}(h_1, h_1) = 0$ and $d_{hllh}(h_1, h_2) \neq d_{hllh}(h_1, h_3) \neq d_{hllh}(h_1, h_4) \neq d_{hllh}(h_1, h_5)$, which is consistent with people's intuitive feeling. That means the results based on proposed distance measure is more reasonable than those of the above mentioned distance measures.

On the other hand, we compare the characteristics of the proposed distance measure with those of the existing distance measures. The results are shown in Table 2.

From Table 2, it can be seen that the proposed distance measure has all listed characteristics, but the mentioned distance measures do not have all of them. This means that the proposed distance measure is superior to the existing distance measures above in many complex situations.

Hesitant fuzzy clustering based on new distance measure

The description of clustering Algorithm

Recently, many studies focus on the clustering analysis of HFSs. Chen and Xu [35] focused on studied the clustering for hesitant fuzzy sets based on the K-means clustering algorithm, which uses the result of hierarchical clustering as the initial clusters. Zhang and Xu [36] proposed a novel hesitant fuzzy agglomerative hierarchical clustering algorithm. The algorithm considers each of the given HFSs as a unique cluster, and then compares each pair of the HFSs by using the weighted Hamming distance or the weighted Euclidean distance. The two clusters with smaller distance are jointed. Repeat the process until the desired number of clusters is achieved.

We focused on studied the hierarchical hesitant fuzzy K-means clustering algorithm, and using the new distance measure to calculate the distance between hesitant fuzzy sets. The specific steps of the hierarchical hesitant fuzzy K-means clustering algorithm are as follows:

step1. (Hierarchical clustering) Consider each hesitant fuzzy set $A_i (i = 1, 2, \dots, n)$ as an independent cluster $\{A_1\}, \{A_2\}, \dots, \{A_n\}$. Then calculate the distance between A_i and A_j , which is denoted by $d_{ij} = d(A_i, A_j)$. The two clusters with smaller distance are jointed by average function, which is given as follows:

$$f(A_1, A_2) = \frac{1}{2} (A_1 \oplus A_2) = \left\{ \langle x_i, \cup_{r_1 \in h_{A_1}(x_i), r_2 \in h_{A_2}(x_i)} \{1 - [(1 - r_1)(1 - r_2)]^{1/2}\} \rangle \mid x_i \in X \right\} \quad (25)$$

This iterative process is repeated until all clusters are aggregated into one cluster.

Table 1 Comparison of different distance measures

	(h_1, h_1)	(h_1, h_2)	(h_1, h_3)	(h_1, h_4)	(h_1, h_5)
d_{xh}	0	0.25	0.25	0.1	0.1
d_{hlh}	0	0.125	0.125	0.05	0.135
d_{hzh}	0	0.175	0.275	0.15	0.15
d_{lth}	0.2	0.35	0.25	0.2	0.23
d_{hllh}	0	0.15	0.2	0.1	0.0575

Table 2 The characteristic comparisons with existing distance measures

Distance measure	Whether to consider hesitant degree	Whether to relax the limitation of lengths of HFEs
Xu and Xia [28]	×	×
Li et al. [32]	✓	×
Zhang and Xu [18]	✓	×
Tang et al. [31]	×	✓
The proposed distance	✓	✓

step2. According to the given number of clusters, select the corresponding result in step 1 as the initial cluster, then calculate the distance between the hesitant fuzzy set $A_i (i = 1, 2, \dots, n)$ and the center of each cluster. Finally classify A_i to the cluster with the closest cluster center.

step3. Recalculate the new cluster center through the average function of the hesitant fuzzy set.

step4. Repeat steps 2 and 3 until all cluster centers are stable.

Illustrative example

A specific example (adapted from Ref. [35]) is given below to illustrate the above algorithm. The proposed hesitant fuzzy distance is applied to the hierarchical hesitant fuzzy K-means clustering algorithm.

There are five tourism resources need to be evaluated and classified. Experts give corresponding evaluation information ($g = 0.1, \theta = \mu = 0.5, \alpha = \beta = 0.5$) to tourism resources from six aspects, namely: scale, environmental conditions, integrity, service, tourist routes and convenient transportation, which is expressed as $X = \{x_1, x_2, \dots, x_6\}$, the evaluation information of the five tourism resources is represented by hesitant fuzzy sets $A_i (i = 1, 2, 3, 4, 5)$, which are listed in Table 3:

step1. Consider each hesitating fuzzy set $A_i (i = 1, 2, 3, 4, 5)$ as an independent cluster: $\{A_1\}, \{A_2\}, \{A_3\}, \{A_4\}$ and $\{A_5\}$. Using the formula 21 calculate the distance between each hesitant fuzzy set and the other four hesitant fuzzy sets:

$$\begin{aligned}
 d(A_1, A_2) &= 0.3326, d(A_1, A_3) = 0.2473 \\
 d(A_1, A_4) &= 0.2256, d(A_1, A_5) = 0.4590 \\
 d(A_2, A_3) &= 0.1797, d(A_2, A_4) = 0.3444 \\
 d(A_2, A_5) &= 0.1955, d(A_3, A_4) = 0.1845 \\
 d(A_3, A_5) &= 0.2293, d(A_4, A_5) = 0.3052
 \end{aligned}$$

Obviously, $\{A_2\}$ and $\{A_3\}$ are the two closest clusters, then calculate the new cluster $\{A_2, A_3\}$ by formula (25). Therefore, the hesitant fuzzy set $A_i (i = 1, 2, 3, 4, 5)$ is divided

Table 3 Hesitant fuzzy assessment information

	x_1	x_2	x_3	x_4	x_5	x_6
A_1	{0.8, 1}	{0.3, 0.4}	{0.4, 0.5, 0.6}	{0.1, 0.3}	{1}	{0.9, 1}
A_2	{0.2, 0.3}	{0.2, 0.4}	{0.2, 0.3}	{0.3, 0.4}	{0.1, 0.3, 0.5}	{0.2, 0.3}
A_3	{0.4, 0.5, 0.6}	{0.4, 0.6}	{0.5, 0.6}	{0.4, 0.5, 0.6}	{0.5}	{0.4, 0.5}
A_4	{0.7, 0.9}	{0.6, 0.8}	{0.7, 0.8}	{0.6, 0.8}	{0.6, 0.7, 0.8}	{0.7, 0.8}
A_5	{0.1, 0.4}	{0.7, 0.8, 0.9}	{0.1, 0.2}	{0.8, 1}	{0.2, 0.3}	{0.2, 0.4}

into the following four clusters: $\{A_1\}$, $\{A_2, A_3\}$, $\{A_4\}$ and $\{A_5\}$. Continue to calculate the distance between each cluster and the other three clusters:

$$\begin{aligned}d(\{A_2, A_3\}, A_1) &= 0.2881 \\d(\{A_2, A_3\}, A_4) &= 0.2644 \\d(\{A_2, A_3\}, A_5) &= 0.2097 \\d(A_1, A_4) &= 0.2256, d(A_1, A_5) = 0.4950 \\d(A_4, A_5) &= 0.3052\end{aligned}$$

Because of $\{A_2, A_3\}$ and $\{A_5\}$ are the two closest clusters, then the hesitant fuzzy sets are divided into the following three clusters: $\{A_2, A_3, A_5\}$, $\{A_1\}$ and $\{A_4\}$. Calculate the new cluster and the distances between each cluster and the other clusters:

$$\begin{aligned}d(A_1, \{A_2, A_3, A_5\}) &= 0.4341 \\d(A_4, \{A_2, A_3, A_5\}) &= 0.3085 \\d(A_4, A_1) &= 0.2256\end{aligned}$$

Where $\{A_1\}$ and $\{A_4\}$ are the two closest clusters, then the hesitant fuzzy sets are divided into two clusters: $\{A_2, A_3, A_5\}$ and $\{A_1, A_4\}$.

In the end, the two clusters merged into one cluster: $\{A_1, A_2, A_3, A_4, A_5\}$.

step2. Assuming number of clusters $c = 3$ is given, according to the result of step1, then $c_1 = \{A_1\}$, $c_2 = \{A_2, A_3, A_5\}$ and $c_3 = \{A_4\}$ are selected as the initial clusters. Next, calculate the distances of each hesitant fuzzy set $A_i (i = 1, 2, \dots, 5)$ between each initial cluster $c_j (j = 1, 2, 3)$ as follows:

$$\begin{aligned}d(A_1, c_1) &= 0, d(A_1, c_2) = 0.4310 \\d(A_1, c_3) &= 0.2806, d(A_2, c_1) = 0.4149 \\d(A_2, c_2) &= 0.2122, d(A_2, c_3) = 0.4500 \\d(A_3, c_1) &= 0.3139, d(A_3, c_2) = 0.1792, \\d(A_3, c_3) &= 0.2222, d(A_4, c_1) = 0.2806 \\d(A_4, c_2) &= 0.2948, d(A_4, c_3) = 0 \\d(A_5, c_1) &= 0.5972, d(A_5, c_2) = 0.1365 \\d(A_5, c_3) &= 0.4000\end{aligned}$$

According to the above calculation results, the clustering result is $c_1 = \{A_1\}$, $c_2 = \{A_2, A_3, A_5\}$ and $c_3 = \{A_4\}$.

step3. The cluster center remains unchanged and the iteration ends.

Comparative analysis

In order to illustrate the performance of the proposed method, we make a comparative analysis with the hierarchical hesitant fuzzy k-means clustering algorithm introduced by Chen et al [35].

Consider each hesitating fuzzy set $A_i (i = 1, 2, 3, 4, 5)$ as an independent cluster: $\{A_1\}$, $\{A_2\}$, $\{A_3\}$, $\{A_4\}$ and $\{A_5\}$. Calculating the distance between each hesitant fuzzy set and the other four hesitant fuzzy sets:

$$\begin{aligned}d(A_1, A_2) &= 0.4194, d(A_1, A_3) = 0.3139 \\d(A_1, A_4) &= 0.2806, d(A_1, A_5) = 0.5972 \\d(A_2, A_3) &= 0.2222, d(A_2, A_4) = 0.4500 \\d(A_2, A_5) &= 0.2444, d(A_3, A_4) = 0.2222 \\d(A_3, A_5) &= 0.3000, d(A_4, A_5) = 0.4000\end{aligned}$$

We can find $d(A_2, A_3) = d(A_3, A_4) = \min \{d(A_i, A_j) \mid i, j = 1, 2, 3, 4, 5 (i \neq j)\} = 0.2222$, there are two options when merging the two clusters into a new cluster. Therefore, the following two cases are considered.

case1: Hesitant fuzzy sets $A_i (i = 1, 2, 3, 4, 5)$ are divided into the following four clusters: $\{A_1\}, \{A_2, A_3\}, \{A_4\}$ and $\{A_5\}$. Calculate the distances between each cluster and the other three clusters. We have $d(\{A_2, A_3\}, A_5)$ is the shortest distance. Merging $\{A_2, A_3\}$ and $\{A_5\}$ into a new cluster, the hesitant fuzzy sets are divided into three clusters: $\{A_2, A_3, A_5\}, \{A_1\}$ and $\{A_4\}$. Calculate the new cluster and the distances between each cluster and the other clusters. We have $d(A_1, A_4)$ is the shortest distance. Therefore, hesitant fuzzy sets are divided into the following two clusters: $\{A_2, A_3, A_5\}$ and $\{A_1, A_4\}$. In the end, the two clusters are merged into one cluster: $\{A_1, A_2, A_3, A_4, A_5\}$.

case2: Hesitant fuzzy sets $A_i (i = 1, 2, 3, 4, 5)$ are divided into the following four clusters: $\{A_1\}, \{A_2\}, \{A_3, A_4\}$ and $\{A_5\}$. Calculate the distance between each cluster and the other three clusters. We have $d(A_2, A_5)$ is the shortest distance. Merging $\{A_2\}$ and $\{A_5\}$ into a new cluster, the hesitant fuzzy set is divided into three clusters: $\{A_1\}, \{A_3, A_4\}$ and $\{A_2, A_5\}$. Calculate the new cluster and the distances between each cluster and the other clusters. We have $d(\{A_3, A_4\}, \{A_2, A_5\})$ is the shortest distance. Therefore, hesitant fuzzy sets are divided into two clusters: $\{A_1\}$ and $\{A_2, A_3, A_4, A_5\}$. In the end, the two clusters are merged into one cluster: $\{A_1, A_2, A_3, A_4, A_5\}$.

Obviously, the clustering results obtained in different cases are different. Next, we analyze the quality of the clustering results of the two cases. Generally, the average distance d_ρ is an indicator to measure the quality of clustering results. The smaller the d_ρ , the better the clustering result. The calculation process is as follows:

$$d(\{A_2, A_3\}, A_2) = 0.1531, d(\{A_2, A_3\}, A_3) = 0.0714$$

$$d_\rho(\{A_2, A_3\}) = \frac{0.1531 + 0.0714}{2} = 0.1123$$

$$d(\{A_3, A_4\}, A_3) = 0.1750, d(\{A_3, A_4\}, A_4) = 0.1069$$

$$d_\rho(\{A_3, A_4\}) = \frac{0.1750 + 0.1069}{2} = 0.1410$$

It can be seen that $d_\rho(\{A_2, A_3\})$ is smaller than $d_\rho(\{A_3, A_4\})$. Therefore, the clustering result of case1 is better than case2.

Results and Discussion

According to the above analysis, the comparison result is shown in Table 4.

From Table 4, we can find that there are two different clustering results using Chen's method introduced in [35]. It is very difficult to decide which one to choose in the clustering process. And even if it can be selected correctly, it will increase the complexity of the algorithm. However, a unique clustering result can be obtained by the proposed method.

Table 4 Comparison of hierarchical clustering results

Category	Ref [35]		The proposed method
	case1	case2	
1	$\{A_1, A_2, A_3, A_4, A_5\}$	$\{A_1, A_2, A_3, A_4, A_5\}$	$\{A_1, A_2, A_3, A_4, A_5\}$
2	$\{A_2, A_3, A_5\}, \{A_1, A_4\}$	$\{A_1\}, \{A_2, A_3, A_4, A_5\}$	$\{A_2, A_3, A_5\}, \{A_1, A_4\}$
3	$\{A_1\}, \{A_4\}, \{A_2, A_3, A_5\}$	$\{A_1\}, \{A_3, A_4\}, \{A_2, A_5\}$	$\{A_1\}, \{A_4\}, \{A_2, A_3, A_5\}$
4	$\{A_1\}, \{A_2, A_3\}, \{A_4\}, \{A_5\}$	$\{A_1\}, \{A_2\}, \{A_3, A_4\}, \{A_5\}$	$\{A_1\}, \{A_2, A_3\}, \{A_4\}, \{A_5\}$
5	$\{A_1\}, \{A_2\}, \{A_3\}, \{A_4\}, \{A_5\}$	$\{A_1\}, \{A_2\}, \{A_3\}, \{A_4\}, \{A_5\}$	$\{A_1\}, \{A_2\}, \{A_3\}, \{A_4\}, \{A_5\}$

And the result is same as the best one obtained by Chen's method. Therefore, the hierarchical hesitant fuzzy k-means clustering method based on the proposed distance measure is more reasonable and effective.

Conclusions

Considering the existing hesitance degrees does not take into account both degree of dispersion and number of the hesitant fuzzy values in the hesitant fuzzy element, a new hesitance degree is defined in this paper, which has better accuracy and applicability. We have elaborated the important role of hesitance degree in hesitant fuzzy distance measure. Further, we proposed some hesitant fuzzy distance measures based on the new hesitance degree, which can overcome the shortcomings of the existing distance measures. Moreover, we applied the new hesitant fuzzy distance to the hierarchical hesitant fuzzy k-means clustering algorithm, and presented an example to illustrate the effectiveness of the proposed method. In addition, we have compared and analyzed with the existing hierarchical hesitant fuzzy k-means clustering algorithm. It has been found that the clustering algorithm based on new distance measure is more reasonable. The proposed distance measure can avoid the original information distortion and have higher resolution. Therefore, it can help decision-makers get the only ideal results in practical problems. In the future, We will apply the proposed distance measure to multi-attribute group decision-making. We will extend this approach to interval valued environment. We will develop the knowledge measure [37] for hesitant fuzzy set.

Acknowledgements

Not applicable.

Declarations The authors declare that the manuscript is our original work. The ideas, views, innovations, and results presented in the above manuscript are totally ours unless otherwise referenced in the text. Works of others necessary for the development of ideas presented in this manuscript are clearly referenced within the text and accurately listed at its end. Also, the authors declare that they consent for publication.

Authors' contributions

Fuping Liao was a major contributor in writing the manuscript. Wu Li conducted experimental analysis and comparison. Both of them were the main authors of the manuscript. Xiaoqiang Zhou and Gang Liu corrected the grammar of the paper. All authors read and approved the final manuscript.

Funding

The work is supported by the Key Research and Development Project of Hunan Province (No. 2019SK2331), the Natural Science Foundation of Hunan Province (Nos: 2018JJ3213, 2019JJ40100, 2019JJ40099), the Key scientific research projects of Hunan Education Department (Nos: 18A317, 19A202) and the Innovation Foundation for Postgraduate of Hunan Institute of Science and Technology (No. YCX2020A34).

Availability of data and materials

All data generated or analysed during this study are included in this published article.

Declarations

Competing interests

The authors declare that they have no competing interests.

Received: 17 December 2021 Accepted: 30 March 2022

Published online: 19 December 2022

References

1. Zadeh LA (1965) Fuzzy sets. *Inf Control* 8(3):338–353
2. Atanassov KT (1986) Intuitionistic fuzzy sets. *Fuzzy Sets Syst* 20(1):87–96
3. Atanassov K, Gargov G (1989) Interval valued intuitionistic fuzzy sets. *Fuzzy Sets Syst* 31(3):343–349
4. Zadeh LA (1975) The concept of a linguistic variable and its application to approximate reasoning. *Inf ences* 8(3):199–249

5. Mizumoto M, Tanaka K (1976) Some properties of fuzzy sets of type 2. *Inf Control* 31(4):312–340
6. YAGER, Ronald R (1986) On the theory of bags. *Int J Gen Syst* 13(1):23–37
7. Mjka B, Pka B, Wd C, Wk D, Zsa B (2020) Bi-parametric distance and similarity measures of picture fuzzy sets and their applications in medical diagnosis. *Egypt Inf J* 22(2):201–212
8. Torra V (2010) Hesitant fuzzy sets. *Int J Intell Syst* 25(6):529–539
9. Zhang, Zhiming (2013) Hesitant fuzzy power aggregation operators and their application to multiple attribute group decision making. *Inf Sci* 234(Complete):150–181
10. Zhu B, Xu Z (2013) Hesitant fuzzy bonferroni means for multi-criteria decision making. *J Oper Res Soc* 64:1831–1840
11. Zhu B, Xu ZS (2012) Hesitant fuzzy geometric bonferroni means. *Inf Sci* 205(1):72–85
12. Wei G (2012) Hesitant fuzzy prioritized operators and their application to multiple attribute decision making. *Knowl-Based Syst* 31:176–182
13. Xu Z, Zhang X (2013) Hesitant fuzzy multi-attribute decision making based on topsis with incomplete weight information. *Knowl-Based Syst* 52(nov):53–64
14. Liao H, Xu Z (2013) A vikor-based method for hesitant fuzzy multi-criteria decision making. *Fuzzy Optim Decis Mak* 12(4):373–392
15. WANG Y, Cuiping Q, Yixin L (2017) Hesitant fuzzy topsis multi-attribute decision method based on prospect theory. *Control Decis* 32(5):864–870
16. Zhang X, Xu Z (2015) Hesitant fuzzy agglomerative hierarchical clustering algorithms. *Int J Syst Sci* 46(3):562–576
17. Liu X, Zhu J, Liu S (2014) Similarity measure of hesitant fuzzy sets based on symmetric cross entropy and its application in clustering analysis. *Control Decis* 29(10):1816–1822
18. Zhang X, Xu Z (2015) Novel distance and similarity measures on hesitant fuzzy sets with applications to clustering analysis. *J Intell Fuzzy Syst* 28(5):2279–2296
19. Chen N, Xu Z, Xia M (2014) Hierarchical hesitant fuzzy k-means clustering algorithm. *Appl Math J Chin Univ* 29(1):1–17
20. Akram M, Adeel A (2019) Novel topsis method for group decision-making based on hesitant m-polar fuzzy model. *J Intell Fuzzy Syst* 37(6):8077–8096
21. Akram M, Adeel A, Al-Kenani AN, Alcantud JCR (2020) Hesitant fuzzy n-soft electre-ii model: a new framework for decision-making. *Neural Comput Appl* 3:1–16
22. Deli I, Karaaslan F (2020) Generalized trapezoidal hesitant fuzzy numbers and their applications to multi criteria decision-making problems. *Soft Comput* 25(1):1017–1032
23. Karaaslan F, Özlü Ş (2019) Some distance measures for type-2 hesitant fuzzy sets and their applications to multi-criteria group decision making problems. *Soft Comput* 24(1):9965–9980
24. Su Z, Xu Z, Liu H, Liu S (2015) Distance and similarity measures for dual hesitant fuzzy sets and their applications in pattern recognition. *J Intell Fuzzy Syst* 29(2):731–745
25. Zeng W, Li D, Qian Y (2016) Distance and similarity measures between hesitant fuzzy sets and their application in pattern recognition 84:267–271
26. Zhang F, Chen S, Li J, Huang W (2018) New distance measures on hesitant fuzzy sets based on the cardinality theory and their application in pattern recognition. *Soft Comput Fusion Found Methodologies Appl* 22(4):1237–1245
27. Li C, Li D, Jin J (2019) *Int J Patt Recogn Artif Intell* 33(12):1–30
28. Xu Z, Xia M (2011) Distance and similarity measures for hesitant fuzzy sets. *Inf ences* 181(11):2128–2138
29. Tong X, Yu L (2016) Madm based on distance and correlation coefficient measures with decision-maker preferences under a hesitant fuzzy environment. *Soft Comput* 20(11):4449–4461
30. Peng D, Gao C, Gao Z (2013) Generalized hesitant fuzzy synergetic weighted distance measures and their application to multiple criteria decision-making. *Appl Math Modell* 37(8):5837–5850
31. Tang X, Peng Z, Ding H, Cheng M, Yang S, Li C, de Oliveira José Valente (2018) Novel distance and similarity measures for hesitant fuzzy sets and their applications to multiple attribute decision making. *J Intell Fuzzy Syst* 34(6):3903–3916
32. Li D, Zeng W, Zhao Y (2015) Note on distance measure of hesitant fuzzy sets. *Inf Sci* 321:103–115
33. Xia M, Xu Z (2011) Hesitant fuzzy information aggregation in decision making. *Int J Approx Reason* 52(3):395–407
34. Hatzimichailidis AG, Papakostas GA, Kaburlasos VG (2012) A novel distance measure of intuitionistic fuzzy sets and its application to pattern recognition problems. *Int J Intell Syst* 27:396–409
35. Chen N, Xu Z, Xia M (2014) Hierarchical hesitant fuzzy k-means clustering algorithm. *Appl Math J Chin Univ* 29(1):1–17
36. Zhang X, Xu Z (2015) Hesitant fuzzy agglomerative hierarchical clustering algorithms. *Int J Syst ence* 46(3):562–576
37. Khan MJ, Kumam P, Shutaywi M (2020) Knowledge measure for the q-rung orthopair fuzzy sets. *Int J Intell Syst* 36(2):628–655

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.