# An orthogonal clustering method under hesitant fuzzy environment 

Yanmin Liu<br>Institute of Sciences, PLA University of Science and Technology<br>Nanjing, Jiangsu 211101, China<br>E-mail:1298958905@qq.com<br>Hua Zhao<br>Institute of Sciences, PLA University of Science and Technology<br>Nanjing, Jiangsu 211101, China<br>E-mail: zhaohua_pla@163.com<br>Zeshui Xu*<br>Business School, Sichuan University, Chengdu, Sichuan 610064, China<br>E-mail: xuzeshui@263.net

Received 15 February 2016
Accepted 13 January 2017


#### Abstract

In this paper, we investigate the cluster techniques of hesitant fuzzy information. Consider that the distance measure is one of the most widely used tools in clustering analysis, we first point out the weakness of the existing distance measures for hesitant fuzzy sets (HFSS), and then put forward a novel distance measure for HFSs, which involves a new hesitation degree. Moreover, we construct the distance matrix and choose different values of $\lambda$ so as to obtain the $\lambda$-cutting matrix, each column of which is treated as a vector. After that, an orthogonal clustering method is developed for HFSs. The main idea of this clustering method is that the orthogonal vectors in the distance matrix should be clustered into the same group, and according to the different values of $\lambda$, the procedure will repeat again and again until all the cases are considered. Finally, two numerical examples are given to demonstrate the effectiveness of our algorithm.


Keywords: Hesitant fuzzy set; distance measure; clustering analysis; orthogonal method.
developed from various angles. In 1986, Atanassov extended fuzzy set to intuitionistic fuzzy set ${ }^{2}$, which takes account of the membership degree, the nonmembership degree and the hesitance degree. Compared to fuzzy set, it includes more details to distinguish different objects. Later, Atanassov and Gargov introduced the concept of interval-valued intuitionistic fuzzy set ${ }^{3}$, in which the membership degree and the non-membership degree are interval-valued. There are

## 1. Introduction

In our daily life, there exist some phenomena that cannot be described accurately in mathematical forms, for example, the words "fast", "big", "beautiful", "rich" and so on. It encourages people to find a more effective way to study and handle these uncertain problems. In 1965, Zadeh ${ }^{1}$ originally put forward the concept of fuzzy set, which opens the door of fuzzy theory research. Since then, fuzzy set theory has been
also some other kinds of fuzzy sets, such as type-2

[^0]fuzzy set ${ }^{4,5}$, interval type-2 fuzzy set ${ }^{6}$, type-n fuzzy set ${ }^{4}$, and hesitant fuzzy linguistic set ${ }^{7,8}$, etc.

The fuzzy sets mentioned above can solve a lot of decision making problems appropriately, but they do not do well when analyzing hesitant fuzzy information ${ }^{7,8}$. In some decision making problems, when considering the degree that an alternative satisfies a criterion, different experts have different opinions. Some experts who are optimistic may assign 0.9 , some experts may give 0.6 , and some 0.1 . Their attitudes cannot be changed and each opinion cannot be ignored. It is clear that this issue cannot be solved by using fuzzy sets talked before. Hesitant fuzzy set ${ }^{7,8}$ proposed by Torra and Narukawa can work out the issue more convincingly. Xia and $\mathrm{Xu}{ }^{9,10}$ gave the mathematical expression of the HFS, and called its components the hesitant fuzzy elements (HFEs). Obviously, utilizing the HFE $\{0.9,0.6,0.1\}$ to describe the above situation is much more reasonable than the interval-valued fuzzy set [ $0.1,0.9]$ or the single value. Therefore, it is necessary and essential to use the HFEs (or HFSs) to describe the hesitant information in the decision making problems. So far, a lot of research has been done to develop the hesitant fuzzy set theory. Torra et.al ${ }^{7,8}$ made a deep exploration about the difference among the HFS and other fuzzy sets, and gave some basic operations of HFSs, such as complement, union and intersection. Xia and $\mathrm{Xu}{ }^{9,10}$ developed some aggregation operators for HFEs, and also studied the distance, similarity and correlation measures of HFSs. Rodriguez et al. ${ }^{11}$ presented an overview on hesitant fuzzy set and pointed out the further research directions in the future. In the existing research, Xu and $\mathrm{Xia}{ }^{10}$ proposed several distance measures to discuss the variances and applied them to clustering analysis. In addition, the distance measures have been widely used in decision making ${ }^{12-}$ ${ }^{14}$, medical diagnosis ${ }^{15}$ and pattern recognition ${ }^{16}$, etc. However, the existing distance measures have some drawbacks, such as changing the original information and ignoring the hesitation degree. To overcome these drawbacks, in this paper, we develop a novel method to calculate the distance involving hesitation degrees and compare it with the existing distance measures, and finally, we apply the proposed distance measure to clustering analysis.

Clustering is a dividing process which divide the set of different kinds of objects into a few groups generally according to their characteristics, which has been widely used in various fields, such as economics, computer sciences, astronomy and so on ${ }^{17-19}$. The similar objects would be clustered into the same group. Based on the properties of the generated clusters, the clustering techniques are generally classified as the partitional clustering method and the hierarchical clustering method ${ }^{21}$. The partitional clustering
algorithm divides the data into several partitions based on certain objective function, where each partition represents a cluster, such as K-means clustering algorithm. While the hierarchical clustering algorithm gathers all the data to form a tree shaped structure, compare the distances or similarities between each pair of clusters in each layer, and form a new layer. Through continuous cycle, we can get the clustering results finally. Recently, some scholars have been giving research on hesitant fuzzy clustering techniques. Chen et al. ${ }^{22}$ constructed a correlation matrix by calculating the correlation coefficients for each pair of HFSs, then formed the correlation coefficients equivalent matrix, and finally clustered the HFSs based on the $\lambda$-cutting matrix. Zhang and $\mathrm{Xu}^{23}$ proposed a minimal spanning tree (MST) clustering technique, while drawing the MST is too complicated. Zhang and $\mathrm{Xu}^{21}$ adopted the traditional agglomerative hierarchical clustering method ${ }^{24}$ to calculate the center of the groups again and again, which needs too much calculational effort too. Chen et al. ${ }^{25}$ put forward a clustering method of HFSs based on K-means clustering algorithm which takes the results of hierarchical clustering as the initial input.

Looking into the clustering algorithms discussed above, we find that some need a large amount of computational effort, some need the complicated transformation, and most of the algorithms take a lot of time to finish clustering. To overcome these issues, in this paper, we will propose a novel orthogonal clustering method for HFSs. In this method, we first construct the distance matrix using our new distance measure for HFSs. After that, we choose the confidence level $\lambda$ to obtain the $\lambda$-cutting matrix, every column of which is treated as a vector. If two vectors are orthogonal, then we cluster them into the same group.

The remainder of the paper is organized as follows: Section 2 reviews some basic knowledge related to HFSs. Section 3 gives a novel method to calculate the distance and defines a new concept of hesitation degree. In Section 4, we put forward the orthogonal clustering method for HFSs. We illustrate the effectiveness of the method via two numerical examples in Section 5. The paper ends with some concluding remarks in Section 6.

## 2. Preliminaries

### 2.1. The basic knowledge related to HFSs

In some decision making problems, many experts are needed to express individual opinions on the same problem. When considering the degree that an alternative satisfies a criterion, different experts have different opinions, and thus, different people may assign
different values to the alternatives. To solve these problems, Torra ${ }^{7}$ generalized the fuzzy set to hesitant fuzzy set (HFS), in which the membership degree of an element to a set is expressed as several possible values between 0 and 1 .

Let $X=\left\{x_{1}, x_{2}, \cdots, x_{m}\right\}$ be a fixed set, a HFS $A$ on $X$ is defined as a function that when applied to $X$ returns a subset of $[0,1]$. Xia and $\mathrm{Xu}^{9}$ expressed it by a mathematical symbol as follows:

$$
A=\left\{\left\langle x, h_{A}(x) \mid x \in X\right\rangle\right\}
$$

where $x$ represents a criterion and $h_{A}(x)$ is a set of some values in $[0,1]$, on behalf of the possible membership degrees of the element $x$ to the set $A$. For convenience, Xu and Xia ${ }^{9}$ defined $h_{A}(x)$ as hesitant fuzzy element (HFE).

Example 1. Let $X=\left\{x_{1}, x_{2}, x_{3}\right\}$ be a fix set, $h_{A}\left(x_{1}\right)=\{0.1,0.2,0.3\} \quad, \quad h_{A}\left(x_{2}\right)=\{0.4,0.5\} \quad, \quad$ and $h_{A}\left(x_{3}\right)=\{0.3,0.5,0.6\}$ are the HFEs of $x_{i}(i=1,2,3)$ to a set $A$. Then the HFS $A$ can be expressed as:

$$
\begin{aligned}
& A=\left\{\left\langle x_{1},(0.1,0.2,0.3)\right\rangle,\left\langle x_{2},(0.4,0.5)\right\rangle,\right. \\
& \left.\left\langle x_{3},(0.3,0.5,0.6)\right\rangle\right\}
\end{aligned}
$$

To compare the HFEs, we introduce the score of the $\mathrm{HFE}^{9}$ :

$$
s(h)=\frac{1}{l_{h}} \sum_{\gamma \in h} \gamma
$$

where $l_{h}$ is the number of the elements in $h$, generally being called the length of $h$, and $\gamma$ is the elements in $h$. In fact, the score of the HFE is the average value of numbers in $h$. For two HFEs $h_{1}$ and $h_{2}$, if $s\left(h_{1}\right)>s\left(h_{2}\right)$, then $h_{1}$ is superior to $h_{2}$, denoted by $h_{1}>h_{2}$. However, if $s\left(h_{1}\right)=s\left(h_{2}\right)$, only with the score of HFEs, then we cannot distinguish which one is bigger. For example, $h_{1}=\{0.1,0.1,0.7\}$ and $h_{2}=\{0.2,0.4\}$ are two HFEs. We can easily get the scores of these two HFEs: $s\left(h_{1}\right)=0.3$ and $s\left(h_{2}\right)=0.3$. Since $s\left(h_{1}\right)=s\left(h_{2}\right)$, then we cannot distinguish these two HFEs with the score. Clearly, even if two HFEs $h_{1}$ and $h_{2}$ have the same score, their deviation degrees may be different. To better compare the HFEs, Chen et al. ${ }^{22}$ defined the concept of deviation degree: Let the HFE $h_{1}(x)=\left\{x_{1}, x_{2} \cdots x_{n}\right\}$, then the deviation degree of $h_{1}(x)$ is expressed as:

$$
\sigma(h)=\sqrt{\frac{\sum_{i=1}^{n}\left(x_{i}-\bar{x}\right)^{2}}{l_{h}}}
$$

where $\bar{x}=\frac{\sum_{i=1}^{n} x_{i}}{n}$. The deviation degree is just the standard variance of $x_{i}$.

Distance measure is an important content in clustering analysis with hesitant fuzzy information. Xu and $\mathrm{Xia}{ }^{10}$ gave its concept as follows:

Definition $1{ }^{\mathbf{1 0}}$. Let $A_{1}$ and $A_{2}$ be two HFSs defined on $X$, then the distance measure between $A_{1}$ and $A_{2}$ is defined as $d\left(A_{1}, A_{2}\right)$, which should satisfy:
(1) $0 \leq d\left(A_{1}, A_{2}\right) \leq 1$.
(2) $d\left(A_{1}, A_{2}\right)=0$ if and only if $A_{1}=A_{2}$.
(3) $d\left(A_{1}, A_{2}\right)=d\left(A_{2}, A_{1}\right)$.

Based on the above properties of distance measure between each pair of HFSs, we can construct the distance matrix below:
Definition $2{ }^{10}$. Let $A_{i}(i=1,2, \cdots, m)$ be $m$ HFSs, then $D=\left(d_{i j}\right)_{m \times m}$ is called a distance matrix, where $d_{i j}=d\left(A_{i}, A_{j}\right)$ is the distance measure between $A_{i}$ and $A_{j}$, and $d_{i j}$ should satisfy the following properties:
(1) $0 \leq d_{i j} \leq 1$ for all $i, j=1,2, \cdots, m$.
(2) $d\left(A_{i}, A_{j}\right)=0$ if and only if $A_{i}=A_{j}$.
(3) $d_{i j}=d_{j i}$ for all $i, j=1,2, \cdots, m$.

Definition $3{ }^{24}$. If $D=\left(d_{i j}\right)_{m \times m}$ is a distance matrix, then we define $D_{\lambda}=\left({ }_{\lambda} d_{i j}\right)_{m \times m}$ as the $\lambda$-cutting matrix of $D$, where

$$
{ }_{\lambda} d_{i j}=\left\{\begin{array}{ll}
0 & \text { if } d_{i j}<\lambda, \\
1 & \text { if } d_{i j} \geq \lambda,
\end{array} \quad i, j=1,2, \cdots, m\right.
$$

### 2.2. The existing distance measures for HFSs

Clustering is a progress which divides different kinds of elements into a few groups. Elements in the same group have something in common. On the contrary, the elements in different groups differ widely. So it is very important to find a suitable method to measure the relationship between different elements. Generally, we estimate this relation by distance, similarity or correlation coefficient. Only a good measure can lead to accurate results, in what follows, we will review some existing distance measures for HFSs.

In earlier research, a lot of approaches have been found to calculate the distance between any two HFSs. Let $h_{A}(x)$ and $h_{B}(x)$ be two HFEs. In most cases, the numbers of values in the HFEs $h_{A}(x)$ and $h_{B}(x)$ may be different. According to Ref. [10], the number of values in the $\operatorname{HFE} h_{A}\left(x_{i}\right)$ is called the length of the HFE $h_{A}\left(x_{i}\right)$, expressed as $l_{h_{A}\left(x_{i}\right)}$. That is to say, the lengths of the HFEs may be different. For convenience, Xu and $\mathrm{Xia}{ }^{10}$ put forward some rules as follows: Arrange the values in $h_{A}(x)$ and $h_{B}(x)$ in descending orders. It means that $h_{A}^{\sigma(j)}(x)$ represents the $j$ th largest element. The equation $h_{A}(x)=h_{B}(x)$ holds if and only if $h_{A}^{\sigma(j)}(x)=h_{B}^{\sigma(j)}(x), j=1,2, \cdots, m$. Generally, if $l_{h_{A}\left(x_{i}\right)}$ and $l_{l_{B}\left(x_{i}\right)}$ represent the lengths of $h_{A}\left(x_{i}\right)$ and $h_{B}\left(x_{i}\right)$, then $l_{x_{i}}=\max \left\{l_{h_{A}\left(x_{i}\right)}, l_{h_{B}\left(x_{i}\right)}\right\}$ for each $x_{i}$ in $X$. Only if they have the same length, can we continue to make deeper research. In the previous algorithms, if the lengths of two HFEs are different, then we add values to the HFEs which have less numbers until they have the same length. The principle of adding numbers reflects the risk preferences of the decision makers. The optimists may add the maximum value to the HFE, while the other who expect negative consequences may add the minimum number to the HFE. In order to avoid its weakness, in this paper, we will extend the shorter one by the maximum value, assuming that the decision makers are all optimistic. For example, there are two HFEs $h_{A}\left(x_{1}\right)=\{0.6,0.5,0.3,0.3,0.3\}$ and $h_{B}\left(x_{1}\right)=$ $\{0.4,0.2\}$, and assume that the decision makers' opinions are optimistic, we extend $h_{B}\left(x_{1}\right)$ to $h_{B}^{\prime}\left(x_{1}\right)=$ $\{0.4,0.4,0.4,0.4,0.2\}$.

Let $A_{1}$ and $A_{2}$ be two HFSs defined on $X=\left\{x_{1}, x_{2}, \cdots x_{n}\right\}, h_{A}\left(x_{i}\right)$ and $h_{B}\left(x_{i}\right)$ be two HFEs. Based on the well-known Hamming distance and Euclidean distance, Xu and $\mathrm{Xia}{ }^{10}$ proposed a generalized hesitant normalized distance:

$$
\begin{equation*}
d_{1}(A, B)=\sum_{i=1}^{n}\left[\frac{1}{l_{x_{i}}} \sum_{j=1}^{l_{x_{j}}}\left|h_{A}^{\sigma(j)}\left(x_{i}\right)-h_{B}^{\sigma(j)}\left(x_{i}\right)\right|^{\chi}\right]^{\frac{1}{\lambda}} \tag{1}
\end{equation*}
$$

where $\lambda>0$. Particularly, if $\lambda=1,2$, then the generalized hesitant normalized distance is reduced to the hesitant Hamming distance and the hesitant Euclidean distance, respectively:

$$
\begin{gather*}
d_{2}(A, B)=\sum_{i=1}^{n}\left[\frac{1}{l_{x_{i}}} \sum_{j=1}^{l_{x_{i}}}\left|h_{A}^{\sigma(j)}\left(x_{i}\right)-h_{B}^{\sigma(j)}\left(x_{i}\right)\right|\right]  \tag{2}\\
d_{3}(A, B)=\left[\sum_{i=1}^{n}\left(\frac{1}{l_{x_{i}}} \sum_{j=1}^{l_{x_{i}}}\left|h_{A}^{\sigma(j)}\left(x_{i}\right)-h_{B}^{\sigma(j)}\left(x_{i}\right)\right|^{2}\right)\right] \tag{3}
\end{gather*}
$$

If the weight $w_{i}$ of each element $x_{i}$ is taken into consideration, then the generalized hesitant weighted distance is defined as follows:

$$
\begin{equation*}
d_{4}(A, B)=\sum_{i=1}^{n} w_{i}\left[\frac{1}{l_{x_{i}}} \sum_{j=1}^{l_{x_{i}}}\left|h_{A}^{\sigma(j)}\left(x_{i}\right)-h_{B}^{\sigma(j)}\left(x_{i}\right)\right|^{\lambda}\right]^{\frac{1}{\lambda}} \tag{4}
\end{equation*}
$$

where $w_{i} \geq 0, i=1,2 \cdots n, \sum_{i=1}^{n} w_{i}=1$.
Example 2. For two HFEs $h_{1}(x)=\{0.8,0.1\}$ and $h_{2}(x)=\{0.8,0.7,0.6,0.5,0.4,0.3,0.2,0.1\}$. Now we calculate the distance between these two HFEs: According to the distance measure proposed above, we first extend $h_{1}(x)$ until it $h_{1}(x)$ and $h_{2}(x)$ have the same length. Suppose that the decision maker is pessimistic, then we should add the minimum number to $h_{1}(x)$. Then $h_{1}(x)$ can be modified to $h_{1}^{\prime}(x)=(0.8,0.1,0.1,0.1,0.1,0.1,0.1,0.1)$. By the hesitant Euclidean distance, we can get the distance between $h_{1}^{\prime}$ and $h_{2}$ as $d_{3}\left(h_{1}, h_{2}\right)=0.096$.

## 3. A novel method to calculate the distance between HFSs

### 3.1. The drawbacks of the existing distance measures for HFSs

In Section 2, we have reviewed some distance measures. Clearly, all of the above distance measures satisfy the three conditions in Definition 1. However, we can easily find that these distance measures have some shortcomings. Most importantly, these distance measures are based on the following assumptions:

- The values in a HFE are arranged in ascending order or increasing order.
- Assume that the lengths of each two corresponding HFEs are the same.
However, in most practical decision making problems, the lengths of the corresponding HFEs may be different. In some cases, they may even have big differences. It is also possible that the values in $h_{A}$ and $h_{B}$ are in order, that is to say, to arrange the values again may be improper. The existing distance measures
have some weakness as described in the introduction. In the following, we will give two short examples to illustrate the weakness of the existing methods:

Example 3. Suppose that we are going to compute the distance between $h_{1}(x)=\{0.8,0.1\}$ and $h_{2}(x)=\{0.8,0.7,0.6,0.5,0.4,0.3,0.2,0.1\}$. In Example 2, we have got the distance as: $d_{3}\left(h_{1}, h_{2}\right)=0.096$. However, in the course of calculation, we have added six numbers to $h_{1}(x)$ which just has two values originally, and the original average value is 0.45 . After adding numbers, the average value is 0.1875 . Clearly, the result is totally different. Furthermore, if we add the maximum number to the shorter one and suppose that the decision maker is optimistic, then we can extend $h_{1}(x)$ to $h_{1}^{\prime}(x)=\{0.8,0.8,0.8,0.8,0.8,0.8,0.8,0.1\}$. So we can measure the distance $d_{3}\left(h_{1}^{\prime}, h_{2}\right)=0.119$, and the average value is 0.7125 . Compared to the results measured before, it is completely different. However, according to the rules given in Ref. [10], these two results are all reasonable even though they differ very much. It is the weakness of the existing methods that change the original information when calculating the distances of HFEs. In the following, we try to propose a novel method to estimate distances so as to avoid the drawback mentioned above.

### 3.2. A novel distance measure for HFSs

Let $A=\left\{<x_{i}, h_{A}\left(x_{i}\right)>\mid x_{i} \in X, i=1,2, \cdots, n\right\} \quad$ and $B=\left\{<x_{i}, h_{B}\left(x_{i}\right)>\mid x_{i} \in X, i=1,2, \cdots, n\right\}$ be two HFSs on $X=\left\{x_{1}, x_{2}, \cdots x_{n}\right\}$, respectively. According to the principle we talked before, if the values are given in disorder, we needn't to arrange the values in HFEs in a decreasing order or increasing order. If the lengths of the corresponding HFEs are different, then adding the minimum number to the short one for computing distances of HFSs is also redundant. To keep the original information, we may consider a new distance measure between the HFSs $A$ and $B$ as:

$$
\begin{equation*}
d_{9}(A, B)=\frac{1}{n} \sum_{i=1}^{n}\left|\frac{1}{l_{h_{A}\left(x_{i}\right)} l_{h_{B}\left(x_{i}\right)}} \sum_{n=1}^{l_{h_{B}\left(x_{i}\right)}} \sum_{m=1}^{l_{A\left(x_{i}\right)}} h_{A}^{m}\left(x_{i}\right)-h_{B}^{n}\left(x_{i}\right)\right| \tag{5}
\end{equation*}
$$

where $l\left(h_{A}\left(x_{i}\right)\right)$ is the length of $h_{A}\left(x_{i}\right)$. Now we need to check whether $d_{9}(A, B)$ satisfies Definition 1 or not. Apparently, $\quad d_{9}(A, B)=d_{9}(B, A)$ and $0 \leq d_{9}(A, B) \leq 1$ hold. What we need to check is $d_{9}(A, B)=0$ if and only if $A=B$. However,

$$
\begin{align*}
d_{9}(A, B) & =\frac{1}{n} \sum_{i=1}^{n}\left|\frac{1}{l_{h_{A}\left(x_{i}\right)} l_{h_{B}\left(x_{i}\right)}} \sum_{n=1}^{l_{\text {bB }}\left(x_{i}\right)} \sum_{m=1}^{l_{h_{A}\left(x_{i}\right)}} h_{A}^{m}\left(x_{i}\right)-h_{B}^{n}\left(x_{i}\right)\right|  \tag{6}\\
& =\frac{1}{n} \sum_{i=1}^{n}\left|l_{h_{B}\left(x_{i}\right)} \sum_{m=1}^{h_{n A(x)}} h_{A}^{m}\left(x_{i}\right)-l_{h_{A}\left(x_{i}\right)} \sum_{n=1}^{l_{h_{B}\left(x_{i}\right)}} h_{B}^{n}\left(x_{i}\right)\right|
\end{align*}
$$

If $l_{h_{B}\left(x_{i}\right)} \sum_{m=1}^{l_{h_{n}\left(x_{i}\right)}} h_{A}^{m}\left(x_{i}\right)-l_{h_{A}\left(x_{i}\right)} \sum_{n=1}^{h_{B B}\left(x_{i}\right)} h_{B}^{n}\left(x_{i}\right)=0$ holds, then $d_{9}(A, B)=0$. While in this case, $A=B$ cannot be guaranteed. For example, $A=\{0.5,0.4,0.3\}$ and $B=\{0.6,0.5,0.4,0.1\}, d_{9}(A, B)=0$ but $A \neq B$. On the contrary, when $A \neq B, d_{9}(A, B)=0$ may also be possible. Thus, $d_{9}(A, B)=0$ if and only if $A=B$ cannot be guaranteed in any cases.

Through analyzing the data, we find that what leads to this conclusion is that we haven't considered the influence of the length and deviation of data on the result. These two parameters are also essential, which involve the hesitance of the decision makers. In the following, we will analyze the importance of hesitance of the decision makers, and develop a novel distance measure considering the hesitation degrees.
Example 4. For two HFEs $h_{1}(x)=\{0.8\}$ and $h_{2}(x)=\{0.8,0.8,0.8,0.8,0.8,0.8\}$. According to the rules ${ }^{10}$, the lengths of these two HFEs are different. For convenience of calculation, we extend $h_{1}(x)$ to $h_{1}(x)=\{0.8,0.8,0.8,0.8,0.8,0.8\}$. So we can conclude that these two HFEs are the same. But considering the amount and distribution of values, the decision makers who provided $h_{2}(x)$ seem to be more hesitant. As a result, the hesitation degrees cannot be ignored.

In the previous research, $\mathrm{Li}^{26}$ defined a concept of hesitation degree only considering the lengths of HFEs:
Definition $4{ }^{26}$. Let $A$ be a HFS on $X=\left\{x_{1}, x_{2}, \cdots x_{n}\right\}$. Then the hesitation degree of the HFS $A$ is defined as:

$$
\begin{equation*}
\mu(A)=\frac{1}{n} \sum_{i=1}^{n} \mu\left(h\left(x_{i}\right)\right) \tag{7}
\end{equation*}
$$

where

$$
\begin{equation*}
\mu\left(h\left(x_{i}\right)\right)=1-\frac{1}{l\left(h\left(x_{i}\right)\right)} \tag{8}
\end{equation*}
$$

and $l\left(h\left(x_{i}\right)\right)$ is the length of $h\left(x_{i}\right)$.
For the $\operatorname{HFE} h\left(x_{i}\right)$, the value of $\mu\left(h\left(x_{i}\right)\right)$ represents the hesitation degree of a decision maker when he or she determines the membership degree. If $l\left(h\left(x_{i}\right)\right)=1$, then $\mu\left(h\left(x_{i}\right)\right)=0$. It means that the decision maker is quite sure about the membership degree. In contrast, if $l\left(h\left(x_{i}\right)\right)$ intends to be infinite, the
hesitation degree $\mu\left(h\left(x_{i}\right)\right)$ intends to be 1 which indicates that the decision maker is quite hesitant. For example, for the HFEs $h_{1}(x)=(0.9,0.8,0.7)$ and
$h_{2}(x)=(0.8,0.5) \quad, \quad \mu\left(h_{1}(x)\right)=1-\frac{1}{3}=\frac{2}{3} \quad, \quad$ and $\mu\left(h_{2}(x)\right)=1-\frac{1}{2}=\frac{1}{2}$.

Superficially, the result mentioned above is somewhat reasonable. It is clear that the hesitation degree is closely associated to the length of HFE. However, to a certain extent, this method is not comprehensive. For example, when talking about the membership degree about the same alternative to meet the same properties, one decision maker puts forward his/her preference by the $\operatorname{HFE} h_{1}(x)=(0.9,0.8,0.7)$, while the other expresses the preference with $h_{2}(x)=(0.9,0.2,0.1)$. Obviously, we can discover that the lengths of $h_{1}(x)$ and $h_{2}(x)$ are the same. If we evaluate the hesitation degree according to the method mentioned above, then the hesitation degrees of the two HFEs will be same as well, namely, $\mu\left(h_{1}(x)\right)=\mu\left(h_{2}(x)\right)$. However, by analyzing the data, we can catch that data in $h_{2}(x)$ spread wider compared with $h_{1}(x)$. It means that the decision maker is doubtful about the membership degree and actually very uncertain. So, in fact $\mu\left(h_{1}(x)\right)<\mu\left(h_{2}(x)\right)$. That is the reason why the decision maker gives $0.9,0.2$ and 0.1 , these three numbers are far away from each other. From the example, we can see that only considering the length of HFE is not enough at all. When analyzing the hesitation degree, it is essential to take the deviation degrees of data into account as well. The wider the data distribute, the bigger the deviation degree is.

Consequently, combined with the standard deviation, we put forward a generalized hesitation degree considering the influence of divergence and the length of HFE.

Definition 5. Let $A$ be a HFS on $X=\left\{x_{1}, x_{2}, \cdots x_{n}\right\}$. The generalized hesitation degree can be expressed as:

$$
\begin{gather*}
\mu(h(x))=\alpha\left(1-\frac{1}{l(h(x))}\right)+\beta \sqrt{\frac{\sum_{i=1}^{n}\left(x_{i}-\bar{x}\right)^{2}}{n}}  \tag{9}\\
\mu(A)=\frac{1}{n} \sum_{i=1}^{n} \mu\left(h\left(x_{i}\right)\right) \tag{10}
\end{gather*}
$$

where $\alpha$ and $\beta$ are the weight coefficients, $0 \leq \alpha, \beta \leq 1$ and $\alpha+\beta=1$. If $\alpha=0$, then it means that we pay no attention to the influence of the length of HFE. In contrast, if $\beta=0$, then it represents that we
ignore the standard deviation. For convenience, here we let $\alpha=\beta=\frac{1}{2}$. So, the generalized hesitation degree has the following form:

$$
\begin{equation*}
\mu(h(x))=\frac{1}{2}\left(1-\frac{1}{l(h(x))}+\sqrt{\frac{\sum_{i=1}^{n}\left(x_{i}-\bar{x}\right)^{2}}{n}}\right) \tag{11}
\end{equation*}
$$

To preserve the original information and by combining the generalized hesitation degrees, we try to define a new distance measure between the HFSs $A$ and $B$ as:

$$
\begin{gather*}
d_{10}\left(h_{A}\left(x_{i}\right), h_{B}\left(x_{i}\right)\right)=\left|\frac{\alpha}{l_{h_{A}\left(x_{i}\right)} l_{k_{B}\left(x_{i}\right)}} \sum_{n=1}^{l_{h_{B}\left(x_{i}\right)}} \sum_{m=1}^{l_{n(x)}} h_{A}^{m}\left(x_{i}\right)-h_{B}^{n}\left(x_{i}\right)\right|+  \tag{12}\\
\beta\left|\mu_{h_{A}}\left(x_{i}\right)-\mu_{h_{B}}\left(x_{i}\right)\right| \\
d_{10}(A, B)=\frac{1}{n} \sum_{i=1}^{n} d\left(h_{A}\left(x_{i}\right), h_{B}\left(x_{i}\right)\right) \tag{13}
\end{gather*}
$$

where $\alpha$ and $\beta$ are the weight coefficients, $0<\alpha, \beta<1$ and $\alpha+\beta=1$. and $l\left(h_{A}\left(x_{i}\right)\right)$ is the length of $h_{A}\left(x_{i}\right) \cdot \mu_{h_{A}}\left(x_{i}\right)$ is the generalized hesitation degree of the HFE $h_{A}\left(x_{i}\right)$. For convenience, suppose that we have the same preference between the hesitation degree and the membership values, then $\alpha=\beta=\frac{1}{2}$, and the distance measure can be expressed as:

$$
d_{10}\left(h_{A}\left(x_{i}\right), h_{B}\left(x_{i}\right)\right)=\frac{1}{2}\left[\begin{array}{l}
\left.\left|\frac{1}{l_{h_{A}\left(x_{i}\right)} l_{h_{B}\left(x_{i}\right)}} \sum_{n=1}^{l_{\text {max }}} \sum_{m=1}^{l_{n}\left(x_{1}\right)} h_{A}^{m}\left(x_{i}\right)-h_{B}^{n}\left(x_{i}\right)\right|+\right] \\
\left|\mu_{h_{A}}\left(x_{i}\right)-\mu_{h_{B}}\left(x_{i}\right)\right|
\end{array}\right]
$$

(14)

It is clear that we have taken the generalized hesitation degree into consideration in the formula (14). Next, we will prove that this distance measure satisfies all the properties in Definition 1.

## Proof.

(1) Since for every HFE, $h_{A}^{m}(x) \geq 0$ and $\mu_{h_{A}}\left(x_{i}\right) \geq 0$ hold. It is obvious that $0 \leq d_{10}(A, B) \leq 1$.
(2) If $A=B$, then $h_{A}^{m}\left(x_{i}\right)=h_{B}^{n}\left(x_{i}\right)$, for any $x_{i} \in X$ and any $m, n$. What's more, their lengths and deviation degrees are both the same, so $\mu_{h_{A}}\left(x_{i}\right)=\mu_{h_{B}}\left(x_{i}\right)$, for any $1 \leq i \leq n$. As a result, $d_{10}(A, B)=0$. On the contrary, if $d_{10}(A, B)=0$, then we can get that for any $1 \leq i \leq n$, $\left|\mu_{h_{A}}\left(x_{i}\right)-\mu_{h_{B}}\left(x_{i}\right)\right|=0$
and $\left|l_{h_{B}\left(x_{i}\right)} \sum_{m=1}^{l_{h_{A}\left(x_{i}\right)}} h_{A}^{m}\left(x_{i}\right)-l_{h_{A}\left(x_{i}\right)} \sum_{n=1}^{l_{h_{B}\left(x_{i}\right)}} h_{B}^{n}\left(x_{i}\right)\right|=0 . \quad$ Since $\left|\mu_{h_{A}}\left(x_{i}\right)-\mu_{h_{B}}\left(x_{i}\right)\right|=0$, then we can get that the lengths of figures and the deviation degrees in $h_{A}\left(x_{i}\right)$ and $h_{B}\left(x_{i}\right)$ are both the same. Thus, $h_{A}^{m}\left(x_{i}\right)=h_{B}^{n}\left(x_{i}\right)$, for any $x_{i} \in X$ and any $m, n$. As a result, we can get that the $\operatorname{HFS} A$ is equal to the $\operatorname{HFS} B$. That is to say, $d_{10}(A, B)=0$ if and only if $A=B$.
(3) It is straightforward that $d_{10}(A, B)=d_{10}(B, A)$.

Consider that the objects $x_{i}(i=1,2, \ldots, n)$ may have different weights in some cases, we propose a weighted form of the distance measure for HFSs. Let $w=\left(w_{1}, w_{2} \cdots w_{n}\right)$ be the weight vector of $x_{i}(i=1,2, \ldots, n)$, with $\quad w_{i} \geq 0, i=1,2, \cdots, n, \quad$ and $\sum_{i=1}^{n} w_{i}=1$. Then the hesitant weighted distance is defined as:

$$
d_{11}(A, B)=\sum_{i=1}^{n}\left\{\frac{w_{i}}{2}\left[\begin{array}{l}
\left|\frac{1}{l_{h_{A}(x)}\left(l_{l_{B}\left(x_{i}\right)}\right.} \sum_{n=1}^{l_{n \in(x)}} \sum_{m=1}^{l_{n}(x)} h_{A}^{m}\left(x_{i}\right)-h_{B}^{n}\left(x_{i}\right)\right|+  \tag{15}\\
\left|\mu_{h_{A}}\left(x_{i}\right)-\mu_{h_{B}}\left(x_{i}\right)\right|
\end{array}\right]\right\}
$$

which also satisfies the properties in Definition 1.
In the following work, we will apply the novel distance measure developed in this paper to clustering analysis. Actually, there are some existing clustering methods, such as the MST clustering algorithm ${ }^{23}$, hierarchical clustering algorithm ${ }^{21}$ and so on. But these methods are somewhat complicated and need so much calculation and lots of transformations. In the following, we will propose a straightforward clustering algorithm called the orthogonal method for clustering the HFSs.

## 4. An orthogonal method for clustering the HFSs

The main idea of the orthogonal clustering algorithm is uncomplicated at all: Firstly, we utilize the developed distance measure to compute the distance between each two HFSs, and then construct a distance matrix $M$; Secondly, we should choose a confidence level $\lambda \in[0,1]$ to get a $\lambda$-cutting matrix $M_{\lambda}$ of the distance matrix $M$, and then take each column of the matrix $M_{\lambda}$ as a vector, so the matrix $M_{\lambda}$ can be expressed as $M_{\lambda}=\left(\overrightarrow{\alpha_{1}}, \overrightarrow{\alpha_{2}}, \cdots, \overrightarrow{\alpha_{n}}\right)$, where $\overrightarrow{\alpha_{j}}=\left(\alpha_{1 j}, \alpha_{2 j}, \cdots, \alpha_{n j}\right)^{T}$. Finally, we utilize the orthogonal relation among the HFSs to cluster the objects. The detailed process can be described as follows:

Step 1. Let $\left\{A_{1}, A_{2}, \cdots, A_{n}\right\}$ be a set of HFSs over $X=\left\{x_{1}, x_{2}, \cdots, x_{m}\right\}$, representing $m$ samples. using the distance measure defined before, we can calculate the distance between each two HFSs, and then construct a distance matrix $M=\left(d_{i j}\right)_{n \times n}$, where $d_{i j}=d\left(A_{i}, A_{j}\right)$.

Step 2. Choose the confidence level $\lambda \in[0,1]$, and then construct the corresponding $\lambda$-cutting matrix $M_{\lambda}$ according to Definition 4. We choose the value of $\lambda$ from the values in the matrix $M$, with the order from the biggest value to the smallest one.

Step 3. After getting the $\lambda$-cutting matrix, we take each column of the matrix $M_{\lambda}$ as a vector. As a result, the matrix $M_{\lambda}$ can be expressed as $M_{\lambda}=\left(\overrightarrow{\alpha_{1}}, \overrightarrow{\alpha_{2}}, \cdots, \overrightarrow{\alpha_{n}}\right)$, where $\overrightarrow{\alpha_{j}}=\left(\alpha_{1 j}, \alpha_{2 j}, \cdots, \alpha_{n j}\right)^{T}$. The inner product of any two column vectors is $\left(\overrightarrow{\alpha_{i}}, \overrightarrow{\alpha_{j}}\right)={\overrightarrow{\alpha_{i}}}^{T} \overrightarrow{\alpha_{j}}$. If $\left(\overrightarrow{\alpha_{i}}, \overrightarrow{\alpha_{j}}\right)={\overrightarrow{\alpha_{i}}}^{T} \overrightarrow{\alpha_{j}}=0$, then we call that these two column vectors are orthogonal.

Step 4. We cluster the objects into a few classes according to the orthogonal method among the column vectors. The detailed procedure is as follows:

○,1If $\left(\overrightarrow{\alpha_{i}}, \overrightarrow{\alpha_{j}}\right) \neq 0$, then we cluster the objects $A_{i}$ and $A_{j}$ into the same class. This is called direct clustering principle.

O,2 If there exist $1 \leq n_{1}, n_{2}, \cdots, n_{s} \leq n$, and $\left(\overrightarrow{\alpha_{i}}, \overrightarrow{\alpha_{n_{1}}}\right)\left(\overrightarrow{\alpha_{n_{1}}}, \overrightarrow{\alpha_{n_{2}}}\right) \cdots\left(\overrightarrow{\alpha_{n_{s}}}, \overrightarrow{\alpha_{j}}\right) \neq 0$, then we cluster the objects $A_{i}$ and $A_{j}$ into the same class. This is called indirect clustering principle.

○, 3if $\left(\overrightarrow{\alpha_{i}}, \overrightarrow{\alpha_{j}}\right)=0$, and for any $1 \leq n_{1}, n_{2}, \cdots, n_{s} \leq n$, $\left(\overrightarrow{\alpha_{i}}, \overrightarrow{\alpha_{n_{1}}}\right)\left(\overrightarrow{\alpha_{n_{1}}}, \overrightarrow{\alpha_{n_{2}}}\right) \cdots\left(\overrightarrow{\alpha_{n_{s}}}, \overrightarrow{\alpha_{j}}\right)=0$, then the objects $A_{i}$ and $A_{j}$ are not in the same group.

Compared with the hesitant fuzzy MST clustering algorithm ${ }^{23}$ and the hesitant fuzzy agglomerative hierarchical clustering algorithm ${ }^{21}$, the hesitant fuzzy orthogonal clustering method we proposed is much easier and can be realized by computer programs. It is practical and can be generalized to the large data environment. While in the hesitant fuzzy agglomerative hierarchical clustering algorithm ${ }^{21}$, we should first divide the alternatives into certain clusters, compute the distance of each pair of clusters, combine the clusters with the minimum distance to form a new cluster, and repeat the above steps until all the alternatives are in the same clusters. Obviously, it is much more complicated and needs a large amount of computational efforts and takes a lot of time to accomplish. In the hesitant fuzzy MST clustering algorithm ${ }^{23}$, after we get the distance of each pair of alternatives, we need draw a hesitant fuzzy
graph where every node represents an alternative and every edge has weight which shows the dissimilarity degree. Then, we make the clustering analysis by using the hesitant fuzzy minimal spanning tree. Although this method is easy to realize, it is not convenient to be computed by the automatic programs, which is a big limitation.

## 5. Applications

In what follows, two numerical examples are given to demonstrate the effectiveness and practicality of the proposed clustering method:

Example 5. Different people have different requirements towards computers. Someone who likes watching movie prefers the computer with clear screen compared to other characters, while game players may like the computer with high speed CPU to control the figures in time.

Now there are six different computers to be evaluated. In order to provide recommendations to the consumers, a digital evaluation website invites six experts to evaluate the performance of these six
computers, mainly from five aspects: price $\left(x_{1}\right)$, electricity consumption $\left(x_{2}\right)$, the speed of $\mathrm{CPU}\left(x_{3}\right)$, the quality of screen $\left(x_{4}\right)$, and design $\left(x_{5}\right)$. The weight of each attribute is not the same and can be changed according to different demands. For business man, electricity consumption is very important, so we can increase the weight of $x_{2}$ to pick out the most suitable computer. In this paper, we assume that the attribute weight vector is $w=(0.1,0.2,0.3,0.25,0.15)^{T}$. The evaluation results for each computer are expressed by hesitant fuzzy numbers shown in Table 1. When the experts evaluate the performance of each computer, the value of the evaluation will be given. For example, the CPU of one computer runs very fast. The experts believe that CPU is very good, they may give a relatively high value, maybe close to 1 . If someone is hesitant in giving membership degree, he/she may give two or more values. Gathering all the evaluation results, we form Table 1. The larger the value of the hesitant fuzzy element in Table 1 is, the better the attribute of the alternative do the experts think.

Table 1. The characteristics information of the computers


Step 2. According to the formula (15), we can calculate the distance with the weighting vector $w=(0.1,0.2,0.3,0.25,0.15)$ between each two computers:
$d_{11}(A, B)=\sum_{i=1}^{n}\left\{\frac{w_{i}}{2}\left[\begin{array}{l}\left.\left\lvert\, \frac{1}{l_{h_{A}\left(x_{i}\right)} l_{h_{B}\left(x_{i}\right)}} \sum_{n=1}^{l_{m_{B}\left(x_{i}\right)}} \sum_{m=1}^{l_{h_{A A}\left(x_{i}\right)}} h_{i}^{m}\right.\right)-\mu_{h_{B}}\left(x_{i}\right) \mid\end{array}\right]\right\}$
Consequently, we can get the distance matrix:
$D=\left(\begin{array}{cccccc}0 & 0.0868 & 0.1017 & 0.0985 & 0.2099 & 0.2468 \\ 0.0868 & 0 & 0.0861 & 0.1097 & 0.1838 & 0.2258 \\ 0.1017 & 0.0861 & 0 & 0.1269 & 0.1935 & 0.2148 \\ 0.0985 & 0.1097 & 0.1269 & 0 & 0.1709 & 0.2417 \\ 0.2099 & 0.1838 & 0.1935 & 0.1709 & 0 & 0.1570 \\ 0.2468 & 0.2258 & 0.2148 & 0.2417 & 0.1570 & 0\end{array}\right)$

Step 3. We choose the confidential level $\lambda$ from the distance matrix to get the $\lambda$-cutting matrix $D_{\lambda}=\left({ }_{\lambda} d_{i j}\right)_{6 \times 6}$ according to the principles given before. For example, if $\lambda=0.0868$, then any values in the distance matrix which are bigger than $\lambda$ would turn into 1 , otherwise, it would be 0 . Consequently, when $\lambda=0.0868$, the $\lambda$-cutting matrix is expressed as:
$D_{\lambda}=\left(\begin{array}{cccccc}0 & 1 & 1 & 1 & 1 & 1 \\ 1 & 0 & 0 & 1 & 1 & 1 \\ 1 & 0 & 0 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 & 1 & 1 \\ 1 & 1 & 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 1 & 1 & 0\end{array}\right)=\left(\overrightarrow{\alpha_{1}}, \overrightarrow{\alpha_{2}}, \overrightarrow{\alpha_{3}}, \overrightarrow{\alpha_{4}}, \overrightarrow{\alpha_{5}}, \overrightarrow{\alpha_{6}}\right)$

If $\lambda=0.2148$, then the $\lambda$-cutting matrix is expressed as:

$$
D_{\lambda}=\left(\begin{array}{llllll}
0 & 0 & 0 & 0 & 0 & 1 \\
0 & 0 & 0 & 0 & 0 & 1 \\
0 & 0 & 0 & 0 & 0 & 1 \\
0 & 0 & 0 & 0 & 0 & 1 \\
0 & 0 & 0 & 0 & 0 & 0 \\
1 & 1 & 1 & 1 & 0 & 0
\end{array}\right)=\left(\overrightarrow{\alpha_{1}}, \overrightarrow{\alpha_{2}}, \overrightarrow{\alpha_{3}}, \overrightarrow{\alpha_{4}}, \overrightarrow{\alpha_{5}}, \overrightarrow{\alpha_{6}}\right)
$$

which takes each column of the matrix $D_{\lambda}$ as a vector. As a result, the matrix $D_{\lambda}$ can be expressed as $D_{\lambda}=\left(\overrightarrow{\alpha_{1}}, \overrightarrow{\alpha_{2}}, \cdots, \overrightarrow{\alpha_{6}}\right)$, where $\overrightarrow{\alpha_{j}}=\left(\alpha_{1 j}, \alpha_{2 j}, \cdots, \alpha_{6 j}\right)^{T}$. The inner product of any two column vectors is $\left(\overrightarrow{\alpha_{i}}, \overrightarrow{\alpha_{j}}\right)={\overrightarrow{\alpha_{i}}}^{T} \overrightarrow{\alpha_{j}}$. If $\left(\overrightarrow{\alpha_{i}}, \overrightarrow{\alpha_{j}}\right)={\overrightarrow{\alpha_{i}}}^{T} \overrightarrow{\alpha_{j}}=0$, then we call that these two column vectors are orthogonal. The orthogonal vectors cannot be clustered into the same group. That is to say, if $\left(\overrightarrow{\alpha_{i}}, \overrightarrow{\alpha_{j}}\right) \neq 0$, then we cluster the computers $A_{i}$ and $A_{j}$ into the same class. For example, when $\lambda=0.2148,\left(\overrightarrow{\alpha_{1}}, \overrightarrow{\alpha_{2}}\right) \neq 0$, then cluster the computers $A_{1}$ and $A_{2}$ into the same class.

Consequently, we can get all the possible classifications of $A_{i}(i=1,2, \cdots, 6)$ as follows:
(1) If $0 \leq \lambda \leq 0.1570$, then $A_{i}(i=1,2, \cdots, 6)$ are clustered into the same group:

$$
\left\{A_{1}, A_{2}, A_{3}, A_{4}, A_{5}, A_{6}\right\}
$$

(2) If $0.1570<\lambda \leq 0.2099$, then $A_{i}(i=1,2, \cdots, 6)$ are clustered into two groups:

$$
\left\{A_{1}, A_{2}, A_{3}, A_{4}\right\},\left\{A_{5}, A_{6}\right\}
$$

(3) If $0.2099<\lambda \leq 0.2148$, then $A_{i}(i=1,2, \cdots, 6)$ are clustered into three groups:

$$
\left\{A_{1}, A_{2}, A_{3}, A_{4}\right\},\left\{A_{5}\right\},\left\{A_{6}\right\}
$$

(4) If $0.2148<\lambda \leq 0.2258$, then $A_{i}(i=1,2, \cdots, 6)$ are clustered into three groups:

$$
\left\{A_{1}, A_{2}, A_{4}\right\},\left\{A_{3}\right\},\left\{A_{5}\right\},\left\{A_{6}\right\}
$$

(5) If $0.2258<\lambda \leq 0.2417$, then $A_{i}(i=1,2, \cdots, 6)$ are clustered into five groups:

$$
\left\{A_{1}, A_{4}\right\},\left\{A_{2}\right\},\left\{A_{3}\right\},\left\{A_{5}\right\},\left\{A_{6}\right\}
$$

(6) If $0.2417<\lambda \leq 1$, then $A_{i}(i=1,2, \cdots, 6)$ are clustered into six groups:

$$
\left\{A_{1}\right\},\left\{A_{2}\right\},\left\{A_{3}\right\},\left\{A_{4}\right\},\left\{A_{5}\right\},\left\{A_{6}\right\}
$$

Obviously, according to the clustering results, the computer market can divide these computers into several groups. The computers in the same group have something similar in some extent. If some consumers regard one attribute as quite important, such as "speed of CPU $\left(x_{3}\right)$ ", then the attribute weight vector can be changed to get the satisfied computers. Different attribute weight vectors lead to different clustering results.

Liu et al. ${ }^{27}$ proposed a hesitant fuzzy netting clustering method, whose process is as follows:
Step 1. According to the similarity measures, we compute the similarity between each two HFSs.
Step 2. Construct the similarity matrix $S=\left(s_{i j}\right)_{m \times m}$, where $s_{i j}=s\left(A_{i}, A_{j}\right), i, j=1,2, \ldots, m$.
Step 3. Delete all the elements above the diagonal of the distance matrix, and replace the elements on the diagonal with the representation of the objects.
Step 4. Choose the confidence level $\lambda$ and construct the corresponding $\lambda$-cutting matrix. Replace ' 1 ' with '*' and delete all the ' 0 ' in the matrix. From the points '*' in the matrix, we can draw the vertical and horizontal line to the representations of the objects on the diagonal. Obviously, each '*' links to two points on the diagonal which represents two objects. Cluster these two objects into the same group. Until go through all the points '*', we can get the clustering result corresponding to the selected $\lambda$. Choosing different
values of $\lambda$, we can get different clustering results until all the objects are clustered into one class.

In the following, we cluster the six kinds of computers with the hesitant fuzzy netting clustering method ${ }^{27}$ again. Here, we utilize the distance measure to estimate the relationship between computers instead of similarity measure. After we get the $\lambda$-cutting matrix, if we use the netting clustering method, then we first should replace ' 1 ' with '*' and delete all the ' 0 ' in the matrix according to Step 4. For example, when $\lambda=0.2258$, the $\lambda$-cutting matrix is expressed as:

$$
D_{\lambda}=\left(\begin{array}{llllll}
0 & 0 & 0 & 0 & 0 & 1 \\
0 & 0 & 0 & 0 & 0 & 1 \\
0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 \\
0 & 0 & 0 & 0 & 0 & 0 \\
1 & 1 & 0 & 1 & 0 & 0
\end{array}\right)
$$

$$
D_{\lambda}=\left(\begin{array}{lllllll}
\mathrm{A}_{1} & & & & & \\
\vdots & \mathrm{~A}_{2} & & & & \\
\vdots & \vdots & \mathrm{~A}_{3} & & & \\
\vdots & \vdots & & \mathrm{~A}_{4} & & \\
\vdots & \vdots & & \vdots & \mathrm{~A}_{5} & \\
* & \ldots & * \ldots . & * & \ldots \ldots & \mathrm{~A}_{6}
\end{array}\right)
$$

In this case, the computers can be clustered into three classes:

$$
\left\{A_{1}, A_{2}, A_{4}\right\},\left\{A_{3}\right\},\left\{A_{5}\right\},\left\{A_{6}\right\}
$$

We can combine the results of these two algorithms as follows:

According to Step 4, the matrix can be adjusted to
Table 2. Clustering results

| Classes | Netting clustering method | Orthogonal clustering method |
| :--- | :--- | :--- |
| 6 | $\left\{A_{1}\right\},\left\{A_{2}\right\},\left\{A_{3}\right\},\left\{A_{4}\right\},\left\{A_{5}\right\},\left\{A_{6}\right\}$ | $\left\{A_{1}\right\},\left\{A_{2}\right\},\left\{A_{3}\right\},\left\{A_{4}\right\},\left\{A_{5}\right\},\left\{A_{6}\right\}$ |
| 5 | $\left\{A_{1}, A_{4}\right\},\left\{A_{2}\right\},\left\{A_{3}\right\},\left\{A_{5}\right\},\left\{A_{6}\right\}$ | $\left\{A_{1}, A_{4}\right\},\left\{A_{2}\right\},\left\{A_{3}\right\},\left\{A_{5}\right\},\left\{A_{6}\right\}$ |
| 4 | $\left\{A_{1}, A_{2}, A_{4}\right\},\left\{A_{3}\right\},\left\{A_{5}\right\},\left\{A_{6}\right\}$ | $\left\{A_{1}, A_{2}, A_{4}\right\},\left\{A_{3}\right\},\left\{A_{5}\right\},\left\{A_{6}\right\}$ |
| 3 | $\left\{A_{1}, A_{2}, A_{3}, A_{4}\right\},\left\{A_{5}\right\},\left\{A_{6}\right\}$ | $\left\{A_{1}, A_{2}, A_{3}, A_{4}\right\},\left\{A_{5}\right\},\left\{A_{6}\right\}$ |
| 2 | $\left\{A_{1}, A_{2}, A_{3}, A_{4}\right\},\left\{A_{5}, A_{6}\right\}$ | $\left\{A_{1}, A_{2}, A_{3}, A_{4}\right\},\left\{A_{5}, A_{6}\right\}$ |
| 1 | $\left\{A_{1}, A_{2}, A_{3}, A_{4}, A_{5}, A_{6}\right\}$ | $\left\{A_{1}, A_{2}, A_{3}, A_{4}, A_{5}, A_{6}\right\}$ |

Obviously, the results derived by these two algorithms are the same. Since the two algorithms utilize the same distance formula and the core idea is similar. However, there still exist some differences between these two methods. In the hesitant fuzzy orthogonal method, we take every column of the distance matrix as a vector. The relationship between each two objects can be easily found through the orthogonal vectors. Thus, this method can be accomplished easily and efficiently by MATLAB programs, which speeds up the clustering progress. While the hesitant fuzzy netting clustering method ${ }^{27}$ is not very convenient to realize with the computer programs. Since the netting is complicated to realize for programs and the netting graph needs people to
recognize, then if the data are very complicated and very big, clustering so many objects is really difficult by using the hesitant fuzzy netting clustering method ${ }^{27}$. Compared to the hesitant fuzzy netting method ${ }^{27}$, the hesitant fuzzy orthogonal method is much more effective and simple.

In order to illustrate the computation complexity, we generate a few HFSs at random for clustering to compare these two algorithms. We measure the computation time before we get the clustering results respectively. The run time of these two methods is shown in Table 3. Considering the practical application, we think the orthogonal clustering method can save much time for big data problem.

Table 3. The running time for each method

| numbers | 6 | 10 | 15 | 20 |
| :---: | :---: | :---: | :---: | :---: |
| The orthogonal <br> method | 0.000164 | 0.000523 | 0.001056 | 0.001619 |
| The netting method | 0.000205 | 0.000852 | 0.001431 | 0.002057 |

In the following example, we compare the clustering results by using our distance measure and the existing distance measure given in Ref. [10]:

Example 6. As we all know, pirate is one of the most important factors threatening the security of merchant shipping. With different social backgrounds, pirates in different oceans have great difference in equipment and
strength. Furthermore, their attack targets and means of crimes are usually different. However, it is common that they all escape quickly from the scene, always before the arrival of modern Navy. In this case, we cannot know its real strength, such as weapons, amounts of people and the areas they always appear in. While fuzzy mathematics can handle these uncertain problems better. According to the features of every accident, we can cluster the pirates in any area into a few groups so that we can easily find out which area is very dangerous and the situations in that area. What's more, through comparing the strengths of pirates in all areas, we can give advice to the passing ships. It contributes greatly to the ship's emergency plan and risk management.

According to the report of IMO (International Maritime Organization), we first extract the features in every attacked accident, including the number of pirates, the damage degree of ships, the loss degree of packages, the damage degree of ship crews and so on.

Since the damage degree cannot be expressed accurately in mathematical forms, it is better to estimate them by HFSs. For every accident, the specialists will make judgements about the damage degrees. According to the given data, we can divide the ocean area and estimate the threatening degrees of any areas and any pirates.

In the recent three months, there happened 10 accidents attacked by pirates. After each crime, some specialists are invited to evaluate the damage degrees of these 10 accidents by HFSs, including the amount of pirates $\left(x_{1}\right)$, the damage degree of ships $\left(x_{2}\right)$, the loss degree of packages $\left(x_{3}\right)$, the damage degree of ship crews $\left(x_{4}\right)$. Now we get the data to make clustering in order to give advice to the passing ships. The attribute weight vector is $w=(0.4,0.2,0.1,0.3)$, and the data are shown below:

Table 4. The characteristic information of pirate crimes

|  | $x_{1}$ | $x_{2}$ | $x_{3}$ | $x_{4}$ |
| :---: | :---: | :---: | :---: | :---: |
| $A_{1}$ | $\{0.1,0.1,0.2,0.4,0.5,0.5\}$ | $\{0.2,0.4\}$ | $\{0.5,0.7,0.8\}$ | $\{0.6,0.6,0.7,0.7\}$ |
| $A_{2}$ | $\{0.1,0.2,0.2,0.4\}$ | $\{0.2,0.4,0.7\}$ | $\{0.3,0.5,0.5,0.6\}$ | $\{0.5,0.5,0.6\}$ |
| $A_{3}$ | $\{0.3,0.4\}$ | $\{0.3,0.3,0.4,0.4,0.5\}$ | $\{0.7,0.8\}$ | $\{0.2,0.2,0.4,0.5,0.5,0.6\}$ |
| $A_{4}$ | $\{0.5,0.6\}$ | $\{0.2\}$ | $\{0.5,0.7,0.7\}$ | $\{0.5,0.5\}$ |
| $A_{5}$ | $\{0.3\}$ | $\{0.4,0.6,0.6\}$ | $\{0.5,0.6,0.7\}$ | $\{0.3,0.4,0.5\}$ |
| $A_{6}$ | $\{0.5,0.6,0.6\}$ | $\{0.3,0.6\}$ | $\{0.4,0.5\}$ | $\{0.6,0.7,0.8\}$ |
| $A_{7}$ | $\{0.3,0.4,0.4,0.5,0.6,0.6\}$ | $\{0.5,0.7\}$ | $\{0.4,0.5\}$ | $\{0.2,0.3,0.6\}$ |
| $A_{8}$ | $\{0.3,0.5\}$ | $\{0.3,0.7,0.8\}$ | $\{0.4,0.7,0.9\}$ | $\{0.7,0.8,0.9,0.9\}$ |
| $A_{9}$ | $\{0.3,0.4,0.5,0.5,0.5\}$ | $\{0.4,0.5,0.7\}$ | $\{0.4,0.5,0.7,0.7,0.8\}$ | $\{0.2,0.6\}$ |
| $A_{10}$ | $\{0.3,0.5\}$ | $\{0.4,0.5\}$ | $\{0.5,0.5\}$ | $\{0.3,0.4,0.5,0.6\}$ |

Step 1. Using the distance measure (15), we can get the distance matrix:

$$
\left.\begin{array}{cccccccccc}
0 & 0.0919 & 0.1391 & 0.1831 & 0.1778 & 0.1284 & 0.1317 & 0.1355 & 0.1248 & 0.1284 \\
0.0919 & 0 & 0.1233 & 0.1942 & 0.1481 & 0.1322 & 0.1303 & 0.1407 & 0.0896 & 0.1125 \\
0.1391 & 0.1233 & 0 & 0.1643 & 0.1163 & 0.1558 & 0.1297 & 0.1249 & 0.1084 & 0.0688 \\
0.1831 & 0.1942 & 0.1643 & 0 & 0.2120 & 0.1403 & 0.1857 & 0.1973 & 0.1632 & 0.1353 \\
0.1778 & 0.1481 & 0.1163 & 0.2120 & 0 & 0.1961 & 0.1665 & 0.1693 & 0.1271 & 0.1261 \\
0.1284 & 0.1322 & 0.1558 & 0.1403 & 0.1961 & 0 & 0.1171 & 0.1153 & 0.1244 & 0.0995 \\
0.1317 & 0.1303 & 0.1297 & 0.1857 & 0.1665 & 0.1171 & 0 & 0.1498 & 0.0616 & 0.0834 \\
0.1355 & 0.1407 & 0.1249 & 0.1973 & 0.1693 & 0.1153 & 0.1498 & 0 & 0.1252 & 0.1077 \\
0.1248 & 0.0896 & 0.1084 & 0.1632 & 0.1271 & 0.1244 & 0.0616 & 0.1252 & 0 & 0.0932 \\
0.1284 & 0.1125 & 0.0688 & 0.1353 & 0.1261 & 0.0995 & 0.0834 & 0.1077 & 0.0932 & 0
\end{array}\right)
$$

Step 2. Choose the confidential level $\lambda$ from the distance matrix to get the $\lambda$-cutting matrix $D_{\lambda}=\left({ }_{\lambda} d_{i j}\right)_{10 \times 10}$ according to the formulas given before.
Finally, by the proposed hesitant fuzzy orthogonal clustering method, we can get the clustering results as follows:
(1) If $0 \leq \lambda \leq 0.1322$, then $A_{i}(i=1,2, \cdots, 10)$ can be clustered into the same group:

$$
\left\{A_{1}, A_{2}, A_{3}, A_{4}, A_{5}, A_{6}, A_{7}, A_{8}, A_{9}, A_{10}\right\}
$$

(2) If $0.1322<\lambda \leq 0.1558$, then $A_{i}(i=1,2, \cdots, 10)$ are clustered into two groups:

$$
\left\{A_{1}, A_{2}, A_{3}, A_{4}, A_{5}, A_{6}, A_{7}, A_{8}, A_{9}\right\},\left\{A_{10}\right\}
$$

(3) If $0.1558<\lambda \leq 0.1643$, then $A_{i}(i=1,2, \cdots, 10)$ are clustered into three groups:

$$
\left\{A_{1}, A_{2}, A_{3}, A_{4}, A_{5}, A_{6}, A_{7}, A_{8}\right\},\left\{A_{9}\right\},\left\{A_{10}\right\}
$$

(4) If $0.1643<\lambda \leq 0.1778$, then $A_{i}(i=1,2, \cdots, 10)$ are clustered into four groups:

$$
\left\{A_{1}, A_{2}, A_{4}, A_{5}, A_{6}, A_{7}, A_{8}\right\},\left\{A_{3}\right\},\left\{A_{9}\right\},\left\{A_{10}\right\}
$$

(5) If $0.1778<\lambda \leq 0.1831$, then $A_{i}(i=1,2, \cdots, 10)$ are clustered into five groups:

$$
\left\{A_{1}, A_{2}, A_{5}, A_{7}, A_{8}\right\},\left\{A_{4}, A_{6}\right\},\left\{A_{3}\right\},\left\{A_{9}\right\},\left\{A_{10}\right\}
$$

(6) If $0.1831<\lambda \leq 0.1857$, then $A_{i}(i=1,2, \cdots, 10)$ are clustered into six groups:

$$
\left\{A_{1}\right\},\left\{A_{2}, A_{5}, A_{7}, A_{8}\right\},\left\{A_{4}, A_{6}\right\},\left\{A_{3}\right\},\left\{A_{9}\right\},\left\{A_{10}\right\}
$$

(7) If $0.1857<\lambda \leq 0.1942$, then $A_{i}(i=1,2, \cdots, 10)$ are clustered into seven groups:

$$
\left\{A_{1}\right\},\left\{A_{2}, A_{5}, A_{8}\right\},\left\{A_{4}, A_{6}\right\},\left\{A_{3}\right\},\left\{A_{7}\right\},\left\{A_{9}\right\},\left\{A_{10}\right\}
$$

(8) If $0.1942<\lambda \leq 0.1961$, then $A_{i}(i=1,2, \cdots, 10)$ are clustered into eight groups:

$$
\left\{A_{1}\right\},\left\{A_{2}\right\},\left\{A_{5}, A_{8}\right\},\left\{A_{4}, A_{6}\right\},\left\{A_{3}\right\},\left\{A_{7}\right\},\left\{A_{9}\right\},\left\{A_{10}\right\}
$$

(9) If $0.1961<\lambda \leq 0.1973$, then $A_{i}(i=1,2, \cdots, 10)$ are clustered into nine groups:
$\left\{A_{1}\right\},\left\{A_{2}\right\},\left\{A_{5}, A_{8}\right\},\left\{A_{4}\right\},\left\{A_{6}\right\},\left\{A_{3}\right\},\left\{A_{7}\right\},\left\{A_{9}\right\},\left\{A_{10}\right\}$
(10) If $0.1973<\lambda \leq 1$, then $A_{i}(i=1,2, \cdots, 10)$ are clustered into ten groups:
$\left\{A_{1}\right\},\left\{A_{2}\right\},\left\{A_{5}\right\},\left\{A_{8}\right\},\left\{A_{4}\right\},\left\{A_{6}\right\},\left\{A_{3}\right\},\left\{A_{7}\right\},\left\{A_{9}\right\},\left\{A_{10}\right\}$
In the following, we will use the existing distance measure ${ }^{10}$ in our hesitant fuzzy orthogonal clustering method:

According to the method proposed by Xu and Xia ${ }^{10}$, if the lengths of the HFEs are different, then we should extent the shorter one by adding the minimum value or the maximum value until they have the same length. Consequently, we can get the improved HFSs as follows:

Table 5. The characteristics information of pirate crimes

| $x_{1}$ |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: |
| $A_{1}$ | $\{0.1,0.1,0.2,0.4,0.5,0.5\}$ | $\{0.2,0.4,0.4,0.4,0.4\}$ | $\{0.5,0.7,0.8,0.8,0.8\}$ | $\{0.6,0.6,0.7,0.7,0.7,0.7\}$ |
| $A_{2}$ | $\{0.1,0.2,0.2,0.4,0.4,0.4\}$ | $\{0.2,0.4,0.7,0.7,0.7\}$ | $\{0.3,0.5,0.5,0.6,0.6\}$ | $\{0.5,0.5,0.6,0.6,0.6,0.6\}$ |
| $A_{3}$ | $\{0.3,0.4,0.4,0.4,0.4,0.4\}$ | $\{0.3,0.3,0.4,0.4,0.5\}$ | $\{0.7,0.8,0.8,0.8,0.8\}$ | $\{0.2,0.2,0.4,0.5,0.6,0.6\}$ |
| $A_{4}$ | $\{0.5,0.6,0.6,0.6,0.6,0.6\}$ | $\{0.2,0.2,0.2,0.2,0.2\}$ | $\{0.5,0.7,0.7,0.7,0.7\}$ | $\{0.5,0.5,0.5,0.5,0.5,0.5\}$ |
| $A_{5}$ | $\{0.3,0.3,0.3,0.3,0.3,0.3\}$ | $\{0.4,0.6,0.6,0.6,0.6\}$ | $\{0.5,0.6,0.7,0.7,0.7\}$ | $\{0.3,0.4,0.5,0.5,0.5,0.5\}$ |
| $A_{6}$ | $\{0.5,0.6,0.6,0.6,0.6,0.6\}$ | $\{0.3,0.6,0.6,0.6,0.6\}$ | $\{0.4,0.5,0.5,0.5,0.5\}$ | $\{0.6,0.7,0.8,0.8,0.8,0.8\}$ |
| $A_{7}$ | $\{0.3,0.4,0.4,0.5,0.6,0.6\}$ | $\{0.5,0.7,0.7,0.7,0.7\}$ | $\{0.4,0.5,0.5,0.5,0.5\}$ | $\{0.2,0.3,0.6,0.6,0.6,0.6\}$ |
| $A_{8}$ | $\{0.3,0.5,0.5,0.5,0.5,0.5\}$ | $\{0.3,0.7,0.8,0.8,0.8\}$ | $\{0.4,0.7,0.9,0.9,0.9\}$ | $\{0.7,0.8,0.9,0.9,0.9,0.9\}$ |
| $A_{9}$ | $\{0.3,0.4,0.5,0.5,0.5,0.5\}$ | $\{0.4,0.5,0.7,0.7,0.7\}$ | $\{0.4,0.5,0.7,0.7,0.8\}$ | $\{0.2,0.6,0.6,0.6,0.6,0.6\}$ |
| $A_{10}$ | $\{0.3,0.5,0.5,0.5,0.5,0.5\}$ | $\{0.4,0.5,0.5,0.5,0.5\}$ | $\{0.5,0.5,0.5,0.5,0.5\}$ | $\{0.3,0.4,0.5,0.6,0.6,0.6\}$ |

Step 1. Calculate the distance $d_{i j}=d_{1}\left(A_{i}, A_{j}\right)$ by the formula (2), and let the attribute vector be $w=(0.4,0.2,0.1,0.3)$. Then we can get the distance matrix $D=\left(d_{i j}\right)_{n \times n}$ as:

$$
\left.\begin{array}{cccccccccc}
0 & 0.1080 & 0.1580 & 0.2013 & 0.1797 & 0.1983 & 0.2057 & 0.1937 & 0.1580 & 0.1627 \\
0.1080 & 0 & 0.1580 & 0.2240 & 0.1237 & 0.2050 & 0.1283 & 0.2123 & 0.1107 & 0.1373 \\
0.1580 & 0.1580 & 0 & 0.1680 & 0.1083 & 0.2470 & 0.1443 & 0.2423 & 0.1267 & 0.1113 \\
0.2013 & 0.2240 & 0.1680 & 0 & 0.2023 & 0.1610 & 0.2017 & 0.2617 & 0.1813 & 0.1487 \\
0.1797 & 0.1237 & 0.1083 & 0.2023 & 0 & 0.2233 & 0.1327 & 0.2347 & 0.1170 & 0.1117 \\
0.1983 & 0.2050 & 0.2470 & 0.1610 & 0.2233 & 0 & 0.1507 & 0.1327 & 0.1523 & 0.1437 \\
0.2057 & 0.1283 & 0.1443 & 0.2017 & 0.1327 & 0.1507 & 0 & 0.1847 & 0.0610 & 0.0797 \\
0.1937 & 0.2123 & 0.2423 & 0.2617 & 0.2347 & 0.1327 & 0.1847 & 0 & 0.1397 & 0.1830 \\
0.1580 & 0.1107 & 0.1267 & 0.1813 & 0.1170 & 0.1523 & 0.0610 & 0.1397 & 0 & 0.0667 \\
0.1627 & 0.1373 & 0.1113 & 0.1487 & 0.1117 & 0.1437 & 0.0797 & 0.1830 & 0.0667 & 0
\end{array}\right)
$$

Step 2. We still use the orthogonal clustering method to make analysis. Firstly, we choose the confidence level $\lambda$ and construct the corresponding $\lambda$-cutting matrix. Then we can group the accidents into several clusters as follows:
(1) If $0<\lambda \leq 0.1813$, then we get

$$
\left\{A_{1}, A_{2}, A_{3}, A_{4}, A_{5}, A_{6}, A_{7}, A_{8}, A_{9}, A_{10}\right\} .
$$

(2) If $0.1813<\lambda \leq 0.1830$, then we get

$$
\left\{A_{1}, A_{2}, A_{3}, A_{4}, A_{5}, A_{6}, A_{7}, A_{8}, A_{10}\right\},\left\{A_{9}\right\} .
$$

(3) If $0.1830<\lambda \leq 0.2017$, then we get

$$
\left\{A_{1}, A_{2}, A_{3}, A_{4}, A_{5}, A_{6}, A_{7}, A_{8}\right\},\left\{A_{9}\right\},\left\{A_{10}\right\} .
$$

(4) If $0.2017<\lambda \leq 0.2123$, then we get

$$
\left\{A_{1}\right\},\left\{A_{2}, A_{3}, A_{4}, A_{5}, A_{6}, A_{8}\right\},\left\{A_{7}\right\},\left\{A_{9}\right\},\left\{A_{10}\right\} .
$$

(5) If $0.2123<\lambda \leq 0.2240$, then we get

$$
\left\{A_{1}\right\},\left\{A_{3}, A_{4}, A_{5}\right\},\left\{A_{2}, A_{6}, A_{8}\right\},\left\{A_{7}\right\},\left\{A_{9}\right\},\left\{A_{10}\right\} .
$$

(6) If $0.2240<\lambda \leq 0.2347$, then we get $\left\{A_{1}\right\},\left\{A_{3}, A_{4}, A_{5}\right\},\left\{A_{2}\right\},\left\{A_{6}, A_{8}\right\},\left\{A_{7}\right\},\left\{A_{9}\right\},\left\{A_{10}\right\}$.
(7) If $0.2347<\lambda \leq 0.2423$, then we get
$\left\{A_{1}\right\},\left\{A_{3}, A_{4}\right\},\left\{A_{5}\right\},\left\{A_{2}\right\},\left\{A_{6}, A_{8}\right\},\left\{A_{7}\right\},\left\{A_{9}\right\},\left\{A_{10}\right\}$.
(8) If $0.2423<\lambda \leq 1$, then we get
$\left\{A_{1}\right\},\left\{A_{3}\right\},\left\{A_{4}\right\},\left\{A_{5}\right\},\left\{A_{2}\right\},\left\{A_{6}\right\},\left\{A_{8}\right\},\left\{A_{7}\right\},\left\{A_{9}\right\},\left\{A_{10}\right\}$.
To compare these two distance measures, we combine the final clustering results in one table. See Table 6.

From the above numerical example, we can find that the clustering results by using these two measures are quite different. Firstly, the accidents can just be clustered into eight groups if using the traditional distance measure, while the results derived by using our measure can be more exquisite. Secondly, there exist some differences among the results. For example, in the class 5, if we use the traditional Euclid distance, then the result is $\left\{A_{1}\right\},\left\{A_{2}, A_{3}, A_{4}, A_{5}, A_{6}, A_{8}\right\},\left\{A_{7}\right\},\left\{A_{9}\right\},\left\{A_{10}\right\}$. While if we use the distance measure proposed in this paper, then the result becomes

$$
\left\{A_{1}, A_{2}, A_{5}, A_{7}, A_{8}\right\},\left\{A_{4}, A_{6}\right\},\left\{A_{3}\right\},\left\{A_{9}\right\},\left\{A_{10}\right\} .
$$

Table 6. Clustering results

| Classes | Distance in this article | Euclid distance [10] |
| :---: | :---: | :---: |
| 10 | $\begin{aligned} & \left\{A_{1}\right\},\left\{A_{2}\right\},\left\{A_{5}\right\},\left\{A_{8}\right\},\left\{A_{4}\right\},\left\{A_{6}\right\}, \\ & \left\{A_{3}\right\},\left\{A_{7}\right\},\left\{A_{9}\right\},\left\{A_{10}\right\} \end{aligned}$ | $\begin{aligned} & \left\{A_{1}\right\},\left\{A_{3}\right\},\left\{A_{4}\right\},\left\{A_{5}\right\},\left\{A_{2}\right\},\left\{A_{6}\right\},\left\{A_{8}\right\}, \\ & \left\{A_{7}\right\},\left\{A_{9}\right\},\left\{A_{10}\right\} \end{aligned}$ |
| 9 | $\begin{aligned} & \left\{A_{1}\right\},\left\{A_{2}\right\},\left\{A_{5}, A_{8}\right\},\left\{A_{4}\right\},\left\{A_{6}\right\},\left\{A_{3}\right\}, \\ & \left\{A_{7}\right\},\left\{A_{9}\right\},\left\{A_{10}\right\} \end{aligned}$ | dull |
| 8 | $\begin{aligned} & \left\{A_{1}\right\},\left\{A_{2}\right\},\left\{A_{5}, A_{2}\right\},\left\{A_{4}, A_{6}\right\},\left\{A_{3}\right\}, \\ & \left\{A_{7}\right\},\left\{A_{9}\right\},\left\{A_{10}\right\} \end{aligned}$ | $\begin{aligned} & \left\{A_{1}\right\},\left\{A_{3}, A_{4}\right\},\left\{A_{5}\right\},\left\{A_{2}\right\},\left\{A_{6}, A_{8}\right\}, \\ & \left\{A_{7}\right\},\left\{A_{9}\right\},\left\{A_{10}\right\} \end{aligned}$ |
| 7 | $\begin{aligned} & \left\{A_{1}\right\},\left\{A_{2}, A_{5}, A_{8}\right\},\left\{A_{4}, A_{6}\right\},\left\{A_{3}\right\},\left\{A_{7}\right\}, \\ & \left\{A_{9}\right\},\left\{A_{10}\right\} \end{aligned}$ | $\begin{aligned} & \left\{A_{1}\right\},\left\{A_{3}, A_{4}, A_{5}\right\},\left\{A_{2}\right\},\left\{A_{6}, A_{8}\right\}, \\ & \left\{A_{7}\right\},\left\{A_{9}\right\},\left\{A_{10}\right\} \end{aligned}$ |
| 6 | $\begin{aligned} & \left\{A_{1}\right\},\left\{A_{2}, A_{5}, A_{7}, A_{8}\right\},\left\{A_{4}, A_{6}\right\}, \\ & \left\{A_{3}\right\},\left\{A_{9}\right\},\left\{A_{10}\right\} \end{aligned}$ | $\begin{aligned} & \left\{A_{1}\right\},\left\{A_{3}, A_{4}, A_{5}\right\},\left\{A_{2}, A_{6}, A_{8}\right\}, \\ & \left\{A_{7}\right\},\left\{A_{9}\right\},\left\{A_{10}\right\} \end{aligned}$ |
| 5 | $\begin{aligned} & \left\{A_{1}, A_{2}, A_{5}, A_{7}, A_{8}\right\},\left\{A_{4}, A_{6}\right\}, \\ & \left\{A_{3}\right\},\left\{A_{9}\right\},\left\{A_{10}\right\} \end{aligned}$ | $\begin{aligned} & \left\{A_{1}\right\},\left\{A_{2}, A_{3}, A_{4}, A_{5}, A_{6}, A_{8}\right\}, \\ & \left\{A_{7}\right\},\left\{A_{9}\right\},\left\{A_{10}\right\} \end{aligned}$ |
| 4 | $\begin{aligned} & \left\{A_{1}, A_{2}, A_{4}, A_{5}, A_{6}, A_{7}, A_{8}\right\}, \\ & \left\{A_{3}\right\},\left\{A_{9}\right\},\left\{A_{10}\right\} \end{aligned}$ | dull |
| 3 | $\begin{aligned} & \left\{A_{1}, A_{2}, A_{3}, A_{4}, A_{5}, A_{6}, A_{7}, A_{8}\right\}, \\ & \left\{A_{9}\right\},\left\{A_{10}\right\} \end{aligned}$ | $\begin{aligned} & \left\{A_{1}, A_{2}, A_{3}, A_{4}, A_{5}, A_{6}, A_{7}, A_{8}\right\}, \\ & \left\{A_{9}\right\},\left\{A_{10}\right\} \end{aligned}$ |
| 2 | $\left\{A_{1}, A_{2}, A_{3}, A_{4}, A_{5}, A_{6}, A_{7}, A_{8}, A_{9}\right\},\left\{A_{10}\right\}$ | $\left\{A_{1}, A_{2}, A_{3}, A_{4}, A_{5}, A_{6}, A_{7}, A_{8}, A_{10}\right\},\left\{A_{9}\right\}$ |
| 1 | $\left\{A_{1}, A_{2}, A_{3}, A_{4}, A_{5}, A_{6}, A_{7}, A_{8}, A_{9}, A_{10}\right\}$ | $\left\{A_{1}, A_{2}, A_{3}, A_{4}, A_{5}, A_{6}, A_{7}, A_{8}, A_{9}, A_{10}\right\}$ |

The main reason is that the distance measure proposed in this paper does not extend the shorter HFE until all the considered HFEs have the same length, and thus, we do not change the original information. Furthermore, in this paper, we take the hesitation degree into account. It is also an essential parameter when comparing and clustering the HFSs. In the following, we give some discussions:
(1) By comparing the original information, we can easily find that $A_{1}$ is close to $A_{7}$ and $A_{2}$ is close to $A_{7}$ too. It indicates that the pirate crimes happened in $A_{1}$, $A_{2}$ and $A_{7}$ are similar by coincidence. That is to say, the pirates who make $A_{1}, A_{2}$ and $A_{7}$ be the same. While, if we extend the shorter HFE using the traditional distance, then the original information will be changed. Thus, $A_{1}$ is totally different from $A_{7}$. In this way, the distance measure proposed in this paper is more convincing. After clustering the accidents, we can analyze the characteristics of these issues, speculate the strengths of these pirates, such as the armed equipment and so on. When handling these information, we can give advice to the passing ships and the local navy. Accurate analysis can reduce much loss and it can guarantee the security of international shipping.
(2) Comparing these two results, we can find that the result calculated by the distance measure proposed in this paper is more detailed and convincing. Since the lengths of HFEs differ very much, adding numbers to the shorter one change the original information greatly. Therefore, the clustering results will be totally different. From this example, we can see that it is essential and important to consider the influence of hesitation degrees and the original information. Only we consider all the influence factors, can the results be more convincing.

## 6. Concluding remarks

In order to keep the original information and consider the influence factors more comprehensively, in this paper, we have proposed a novel distance measure which combines the improved hesitation degree. What's more, a new method called orthogonal clustering method has been proposed and applied to cluster hesitant fuzzy information. Computational tests on the novel distance measure and the new clustering method have shown that the orthogonal clustering method is available and efficient. Furthermore, compared to the hesitant fuzzy netting clustering method ${ }^{27}$ in the numerical example, we have found that the two methods have the same clustering results, while the hesitant fuzzy netting clustering method ${ }^{27}$ needs people to
recognize, which is much more inconvenient. In addition, the distance measure proposed in this paper take the hesitation degrees into consideration, which has not changed the original information and thus is more reasonable and convincing.

## Acknowledgements

The authors would like to thank the anonymous reviewers for their insightful and constructive commendations that have led to an improved version of this paper. The work was supported by the National Natural Science Foundation of China (No. 71571123).

## References

1. L. A. Zadeh. Fuzzy sets. Inform and control 8 (1965) 338-353
2. K. Atanassov. Intuitionistic fuzzy sets. Fuzzy Sets and Systems 20 (1986) 87-96
3. K. Atanassov and G. Gargov. Interval-valued intuitionistic fuzzy sets. Fuzzy Sets and Systems 31 (1989) 343-349
4. D. Dubois and H. Prade. Fuzzy Sets and Systems: Theory and Applications. New York: Academic Press, 1980
5. S. Miyamoto. Remarks on basics of fuzzy sets and fuzzy multisets. Fuzzy Sets and Systems 156 (2005) 427-431
6. J.M. Mendel, R.I. John and F. Liu. Interval type-2 fuzzy sets made simple. IEEE Transactions on Fuzzy Systems 14 (6) (2006) 808-821
7. V. Torra. Hesitant Fuzzy Sets. International Journal of Intelligent Systems 25 (2010) 529-539
8. V. Torra and Y. Narukawa. On hesitant fuzzy sets and decision. The $18^{\text {th }}$ IEEE International Conference on Fuzzy Systems, Jeju Island, Korea, 2009, 1378-1382
9. M.M. Xia and Z.S. Xu. Hesitant fuzzy information aggregation in decision making. International Journal of Approximate Reasoning 52 (2011) 395-407
10. Z.S. Xu and M.M. Xia. Distance and similarity measures for hesitant fuzzy sets. Information Sciences 181 (2011) 2128-2138
11. R. M. Rodríguez, L. Martínez, V. Torra, Z.S. Xu and F. Herrera, "Hesitant fuzzy sets: State of the art and future directions", International Journal of Intelligent Systems, vol. 29, no. 6: Wiley, pp. 495-524, 2014
12. H.C. Liao, Z.S. Xu and X.J. Zeng. Distance and similarity measures for hesitant fuzzy linguistic term sets and their application in multi-criteria decision making. Information Sciences 271 (2014) 125-142
13. Z.S. Xu. Hesitant Fuzzy Sets Theory. Berlin, Heidelberg: Springer-Verlag, 2015
14. R. M. Rodríguez, B. Bedregal, H. Bustince, Y. Dong, B.

Farhadinia, C. Kahraman, L. Martínez, V. Torra, Y.J. Xu, Z.S. Xu, et al., "A Position and Perspective Analysis of Hesitant Fuzzy Sets on Information Fusion in Decision Making. Towards High Quality Progress", Information Fusion, vol. 29, pp. 89-97, 2016
15. E. Szmidt and J. Kacprzyk. A similarity measure for intuitionistic fuzzy sets and its application in supporting medical diagnostic reasoning. The 7th International Conference on Artificial Intelligence and Soft Computing. Zakopane, 2004, 388-393
16. D. F. Li and C. T. Cheng. New similarity measures of intuitionistic fuzzy sets and application to pattern recognitions. Pattern Recognition Letters 23 (2002) 221225
17. C.C. Wong and H.R. Lai. A grey-based clustering algorithm and its application on fuzzy system design. International Journal of Systems Science 34 (2003) 269281
18. B. Everitt, S. Landau and M. Leese. Cluster Analysis, 4th edition, London: Arnold 2001
19. C. Alzate and J.A.K. Suykens. Hierarchical kernel spectral clustering, Neural Networks, 35 (2012) 21-30
20. K. Polat. Classification of Parkinson's disease using feature weighting method on the basis of fuzzy C-means clustering. International Journal of Systems Science. 43 (2012) 597-609
21. X.L. Zhang and Z.S. Xu. Hesitant fuzzy agglomerative hierarchical clustering algorithms. International Journal of Systems Science. 46 (2013) 562-576
22. N. Chen, Z.S. Xu and M.M. Xia. Correlation coefficients of hesitant fuzzy set and their applications to clustering analysis. Applied Mathematical Modelling 37 (2013) 2197-2211
23. X.L. Zhang and Z.S. Xu. A MST clustering analysis method under hesitant fuzzy environment. Control and Cybernetics 41 (2012) 645-666
24. S. Miyamoto. Fuzzy sets in information retrieval and cluster analysis. Dordrecht: Kluwer, 1990: 181-192
25. N. Chen, Z.S. Xu and M.M. Xia. Hierarchical hesitant fuzzy K-means clustering algorithm. Applied Mathematics-A Journal of Chinese Universities 2014,29(1):1-17
26. D. Q. Li, W. Y. Zeng and J. H. Li. New distance and similarity measures on hesitant fuzzy sets and their applications in multiple criteria decision making. Engineering Applications of Artificial Intelligence 40 (2015) 11-16
27. X.D. Liu, J. J. Zhu, S.F. Liu. Similarity measure of hesitant fuzzy sets based on symmetric cross entropy and its application in clustering analysis. Control and Decision 29 (2014) 1816-182


[^0]:    * Corresponding author.

