

## New distance and similarity measures for hesitant fuzzy soft sets

K. Rezaei<sup>1</sup> and H. Rezaei<sup>1</sup>

<sup>1</sup>*Department of Computer Science, University of Sistan and Baluchestan, Zahedan, Iran*

kamranrezaei@pgs.usb.ac.ir, hrezaei@cs.usb.ac.ir

### Abstract

The hesitant fuzzy soft set (HFSS), as a combination of hesitant fuzzy and soft sets, is regarded as a useful tool for dealing with the uncertainty and ambiguity of real-world problems. In HFSSs, each element is defined in terms of several parameters with arbitrary membership degrees. In addition, distance and similarity measures are considered as the important tools in different areas such as pattern recognition, clustering, medical diagnosis, and the like. For this purpose, the present study aimed to evaluate the distance and similarity measures for HFSSs by using well-known Hamming, Euclidean, and Minkowski distance measures. Further, some examples were used to demonstrate that these measures fail to perform well in some applications. Accordingly, new distance and similarity measures were proposed by considering a hesitance index for HFSSs and the effect of considering hesitance index was shown by using an example of pattern recognition. Finally, the application of the proposed measures and hesitance index was investigated in the clustering and decision-making problem, respectively. In conclusion, the use of the proposed measures in clustering and hesitance index in decision-making can provide better and more reasonable results.

**Keywords:** Hesitant fuzzy set, Hesitant fuzzy soft set, Hesitance index, Distance measure, Similarity measure, Clustering

**2010 MSC:** 03B52, 03E72

## 1 Introduction

Nowadays, we are encountering with ambiguous phenomena which cannot be easily described by using the classical theory of sets. Thus, Zadeh [44] introduced the fuzzy sets as a development of classical sets. In fuzzy sets, each element is related to that set with a degree in the interval  $[0,1]$ . A large number of studies have focused on fuzzy sets in a variety of fields and accordingly several extensions have been introduced in this regard. For example, Zadeh [45] proposed the type-2 fuzzy set and interval-valued fuzzy set. The membership degree of an element in a type-2 fuzzy set and an interval-valued fuzzy set can be defined in the form of a type-1 fuzzy set, and a subinterval of the interval  $[0,1]$ , respectively. Dubois and Prade [15] introduced type- $n$  fuzzy sets, which are the generalizations of type-2 fuzzy sets, and the membership degrees can be considered as a type- $n - 1$  fuzzy set. Atanassov [4] suggested intuitionistic fuzzy sets which are more appropriate than fuzzy sets when faced with uncertainty and ambiguity, due to the membership, non-membership, and hesitation degrees for each element. Recently, Torra and Narukawa [36] and Torra [35] emphasized that a set of values may be available as the membership degree for an element, and accordingly the concept of hesitant fuzzy set (HFS) was introduced. Some researchers have focused on these sets in various fields. For example, Xu et al. [41] proposed a new hesitant fuzzy programming method for hybrid multiple attribute decision-making problems. In another study, Dong et al. [14] developed an extended VIKOR method for solving multiple criteria decision-making problems with linguistic HFSs. Wan et al. [37] proposed a new hesitant fuzzy mathematical programming method for hybrid multi-criteria group decision-making with hesitant fuzzy truth degrees and incomplete criteria weight information. In addition, Sun et al. [32] presented an innovative TOPSIS based on a novel synthetic correlation coefficient between HFSs. Guan et al. [20] proposed a synthetic correlation coefficient between the HFSs with respect to the integrality,

distribution, and length of the membership. Further, the grey relational analysis for HFSSs and interval-valued HFSSs were considered in several studies [33, 19, 34].

According to Molodtsov [28], there are some difficulties in different methods such as fuzzy sets, which are used to deal with ambiguity, maybe due to the lack of a parameterization tool for these methods. Hence, a mathematical tool was introduced to deal with uncertainty called “soft sets”. The use of these sets has attracted the researchers’ attentions. For example, Maji et al. [25] introduced fuzzy soft set by combining soft set and fuzzy set. In addition, Maji represented the concept of an intuitionistic fuzzy soft set using the combination of soft sets and intuitionistic fuzzy sets [26]. Yang et al. [43] proposed the interval-valued fuzzy soft set by combining the interval-valued fuzzy set and the soft set. In another study, Jiang et al. [22] suggested the interval-valued intuitionistic fuzzy soft theory. Further, Feng et al. [17] developed the concept of the rough set [29] and introduced the soft rough set.

Recently, the hesitant fuzzy soft set (HFSS) was defined as a combination of soft set and hesitant fuzzy set [5]. It is considered as a powerful tool for dealing with uncertainty in real world problems, especially those related to decision-making. For example, candidates should be evaluated by a group of decision makers on some criteria in deciding whether to assign a job to a person from among several candidates. Some decision makers are often hesitant about giving their opinions on the candidates for different criteria, which may be related to inadequate knowledge about some criteria. Further, some may regard a set of criteria as more important than others, while some others consider another relevant set of criteria. In some circumstances, decision makers are interested in considering a subset of criteria rather than the whole criteria for evaluating candidates. In such a case, the use of HFSS plays a significant role in collecting comments and solving decision-making problem. These sets have been highlighted by many researchers [38, 12, 8, 7, 13, 2]. For example, Bin [8] introduced generalized hesitant fuzzy soft sets and some operators related to these sets. Beg and Rashid [7] developed the fuzzy TOPSIS method for HFSSs. Das et al. [13] represented the concept of correlation coefficient and some of its properties for HFSSs, as well as an algorithm for applying this correlation coefficient to a decision-making problem. In another study, Alshehri et al. [2] applied the notion of HFSS for dealing with several kinds of theories in BCK algebras, and evaluated the application of HFSSs in decision-making.

The distance and similarity measures have been considered as important tools in many areas such as pattern recognition, clustering, decision-making, and medical diagnosis and the like, among which the well-known measures for evaluating the distance between two fuzzy sets  $A$  and  $B$  on the universal set  $U = \{u_1, u_2, \dots, u_n\}$  are as follows [23]:

- the Hamming distance:  $d_h(A, B) = \sum_{i=1}^n | \mu_A(u_i) - \mu_B(u_i) |$
- the normalized Hamming distance:  $d_{nh}(A, B) = \frac{1}{n} \sum_{i=1}^n | \mu_A(u_i) - \mu_B(u_i) |$
- the Euclidean distance:  $d_e(A, B) = (\sum_{i=1}^n | \mu_A(u_i) - \mu_B(u_i) |^2)^{1/2}$
- the normalized Euclidean distance:  $d_{ne}(A, B) = (\frac{1}{n} \sum_{i=1}^n | \mu_A(u_i) - \mu_B(u_i) |^2)^{1/2}$

A large number of studies have been conducted on developing distance and similarity measures of fuzzy, type-2 fuzzy, interval-valued fuzzy, intuitionistic fuzzy sets, and soft sets [39, 6, 21, 3, 18, 9, 27]. Regarding HFSSs, several distance and similarity measures have been proposed. For example, Xu and Xia [42], introduced various distance measures and accordingly provided similarity measures of hesitant fuzzy sets. Farhadinia [16] investigated the relationship between entropy, similarity, and distance for hesitant fuzzy sets and interval-valued hesitant fuzzy sets. Zhang and Xu [48] proposed the distance measures and two new similarity measures of hesitant fuzzy sets by considering a new hesitant index for hesitant fuzzy sets. In another study, Zeng et al. [46] defined several new similarity and distance measures of hesitant fuzzy sets by considering the degree of hesitation for these sets. Recently, some studies emphasized the imprecise and vague data clustering in which similarity and distance measures were taken into consideration. For example, Zeshui [47] provided an intuitionistic fuzzy hierarchical clustering algorithm. In addition, Chen et al. [10] proposed a clustering algorithm for the HFSSs. Aliahmadipour and Eslami [1] represented a new hierarchical clustering algorithm for clustering generalized hesitant fuzzy sets introduced in [30]. However, to the best of our knowledge, no work has been conducted on HFSSs.

In this study, the well-known Hamming, Euclidean, and Minkowski distance measures for HFSSs were evaluated, along with their corresponding similarity measures. In addition, an example was provided to demonstrate the lack of appropriate performance for these measures in some cases. Further, distance and similarity measures play an important role in many areas and their inappropriate performance can disturb the final result. In fact, the inappropriate performance of these measures is related to the lack of considering the hesitation in the HFSS. Therefore, this problem has motivated the authors to propose new distance and similarity measures for HFSSs by considering a hesitance index for these sets. Then, the use of proposed measures led to more reasonable results based on a pattern recognition example. Furthermore, several projects evaluated by several evaluators in the form of HFSSs, were clustered based on

the single linkage hierarchical clustering algorithm [31]. Additionally, the proposed hesitance index was used to improve a decision-making algorithm.

This paper is organized as follows:

Section 2 investigates some required definitions related to hesitant fuzzy sets, soft sets, and hesitant fuzzy soft sets. Section 3 indicates the well-known distance and similarity measures for HFSSs, as well as demonstrating the drawback of these measures by an example. Section 4, presents the new distance and similarity measures of HFSSs. Hesitant fuzzy soft data clustering, along with the application of the proposed hesitance index in decision-making is evaluated in Section 5. Finally, the conclusion is elaborated in Section 6.

## 2 Preliminaries

Some definitions required for the next sections will be presented in this section. Throughout this paper, it is assumed that  $U = \{u_1, u_2, \dots, u_n\}$  is a universal set and  $E = \{e_1, e_2, \dots, e_m\}$  is a set of parameters.

**Definition 2.1.** [35] The hesitant fuzzy set  $H$  over  $U$  is a function that, returns a subset of the interval  $[0, 1]$  when applied to  $U$ .

In [40], for a simpler understanding, a hesitant fuzzy set, like  $H$  over the universal set  $U$ , is represented using the mathematical symbol as follows:

$$H = \{ \langle u, h_H(u) \rangle \mid u \in U \} \quad (1)$$

Where  $h_H(u)$  is a set of possible membership degrees of the element  $u \in U$  in the set  $H$ .  $h = h_H(u)$  is called a hesitant fuzzy element (HFE).

Since the number of membership degrees may be different in different HFEs, the number of elements in the HFE  $h_H(u)$  is shown by  $l(h_H(u))$ .

**Definition 2.2.** [40] Let  $h(u) = \{\alpha_i\}_{i=1}^{l(h(u))}$  be a HFE,  $s(h(u))$  is the score of  $h(u)$  and is defined as:

$$s(h(u)) = \frac{1}{l(h(u))} \sum_{i=1}^{l(h(u))} \alpha_i \quad (2)$$

**Definition 2.3.** [11] Let  $h(u) = \{\alpha_i\}_{i=1}^{l(h(u))}$  be a HFE,  $\sigma(h(u))$  is the deviation degree of  $h(u)$  and is defined as:

$$\sigma(h(u)) = \sqrt{\frac{1}{l(h(u))} \sum_{i=1}^{l(h(u))} (\alpha_i - s(h(u)))^2} \quad (3)$$

In fact, the deviation degree presented in Definition 2.3 reflects the same standard deviation for the HFE  $h(u)$ .

In [11], the score and deviation degree of the HFEs are used for comparison of these elements. Suppose that  $h_1$  and  $h_2$  are two HFEs,  $s(h_1) < s(h_2)$ , implies that  $h_1 < h_2$  and assuming  $s(h_1) = s(h_2)$ , then the following relationships are used:

- (i)  $\sigma(h_1) = \sigma(h_2) \Rightarrow h_1 = h_2$
- (ii)  $\sigma(h_1) > \sigma(h_2) \Rightarrow h_1 < h_2$
- (iii)  $\sigma(h_1) < \sigma(h_2) \Rightarrow h_1 > h_2$

**Assumption 2.1.** [42] Assume that  $h_1$  and  $h_2$  be two HFEs over universal set  $U$ , it might be for  $u \in U$ ,  $l(h_1(u)) \neq l(h_2(u))$ . In this case,  $l_u$  is defined as  $l_u = \max\{l(h_1(u)), l(h_2(u))\}$ . To make a comparison between the two HFEs, their membership degrees must first be ordered. Also, it is proposed that a HFE with the fewer number of membership degrees to be developed to perform a correct comparison between two HFEs so that the number of membership degrees to be equal in both elements.

In this paper, the membership degrees in the elements are ordered in descending order. Thus,  $h_{H_1}^{\tau(j)}(u)$  and  $h_{H_2}^{\tau(j)}(u)$  represent the  $j$ -th largest value in  $h_{H_1}(u)$  and  $h_{H_2}(u)$ , respectively. Furthermore, a HFE with the fewer number of membership degrees is developed by adding the smallest membership degree in it, until their number is equal with the membership degrees of other HFE.

Farhadinia [16] proposed two methods to order the HFSs as follows:

**Definition 2.4.** [16] Let  $H_1$  and  $H_2$  be two HFSs over  $U$ , then the component-wise and total ordering methods for them are defined as:

(i) (Component-wise ordering)  $H_1 \leq H_2$  iff:

$$h_{H_1}^{\tau(j)}(u_i) \leq h_{H_2}^{\tau(j)}(u_i), \quad i = 1, \dots, n, \quad j = 1, \dots, l_{u_i},$$

(ii) (Total ordering)  $H_1 \preceq H_2$  iff:

$$s(H_1) \leq s(H_2)$$

In this definition,  $s(H_1)$  and  $s(H_2)$ , respectively, represent the score of the HFSs  $H_1$  and  $H_2$  and are defined as:

$$s(H_1) = \frac{1}{n} \sum_{i=1}^n \left( \frac{1}{l_{u_i}} \sum_{j=1}^{l_{u_i}} h_{H_1}^{\tau(j)}(u_i) \right), \quad s(H_2) = \frac{1}{n} \sum_{i=1}^n \left( \frac{1}{l_{u_i}} \sum_{j=1}^{l_{u_i}} h_{H_2}^{\tau(j)}(u_i) \right).$$

In [48], for a HFE, like  $h_H$  of a hesitant fuzzy set  $H$ , a hesitance index has been defined, which is indicated by  $\hbar_H$  and shows the hesitance degree in  $h_H$ . Then, by considering the hesitance index, two new ordering methods have been defined for the hesitant fuzzy sets.

**Definition 2.5.** [48] Let  $H_1$  and  $H_2$  be two HFSs over the universal set  $U = \{u_1, u_2, \dots, u_n\}$ , then the component-wise and total ordering methods for them are defined as:

(i) (Component-wise ordering)  $H_1 \leq H_2$  if and only if for  $i = 1, \dots, n$  and for  $j = 1, \dots, l_{u_i}$ , the following condition holds:

$$h_{H_1}^{\tau(j)}(u_i) \leq h_{H_2}^{\tau(j)}(u_i) \quad \& \quad \hbar_{H_2}(u_i) \leq \hbar_{H_1}(u_i)$$

(ii) (Total ordering)  $H_1 \preceq H_2$  if and only if the following condition holds:

$$s(H_1) < s(H_2) \quad \text{or} \quad s(H_1) = s(H_2) \quad \text{and} \quad \hbar(H_2) \leq \hbar(H_1)$$

**Definition 2.6.** [28] Let  $A \subseteq E$ , A pair  $(G, A)$  is called a soft set over  $U$  if and only if  $G$  is a mapping from  $A$  to the power set  $P(U)$  (the set of all subsets of  $U$ ). In other words, it can be said that the soft set  $(G, A)$  is a parameterized family of subsets of the universal set  $U$ .

**Definition 2.7.** [25] Let  $A \subseteq E$  and  $\tilde{P}(U)$  be the set of all fuzzy subsets of  $U$ . A pair  $(\tilde{G}, A)$  is called a fuzzy soft set over  $U$ , where  $\tilde{G}$  is a mapping from  $A$  to the  $\tilde{P}(U)$ .

**Definition 2.8.** [5] Let  $A \subseteq E$  and  $H(U)$  be the set of all hesitant fuzzy sets in  $U$ . A pair  $(\tilde{F}, A)$  is called a hesitant fuzzy soft set over  $U$ , where  $\tilde{F}$  is a mapping from  $A$  to the  $H(U)$ .

**Definition 2.9.** [12] Let  $M = (\tilde{F}, A)$  be a HFSS over  $U$  and  $A \subseteq E$ , the score of  $M$  is defined as:

$$s(M) = \frac{1}{m} \sum_{i=1}^m \frac{1}{n} \sum_{j=1}^n \frac{1}{l_{e_i, u_j}} \sum_{k=1}^{l_{e_i, u_j}} h_M^{\tau(k)}(e_i)(u_j) \quad (4)$$

Where  $h_M(e_i)(u_j)$  is the set of membership degrees of  $u_j$  in  $\tilde{F}(e_i)$ .

**Assumption 2.2.** Let  $M$  and  $N$  be two HFSSs over  $U$ , it might be for  $u \in U$  and  $e \in E$ ,  $l(h_M(e)(u)) \neq l(h_N(e)(u))$ . Hence,  $l_{e,u}$  is defined as  $l_{e,u} = \max\{l(h_M(e)(u)), l(h_N(e)(u))\}$ . Also, to make a comparison between  $h_M(e)(u)$  and  $h_N(e)(u)$ , their membership degrees must be ordered. In this paper, the membership degrees are ordered in descending order. Therefore,  $h_M^{\tau(k)}(e)(u)$  and  $h_N^{\tau(k)}(e)(u)$  represent the  $k$ -th largest value in  $h_M(e)(u)$  and  $h_N(e)(u)$ , respectively. To make a correct comparison between  $h_M(e)(u)$  and  $h_N(e)(u)$  the number of membership degrees of them must be equal. For example, if there are a fewer number of member degrees in  $h_M(e)(u)$ , then  $h_M(e)(u)$  is developed by adding its smallest membership degree until it has the same length with  $h_N(e)(u)$ .

### 3 Well-known distance and similarity measures for hesitant fuzzy soft sets

In this section, the well-known Hamming, Euclidean, and Minkowski distance measures and their corresponding similarity measures are investigated for the HFSSs and illustrated by an example; these measures may not have a good performance in some cases.

**Definition 3.1.** Let  $M = (\tilde{F}, A)$  and  $N = (\tilde{G}, B)$  be two HFSSs over the universal set  $U$  and  $A$  and  $B$  be two subsets of the parameters set  $E$ ,  $d(M, N)$  is a distance measure between  $M$  and  $N$ , if it satisfies the following properties:

$$(D1) \quad 0 \leq d(M, N) \leq 1$$

$$(D2) \quad d(M, N) = 0 \text{ iff } M = N$$

$$(D3) \quad d(M, N) = d(N, M)$$

And if  $Q = (\tilde{Z}, C)$  be also another HFSS defined over  $U$  and  $C \subseteq E$  such that  $M \leq N \leq Q$ , then The following conditions must be satisfied:

$$(D4) \quad d(M, N) \leq d(M, Q), d(N, Q) \leq d(M, Q)$$

Also,  $s(M, N)$  is the similarity measure between  $M$  and  $N$ , if it satisfies the following properties:

$$(S1) \quad 0 \leq s(M, N) \leq 1$$

$$(S2) \quad s(M, N) = 1 \text{ iff } M = N$$

$$(S3) \quad s(M, N) = s(N, M)$$

And again, taking into account the set  $Q = (\tilde{Z}, C)$  such that  $M \leq N \leq Q$ , the following condition must be satisfied:

$$(S4) \quad s(M, Q) \leq s(M, N), s(M, Q) \leq s(N, Q)$$

**Remark 3.1.** From the definition 3.1 it is concluded that  $s(M, N) = 1 - d(M, N)$ .

**Remark 3.2.** Let  $M = (\tilde{F}, A)$  and  $N = (\tilde{G}, B)$  be two HFSSs over  $U$  such that  $A \subseteq E$  and  $B \subseteq E$ . To calculate the distance between these two sets, the set of their parameters is firstly developed as follows (Suppose that  $C = A \cap B$ ):

- $\forall e \in B \setminus C, \forall i = 1, \dots, n, \tilde{F}(e)(u_i) = \{0\}$
- $\forall e \in A \setminus C, \forall i = 1, \dots, n, \tilde{G}(e)(u_i) = \{0\}$

Therefore in the following, without loss of generality, the HFSSs will be considered on the set of main parameters, i.e.,  $E$ .

Considering the Remark 3.2 and the Hamming and Euclidean distances defined in [12], the hesitant soft normalized Hamming distance ( $d_{hsnh}$ ), the hesitant soft normalized Euclidean distance ( $d_{hsne}$ ) and the hesitant soft normalized minkowski distance ( $d_{hsnm}$ ) for the HFSSs  $M$  and  $N$  over  $U$  are defined as follows:

$$d_{hsnh}(M, N) = \frac{1}{m} \sum_{i=1}^m \left\{ \frac{1}{n} \sum_{j=1}^n \left[ \frac{1}{l_{e_i, u_j}} \sum_{k=1}^{l_{e_i, u_j}} \left| h_M^{\tau(k)}(e_i)(u_j) - h_N^{\tau(k)}(e_i)(u_j) \right| \right] \right\} \quad (5)$$

$$d_{hsne}(M, N) = \left( \frac{1}{m} \sum_{i=1}^m \left\{ \frac{1}{n} \sum_{j=1}^n \left[ \frac{1}{l_{e_i, u_j}} \sum_{k=1}^{l_{e_i, u_j}} \left| h_M^{\tau(k)}(e_i)(u_j) - h_N^{\tau(k)}(e_i)(u_j) \right|^2 \right] \right\} \right)^{\frac{1}{2}} \quad (6)$$

$$d_{hsnm}(M, N) = \left( \frac{1}{m} \sum_{i=1}^m \left\{ \frac{1}{n} \sum_{j=1}^n \left[ \frac{1}{l_{e_i, u_j}} \sum_{k=1}^{l_{e_i, u_j}} \left| h_M^{\tau(k)}(e_i)(u_j) - h_N^{\tau(k)}(e_i)(u_j) \right|^\lambda \right] \right\} \right)^{\frac{1}{\lambda}} \quad (7)$$

Where  $\lambda > 0$ ,  $l_{e_i, u_j} = \max\{l(h_M(e_i)(u_j)), l(h_N(e_i)(u_j))\}$  and  $h_M^{\tau(k)}(e_i)(u_j)$  and  $h_N^{\tau(k)}(e_i)(u_j)$  are respectively the  $k$ -th largest value in  $h_M(e_i)(u_j)$  and  $h_N(e_i)(u_j)$ .

According to Remark 3.1, the similarity measures for two HFSSs  $M$  and  $N$  are also defined as  $s_{hsnh}(M, N) = 1 - d_{hsnh}(M, N)$ ,  $s_{hsne}(M, N) = 1 - d_{hsne}(M, N)$  and  $s_{hsnm}(M, N) = 1 - d_{hsnm}(M, N)$ .

In some cases, the measures expressed above may not have an appropriate performance, as indicated in the following example.

	$e_1$	$e_2$	$e_3$
$u$	{0.7, 0.6, 0.5}	{0.6, 0.5, 0.4}	{0.8, 0.7}

Table 1: The HFSS  $M$ 

	$e_1$	$e_2$	$e_3$
$u$	{0.5, 0.45, 0.4}	{0.4, 0.3, 0.2}	{0.6, 0.5}

Table 2: The HFSS  $N$ 

	$e_1$	$e_2$	$e_3$
$u$	{0.9, 0.5, 0.35}	{0.8, 0.3, 0.2}	{1.0, 0.5}

Table 3: The HFSS  $O$ 

**Example 3.1.** Let  $M = (\tilde{F}, E)$ ,  $N = (\tilde{G}, E)$  and  $O = (\tilde{H}, E)$  be three HFSSs over  $U$ , which have been defined in Table 1, Table 2 and Table 3, respectively.

Suppose that we have to find a set between the sets  $N$  and  $O$  which is more similar to the set  $M$ . Using each of the similarity measures  $s_{hsnh}$ ,  $s_{hsne}$  and  $s_{hsnm}$ , it is concluded that the similarity of the sets  $M$  and  $N$  is equal to that of  $M$  and  $O$ :

$$s_{hsnh}(M, N) = s_{hsnh}(M, O) = 0.8167$$

$$s_{hsne}(M, N) = s_{hsne}(M, O) = 0.8137$$

$$s_{hsnm}(M, N) = s_{hsnm}(M, O) = 0.8073, \quad (\lambda = 6)$$

$$s_{hsnm}(M, N) = s_{hsnm}(M, O) = 0.8048, \quad (\lambda = 10)$$

however, it is observed that the hesitancy of  $\tilde{G}(e_2)(u)$  and  $\tilde{G}(e_3)(u)$  are respectively similar to the hesitancy of  $\tilde{F}(e_2)(u)$  and  $\tilde{F}(e_3)(u)$ . Therefore, taking into account the equal similarity of the sets  $N$  and  $O$  to the set  $M$  is not logical.

## 4 New distance and similarity measures for hesitant fuzzy soft sets

As observed, the distance and similarity measures introduced so far may not have desirable outcomes in some cases. Hence, taking into account the concept of standard deviation, new distance and similarity measures for HFSSs are proposed. First, some definitions related to HFSSs are expressed.

**Definition 4.1.** Let  $M = (\tilde{F}, A)$  be a HFSS over  $U$ , the standard deviation for  $\tilde{F}(e)(u) = \{\gamma_i\}_{i=1}^{l(\tilde{F}(e)(u))}$ ,  $e \in A$  and  $u \in U$ , is defined as:

$$\sigma(\tilde{F}(e)(u)) = \sqrt{\left(\frac{1}{l(\tilde{F}(e)(u))}\right) \sum_{i=1}^{l(\tilde{F}(e)(u))} (\gamma_i - s(\tilde{F}(e)(u)))^2} \quad (8)$$

Where  $l(\tilde{F}(e)(u))$  denotes the number of elements in  $\tilde{F}(e)(u)$  and  $s(\tilde{F}(e)(u)) = \frac{1}{l(\tilde{F}(e)(u))} \sum_{i=1}^{l(\tilde{F}(e)(u))} \gamma_i$  is the score of  $\tilde{F}(e)(u)$ .

**Definition 4.2.** According to the standard deviation defined for  $\tilde{F}(e)(u)$ , the standard deviation for  $\tilde{F}(e)$ , for each  $e \in A$ , as well as for the HFSS  $M$ , is defined by equations 9 and 10, respectively:

$$\sigma(\tilde{F}(e)) = \frac{1}{n} \sum_{j=1}^n \sigma(\tilde{F}(e)(u_j)) \quad (9)$$

$$\sigma(M) = \frac{1}{m} \sum_{i=1}^m \sigma(\tilde{F}(e_i)) \quad (10)$$

As is clear from Definition 4.1, the greater standard deviation for  $\tilde{F}(e)(u)$  indicates the greater distance of their membership degrees to their average. Therefore, the standard deviation defined for  $\tilde{F}(e)(u)$  is considered as a hesitance

index for it. Therefore, the greater standard deviation for  $\tilde{F}(e)(u)$  results in the greater hesitation of its membership degrees. Considering this hesitance index, and using the Definition 2.5, the ordering for the two HFSSs  $M = (\tilde{F}, A)$  and  $N = (\tilde{G}, B)$  is presented in the form of Definition 4.3.

**Definition 4.3.** Let  $M = (\tilde{F}, A)$  and  $N = (\tilde{G}, B)$  be two HFSSs over  $U$ . Then the two component-wise and total ordering methods for them are defined as:

- (i) (Component-wise ordering)  $M \leq N$  if and only if for  $i = 1, \dots, m$ ,  $j = 1, \dots, n$ , and for  $k = 1, \dots, l_{e_i, u_j}$  the following conditions are satisfied:

$$A \subseteq B$$

$$h_M^{\tau(k)}(e_i)(u_j) \leq h_N^{\tau(k)}(e_i)(u_j) \quad \& \quad \sigma(\tilde{G}(e_i)(u_j)) \leq \sigma(\tilde{F}(e_i)(u_j))$$

- (ii) (Total ordering)  $M \preceq N$  if and only if the following condition is satisfied:

$$s(M) < s(N) \quad \text{or} \quad s(M) = s(N) \quad \text{and} \quad \sigma(N) \leq \sigma(M)$$

Where  $s(M)$  and  $s(N)$  represent the scores of the HFSSs  $M$  and  $N$ , respectively.

**Definition 4.4.** Let  $M = (\tilde{F}, A)$  and  $N = (\tilde{G}, B)$  be two HFSSs. It is said that sets  $M$  and  $N$  are equal if and only if  $M \leq N$  and  $N \leq M$ .

In the following, the new distance and similarity measures for HFSSs are proposed.

**Definition 4.5.** Let  $M = (\tilde{F}, E)$  and  $N = (\tilde{G}, E)$  be two HFSSs over  $U$ , The three distance measures, the new hesitant soft normalized Hamming distance ( $d_{nhshnh}$ ), the new hesitant soft normalized Euclidean distance ( $d_{nhshne}$ ) and the new hesitant soft normalized Minkowski distance ( $d_{nhshnm}$ ) are defined as follows:

$$d_{nhshnh}(M, N) = \frac{1}{m} \sum_{i=1}^m \left\{ \frac{1}{2n} \sum_{j=1}^n \left[ \left| \sigma(\tilde{F}(e_i)(u_j)) - \sigma(\tilde{G}(e_i)(u_j)) \right| + \frac{1}{l_{e_i, u_j}} \sum_{k=1}^{l_{e_i, u_j}} \left| h_M^{\tau(k)}(e_i)(u_j) - h_N^{\tau(k)}(e_i)(u_j) \right| \right] \right\} \quad (11)$$

$$d_{nhshne}(M, N) = \left( \frac{1}{m} \sum_{i=1}^m \left\{ \frac{1}{2n} \sum_{j=1}^n \left[ \left| \sigma(\tilde{F}(e_i)(u_j)) - \sigma(\tilde{G}(e_i)(u_j)) \right|^2 + \frac{1}{l_{e_i, u_j}} \sum_{k=1}^{l_{e_i, u_j}} \left| h_M^{\tau(k)}(e_i)(u_j) - h_N^{\tau(k)}(e_i)(u_j) \right|^2 \right] \right\} \right)^{\frac{1}{2}} \quad (12)$$

$$d_{nhshnm}(M, N) = \left( \frac{1}{m} \sum_{i=1}^m \left\{ \frac{1}{2n} \sum_{j=1}^n \left[ \left| \sigma(\tilde{F}(e_i)(u_j)) - \sigma(\tilde{G}(e_i)(u_j)) \right|^\lambda + \frac{1}{l_{e_i, u_j}} \sum_{k=1}^{l_{e_i, u_j}} \left| h_M^{\tau(k)}(e_i)(u_j) - h_N^{\tau(k)}(e_i)(u_j) \right|^\lambda \right] \right\} \right)^{\frac{1}{\lambda}} \quad (13)$$

According to Remark 3.1, the new similarity measures for HFSSs are also defined as:

$$\begin{aligned} s_{nhshnh}(M, N) &= 1 - d_{nhshnh}(M, N) \\ &= 1 - \frac{1}{m} \sum_{i=1}^m \left\{ \frac{1}{2n} \sum_{j=1}^n \left[ \left| \sigma(\tilde{F}(e_i)(u_j)) - \sigma(\tilde{G}(e_i)(u_j)) \right| + \frac{1}{l_{e_i, u_j}} \sum_{k=1}^{l_{e_i, u_j}} \left| h_M^{\tau(k)}(e_i)(u_j) - h_N^{\tau(k)}(e_i)(u_j) \right| \right] \right\} \end{aligned} \quad (14)$$

$$\begin{aligned} s_{nhshne}(M, N) &= 1 - d_{nhshne}(M, N) \\ &= 1 - \left( \frac{1}{m} \sum_{i=1}^m \left\{ \frac{1}{2n} \sum_{j=1}^n \left[ \left| \sigma(\tilde{F}(e_i)(u_j)) - \sigma(\tilde{G}(e_i)(u_j)) \right|^2 + \frac{1}{l_{e_i, u_j}} \sum_{k=1}^{l_{e_i, u_j}} \left| h_M^{\tau(k)}(e_i)(u_j) - h_N^{\tau(k)}(e_i)(u_j) \right|^2 \right] \right\} \right)^{\frac{1}{2}} \end{aligned} \quad (15)$$

$$\begin{aligned} s_{nhshnm}(M, N) &= 1 - d_{nhshnm}(M, N) \\ &= 1 - \left( \frac{1}{m} \sum_{i=1}^m \left\{ \frac{1}{2n} \sum_{j=1}^n \left[ \left| \sigma(\tilde{F}(e_i)(u_j)) - \sigma(\tilde{G}(e_i)(u_j)) \right|^\lambda + \frac{1}{l_{e_i, u_j}} \sum_{k=1}^{l_{e_i, u_j}} \left| h_M^{\tau(k)}(e_i)(u_j) - h_N^{\tau(k)}(e_i)(u_j) \right|^\lambda \right] \right\} \right)^{\frac{1}{\lambda}} \end{aligned} \quad (16)$$

**Theorem 4.1.** Let  $M = (\tilde{F}, E)$ ,  $N = (\tilde{G}, E)$  and  $Z = (\tilde{H}, E)$  be three HFSSs, the new distance and similarity measures satisfy the properties mentioned in Definition 3.1.

*Proof.* The proof is performed for the new hesitant soft normalized Minkowski distance ( $d_{nhsnm}$ ). Clearly, by placing  $\lambda = 1$  and  $\lambda = 2$ , the proof will be obtained for  $d_{nhshh}$  and  $d_{nhshs}$ , respectively.

• **Property (D1):**

It is straightforward.

• **Property (D2):**

Suppose that  $d_{nhsnm}(M, N) = 0$ , that is:

$$\left( \frac{1}{m} \sum_{i=1}^m \left\{ \frac{1}{2n} \sum_{j=1}^n \left[ \left| \sigma(\tilde{F}(e_i)(u_j)) - \sigma(\tilde{G}(e_i)(u_j)) \right|^\lambda + \frac{1}{l_{e_i, u_j}} \sum_{k=1}^{l_{e_i, u_j}} \left| h_M^{\tau(k)}(e_i)(u_j) - h_N^{\tau(k)}(e_i)(u_j) \right|^\lambda \right] \right\} \right)^{\frac{1}{\lambda}} = 0$$

It is clear that for  $i = 1, \dots, m$ ,  $j = 1, \dots, n$  and  $k = 1, \dots, l_{e_i, u_j}$ :

$$\sigma(\tilde{F}(e_i)(u_j)) - \sigma(\tilde{G}(e_i)(u_j)) = 0 \Rightarrow \sigma(\tilde{F}(e_i)(u_j)) = \sigma(\tilde{G}(e_i)(u_j))$$

$$h_M^{\tau(k)}(e_i)(u_j) - h_N^{\tau(k)}(e_i)(u_j) = 0 \Rightarrow h_M^{\tau(k)}(e_i)(u_j) = h_N^{\tau(k)}(e_i)(u_j)$$

therefore,  $M \leq N$  and  $N \leq M$ , and according to Definition 4.4, it follows that  $N = M$ .

On the other hand, if  $N = M$ , then  $M \leq N$  and  $N \leq M$  which according to the Definition 4.3 for  $i = 1, \dots, m$ ,  $j = 1, \dots, n$  and  $k = 1, \dots, l_{e_i, u_j}$  it follows that:

$$\sigma(\tilde{F}(e_i)(u_j)) = \sigma(\tilde{G}(e_i)(u_j))$$

$$h_M^{\tau(k)}(e_i)(u_j) = h_N^{\tau(k)}(e_i)(u_j)$$

So clearly  $d_{nhsnm}(M, N)$  will be zero.

• **Property (D3):**

Since the following relationships are established:

$$\left| \sigma(\tilde{F}(e_i)(u_j)) - \sigma(\tilde{G}(e_i)(u_j)) \right|^\lambda = \left| \sigma(\tilde{G}(e_i)(u_j)) - \sigma(\tilde{F}(e_i)(u_j)) \right|^\lambda$$

$$\left| h_M^{\tau(k)}(e_i)(u_j) - h_N^{\tau(k)}(e_i)(u_j) \right|^\lambda = \left| h_N^{\tau(k)}(e_i)(u_j) - h_M^{\tau(k)}(e_i)(u_j) \right|^\lambda$$

Proof of this property is clearly established.

• **Property (D4):**

Suppose that  $M \leq N \leq Z$ . According to Definition 4.3 for  $i = 1, \dots, m$ ,  $j = 1, \dots, n$  and  $k = 1, \dots, l_{e_i, u_j}$  the following conditions are established:

$$1) A \subseteq B \subseteq C$$

$$2) h_M^{\tau(k)}(e_i)(u_j) \leq h_N^{\tau(k)}(e_i)(u_j) \ \& \ \sigma(\tilde{F}(e_i)(u_j)) \geq \sigma(\tilde{G}(e_i)(u_j))$$

$$3) h_N^{\tau(k)}(e_i)(u_j) \leq h_Z^{\tau(k)}(e_i)(u_j) \ \& \ \sigma(\tilde{G}(e_i)(u_j)) \geq \sigma(\tilde{H}(e_i)(u_j))$$

It is first proved that  $d_{nhsnm}(M, N) \leq d_{nhsnm}(M, Z)$ :

As above, it follows that:



$$\begin{aligned} & \left| h_M^{\tau(k)}(e_i)(u_j) - h_N^{\tau(k)}(e_i)(u_j) \right|^\lambda \leq \left| h_M^{\tau(k)}(e_i)(u_j) - h_Z^{\tau(k)}(e_i)(u_j) \right|^\lambda \\ \Rightarrow & \frac{1}{l_{e_i, u_j}} \sum_{k=1}^{l_{e_i, u_j}} \left| h_M^{\tau(k)}(e_i)(u_j) - h_N^{\tau(k)}(e_i)(u_j) \right|^\lambda \leq \frac{1}{l_{e_i, u_j}} \sum_{k=1}^{l_{e_i, u_j}} \left| h_M^{\tau(k)}(e_i)(u_j) - h_Z^{\tau(k)}(e_i)(u_j) \right|^\lambda \end{aligned}$$

Also:

$$\left| \sigma(\tilde{F}(e_i)(u_j)) - \sigma(\tilde{G}(e_i)(u_j)) \right|^\lambda \leq \left| \sigma(\tilde{F}(e_i)(u_j)) - \sigma(\tilde{H}(e_i)(u_j)) \right|^\lambda$$

Then:

$$\begin{aligned} & \left( \frac{1}{m} \sum_{i=1}^m \left\{ \frac{1}{2n} \sum_{j=1}^n \left[ \left| \sigma(\tilde{F}(e_i)(u_j)) - \sigma(\tilde{G}(e_i)(u_j)) \right|^\lambda + \frac{1}{l_{e_i, u_j}} \sum_{k=1}^{l_{e_i, u_j}} \left| h_M^{\tau(k)}(e_i)(u_j) - h_N^{\tau(k)}(e_i)(u_j) \right|^\lambda \right] \right\} \right)^{\frac{1}{\lambda}} \\ & \leq \left( \frac{1}{m} \sum_{i=1}^m \left\{ \frac{1}{2n} \sum_{j=1}^n \left[ \left| \sigma(\tilde{F}(e_i)(u_j)) - \sigma(\tilde{H}(e_i)(u_j)) \right|^\lambda + \frac{1}{l_{e_i, u_j}} \sum_{k=1}^{l_{e_i, u_j}} \left| h_M^{\tau(k)}(e_i)(u_j) - h_Z^{\tau(k)}(e_i)(u_j) \right|^\lambda \right] \right\} \right)^{\frac{1}{\lambda}} \end{aligned}$$

Therefore  $d_{nh snm}(M, N) \leq d_{nh snm}(M, Z)$ . It is proved in the same way that  $d_{nh snm}(N, Z) \leq d_{nh snm}(M, Z)$ . □

The proof is clearly established for the proposed similarity measures.

**Example 4.1.** in Example 3.1, it was shown that with none of the similarity measures of  $s_{h snh}$ ,  $s_{h sne}$  and  $s_{h snm}$ , detection of the more similar set to  $M$  between the sets  $N$  and  $O$  is not possible. However, if the new similarity measures  $s_{nh snh}$ ,  $s_{nh sne}$  and  $s_{nh snm}$  are used, then the following result will be obtained:

$$\begin{aligned} s_{nh snh}(M, N) &= 0.9015, & s_{nh snh}(M, O) &= 0.8198 \\ s_{nh sne}(M, N) &= 0.8672, & s_{nh sne}(M, O) &= 0.8177 \\ s_{nh snm}(M, N) &= 0.8283, & s_{nh snm}(M, O) &= 0.8121, & (\lambda = 6) \\ s_{nh snm}(M, N) &= 0.8179, & s_{nh snm}(M, O) &= 0.8091, & (\lambda = 10) \end{aligned}$$

It can be observed that all the used measures result in the similarity of the set  $N$  to the set  $M$ .

**Example 4.2. (pattern recognition)** Let  $A_1 = (\tilde{I}, E)$ ,  $A_2 = (\tilde{J}, E)$  and  $A_3 = (\tilde{K}, E)$  be three patterns are presented using the HFSSs and are shown in Table 4, Table 5 and Table 6, respectively. A hesitant fuzzy soft pattern  $P = (\tilde{L}, E)$  given in Table 7, it should be detected which of the above patterns it is similar to.

$U$	$e_1$	$e_2$	$e_3$
$u_1$	{1, 0.4}	{0.9, 0.3}	{0.7, 0.4}
$u_2$	{0.7, 0.3}	{0.8, 0.3}	{1, 0.5}

Table 4: The HFSS  $A_1$

$U$	$e_1$	$e_2$	$e_3$
$u_1$	{0.6, 0.4}	{0.4, 0.2}	{0.7, 0.4}
$u_2$	{0.4, 0.2}	{0.4, 0.3}	{0.6, 0.5}

Table 5: The HFSS  $A_2$

Using the similarity measures presented in section 3, sample  $P$  has the lowest similarity to the pattern  $A_3$ . However, using these similarity measures, one cannot distinguish between the similarity of the sample  $P$  to the patterns  $A_1$  or  $A_2$ , since all the similarity measures in section 3 result in the same values for the similarity between sample  $P$  and pattern  $A_1$ , as well as between sample  $P$  and pattern  $A_2$ . The results are presented in Table 8. But, using the Eq. 10 the hesitance index of  $A_1$ ,  $A_2$  and  $P$  are obtained as follows:

$$\sigma(P) = 0.0583, \quad \sigma(A_1) = 0.2417, \quad \sigma(A_2) = 0.0917$$

$U$	$e_1$	$e_2$	$e_3$
$u_1$	{0.1, 0.5}	{0, 0.3}	{0.1, 0.4}
$u_2$	{0.1, 0.3}	{0.7, 0.2}	{0, 0.4}

Table 6: The HFSS  $A_3$ 

$U$	$e_1$	$e_2$	$e_3$
$u_1$	{0.8, 0.7}	{0.6, 0.5}	{0.9, 0.7}
$u_2$	{0.5, 0.4}	{0.6, 0.5}	{0.8, 0.7}

Table 7: The HFSS  $P$ 

Similarity	$s(P, A_1)$	$s(P, A_2)$	$s(P, A_3)$	Result
$s_{hshh}$	0.7833	0.7833	0.6000	$s(P, A_3) < s(P, A_1) = s(P, A_2)$
$s_{hsne}$	0.7764	0.7764	0.5641	$s(P, A_3) < s(P, A_1) = s(P, A_2)$
$s_{hshnm}$ ( $\lambda = 6$ )	0.7534	0.7534	0.4757	$s(P, A_3) < s(P, A_1) = s(P, A_2)$
$s_{hshnm}$ ( $\lambda = 10$ )	0.7377	0.7377	0.4314	$s(P, A_3) < s(P, A_1) = s(P, A_2)$

Table 8: Results obtained using the similarity measures of section 3

It is observed that the hesitance index of  $P$  is closer to hesitance index of  $A_2$ . Therefore the results in Table 8 are not logical.

However, in case of using the similarity measures of  $s_{hshh}$ ,  $s_{hsne}$  and  $s_{hshnm}$ , introduced in section 4, it is concluded that the sample  $P$  has the highest similarity to the pattern  $A_2$ . The results obtained using these similarity measures are presented in Table 9.

Similarity	$s(P, A_1)$	$s(P, A_2)$	$s(P, A_3)$	Result
$s_{hshh}$	0.8000	0.8750	0.7417	$s(P, A_3) < s(P, A_1) < s(P, A_2)$
$s_{hsne}$	0.7898	0.8393	0.6785	$s(P, A_3) < s(P, A_1) < s(P, A_2)$
$s_{hshnm}$ ( $\lambda = 6$ )	0.7658	0.7803	0.5328	$s(P, A_3) < s(P, A_1) < s(P, A_2)$
$s_{hshnm}$ ( $\lambda = 10$ )	0.7501	0.7552	0.4694	$s(P, A_3) < s(P, A_1) < s(P, A_2)$

Table 9: Results obtained using the new similarity measures

## 5 Applications

### 5.1 Hesitant fuzzy soft data clustering

We know that the distance and similarity concepts play an important role in data clustering. Hence, hesitant fuzzy soft data clustering using the single linkage hierarchical clustering algorithm is discussed in this section. In this clustering algorithm, each data is first placed in a cluster, then by calculating the distance between the clusters, in each step, the closest clusters are merged and the process continues to reach one cluster. Suppose that  $H_1, H_2, \dots, H_r$  be the  $r$  HFSSs defined over the universal set  $U$ . The following steps are used to clustering these sets with the single linkage method (the symbol  $C_i$  is used to represent the cluster  $i$ ).

**Step 1)** Consider each of the HFSSs as a cluster:

$$C_1 = \{H_1\}, C_2 = \{H_2\}, \dots, C_r = \{H_r\}$$

**Step 2)** Using a distance measure, form the distance matrix  $D = [d_{ij}]_{r \times r}$ , which  $d_{ij}$  is the distance between the clusters  $C_i$  and  $C_j$  and is calculated using Eq. (17).

$$d_{ij} = \text{distance}(C_i, C_j) = \min \{ \text{distance}(a, b) \mid a \in C_i, b \in C_j \} \quad (17)$$

**Step 3)** Find the two closest clusters (corresponding clusters with the smallest value in the distance matrix) and integrate them together.

**Step 4)** Update the distance matrix after the cluster integration,

**Step 5)** Repeat steps 3 and 4 until reaching one cluster.

To illustrate the clustering process of the hesitant fuzzy soft data, the following example, which includes the clustering of some energy projects, is presented. The idea used to consider the projects in this example is taken from [8, 24, 42]. To show the effect of considering the hesitance index in calculating the distance, once the distance measure  $d_{h_{snh}}$  and once the distance measure  $d_{nh_{snh}}$  are used to calculate the distance between the hesitant fuzzy soft data, then the results are compared together.

**Example 5.1.** The example includes the clustering of the performance of five different energy projects that are evaluated by several evaluators. Assume that  $A_i$  ( $i = 1, \dots, 5$ ) are five energy projects; each energy project consists of two steps  $u_1$  and  $u_2$  and four features including  $e_1$ : technological,  $e_2$ : environmental,  $e_3$ : socio-political and  $e_4$ : economic. Several evaluators were asked to comment on the performance of these projects, and their comments for each project were collected as a HFSS. Some values in these comments may be repeated more than once, which do not indicate their importance relative to the values with fewer repetitions. For example, the value which existed once, may be presented by an evaluator which is elite in the field, and the value repeated more than once may be presented by evaluators which are not sufficiently familiar with that field. The collected comments are shown in Table 10 to Table 14, respectively. These five HFSSs are shown respectively with  $H_1$ ,  $H_2$ ,  $H_3$ ,  $H_4$  and  $H_5$ .

$U$	$e_1$	$e_2$	$e_3$	$e_4$
$u_1$	{0.8, 0.7, 0.6}	{0.8, 0.75}	{0.6, 0.5, 0.4}	{0.8, 0.7}
$u_2$	{0.75, 0.7, 0.6}	{0.7, 0.65, 0.6}	{0.7, 0.5, 0.6}	{0.8, 0.7}

Table 10: The HFSS  $H_1$  for the project  $A_1$

$U$	$e_1$	$e_2$	$e_3$	$e_4$
$u_1$	{0.8, 0.6, 0.5}	{0.8, 0.65, 0.25}	{0.6, 0.1, 0.2}	{0.8, 0.5}
$u_2$	{0.75, 0.3, 0.4}	{0.7, 0.25, 0.2}	{0.9, 0.5, 0.6}	{0.8, 0.3}

Table 11: The HFSS  $H_2$  for the project  $A_2$

$U$	$e_1$	$e_2$	$e_3$	$e_4$
$u_1$	{0.8, 0.75, 0.7}	{0.7, 0.6}	{0.5, 0.45, 0.4}	{0.8, 0.7}
$u_2$	{0.65, 0.6}	{0.7, 0.55, 0.5}	{0.6, 0.5, 0.4}	{0.8, 0.7}

Table 12: The HFSS  $H_3$  for the project  $A_3$

$U$	$e_1$	$e_2$	$e_3$	$e_4$
$u_1$	{0.7, 0.6}	{0.8, 0.7, 0.5}	{0.4, 0.3}	{0.7, 0.6}
$u_2$	{0.6, 0.5}	{0.5, 0.45, 0.4}	{0.8, 0.7}	{0.8, 0.5}

Table 13: The HFSS  $H_4$  for the project  $A_4$

$U$	$e_1$	$e_2$	$e_3$	$e_4$
$u_1$	{0.8, 0.5, 0.45}	{0.9, 0.6, 0.4}	{0.5, 0.2, 0.15}	{0.8, 0.5}
$u_2$	{0.7, 0.55}	{0.7, 0.35, 0.3}	{1.0, 0.5, 0.4}	{0.8, 0.3}

Table 14: The HFSS  $H_5$  for the project  $A_5$ **Clustering using the distance measure  $d_{hsnh}$ :**

Step 1) Each of the HFSSs is considered as a cluster. Hence, the following clusters will be achieved in this step:

$$C_1 = \{H_1\}, C_2 = \{H_2\}, C_3 = \{H_3\}, C_4 = \{H_4\}, C_5 = \{H_5\}$$

Step 2) Using the distance measure  $d_{hsnh}$ , the distance matrix  $D$  is formed as follows:

$$D = \begin{bmatrix} 0.0000 & 0.1625 & 0.0573 & 0.1271 & 0.1604 \\ 0.1625 & 0.0000 & 0.1708 & 0.1271 & 0.0646 \\ 0.0573 & 0.1708 & 0.0000 & 0.1219 & 0.1354 \\ 0.1271 & 0.1271 & 0.1219 & 0.0000 & 0.1219 \\ 0.1604 & 0.0646 & 0.1354 & 0.1219 & 0.0000 \end{bmatrix}$$

Step 3) It can be seen that the smallest distance in the matrix  $D$  is 0.0573, which corresponds to clusters  $C_1$  and  $C_3$ . So in this step, the clusters  $C_1$  and  $C_3$  are merged and the following clustering will be obtained:

$$C_1 = \{H_1, H_3\}, C_2 = \{H_2\}, C_3 = \{H_4\}, C_4 = \{H_5\}$$

Step 4) the distance matrix is updated for the clusters obtained in step 3:

$$D = \begin{bmatrix} 0.0000 & 0.1625 & 0.1219 & 0.1354 \\ 0.1625 & 0.0000 & 0.1271 & 0.0646 \\ 0.1219 & 0.1271 & 0.0000 & 0.1219 \\ 0.1354 & 0.0646 & 0.1219 & 0.0000 \end{bmatrix}$$

Since one cluster has not yet been obtained, steps 3 and 4 are repeated. It is observed that this time, the smallest distance is 0.0646, which corresponds to clusters  $C_2$  and  $C_4$ . Therefore, the obtained clustering is:

$$C_1 = \{H_1, H_3\}, C_2 = \{H_2, H_5\}, C_3 = \{H_4\}$$

Therefore, the distance matrix is obtained as follows:

$$D = \begin{bmatrix} 0.0000 & 0.1354 & 0.1219 \\ 0.1354 & 0.0000 & 0.1219 \\ 0.1219 & 0.1219 & 0.0000 \end{bmatrix}$$

Since one cluster has not yet been obtained, steps 3 and 4 are repeated. It is observed that this time, the smallest distance is 0.1219, which corresponds to clusters  $C_1$  and  $C_3$ , as well as the clusters  $C_2$  and  $C_3$ . Therefore, it is not possible to make a definite decision to merge clusters in this step, because each of the following two clusters is possible:

$$C_1 = \{H_1, H_3\}, C_2 = \{H_2, H_4, H_5\}$$

$$C_1 = \{H_1, H_3, H_4\}, C_2 = \{H_2, H_5\}$$

By performing each of the above clusters, the distance matrix will be updated as follows:

$$D = \begin{bmatrix} 0.0000 & 0.1219 \\ 0.1219 & 0.0000 \end{bmatrix}$$

Therefore, clusters  $C_1$  and  $C_2$  merge together and the algorithm terminates after reaching one cluster. Hence, the final cluster is as follows:

$$C = \{H_1, H_2, H_3, H_4, H_5\}$$

As can be seen, if the distance measure  $d_{hsnh}$  is used in the step of forming two clusters, the algorithm faces a problem and a cluster must be selected randomly. This problem also exists in clustering in case of using the distances  $d_{hsne}$  and  $d_{hsnm}$ .

**Clustering using the distance measure  $d_{nhsnh}$ :**

Step 1) Each of the HFSSs is considered as a cluster. Hence, the following clusters will be achieved in this step:

$$C_1 = \{H_1\}, C_2 = \{H_2\}, C_3 = \{H_3\}, C_4 = \{H_4\}, C_5 = \{H_5\}$$

Step 2) Using the distance measure  $d_{nhsnh}$  the distance matrix D is formed as follows:

$$D = \begin{bmatrix} 0.0000 & 0.1492 & 0.0404 & 0.0827 & 0.1400 \\ 0.1492 & 0.0000 & 0.1565 & 0.1257 & 0.0557 \\ 0.0404 & 0.1565 & 0.0000 & 0.0793 & 0.1306 \\ 0.0827 & 0.1257 & 0.0793 & 0.0000 & 0.1150 \\ 0.1400 & 0.0557 & 0.1306 & 0.1150 & 0.0000 \end{bmatrix}$$

Step 3) It can be seen that the smallest distance in the matrix D is 0.0404, which corresponds to clusters  $C_1$  and  $C_3$ . So in this step, the clusters  $C_1$  and  $C_3$  are merged and the following clustering will be obtained:

$$C_1 = \{H_1, H_3\}, C_2 = \{H_2\}, C_3 = \{H_4\}, C_4 = \{H_5\}$$

Step 4) the distance matrix is updated for the clusters obtained in step 3:

$$D = \begin{bmatrix} 0.0000 & 0.1492 & 0.0793 & 0.1306 \\ 0.1492 & 0.0000 & 0.1257 & 0.0557 \\ 0.0793 & 0.1257 & 0.0000 & 0.1150 \\ 0.1306 & 0.0557 & 0.1150 & 0.0000 \end{bmatrix}$$

Since one cluster has not yet been obtained, steps 3 and 4 are repeated again. It is observed that this time, the smallest distance is 0.0557, which corresponds to clusters  $C_2$  and  $C_4$ . Therefore, the obtained clustering is:

$$C_1 = \{H_1, H_3\}, C_2 = \{H_2, H_5\}, C_3 = \{H_4\}$$

Therefore, the distance matrix is obtained as follows:

$$D = \begin{bmatrix} 0.0000 & 0.1306 & 0.0793 \\ 0.1306 & 0.0000 & 0.1150 \\ 0.0793 & 0.1150 & 0.0000 \end{bmatrix}$$

Since one cluster has not yet been obtained, steps 3 and 4 are repeated again. It is observed that this time, the smallest distance is 0.0793, which corresponds to clusters  $C_1$  and  $C_3$ . Therefore, the obtained clustering is:

$$C_1 = \{H_1, H_3, H_4\}, C_2 = \{H_2, H_5\}$$

Therefore, the distance matrix is obtained as follows:

$$D = \begin{bmatrix} 0.0000 & 0.1150 \\ 0.1150 & 0.0000 \end{bmatrix}$$

Therefore, clusters  $C_1$  and  $C_2$  merge together and the algorithm terminates after reaching one cluster. Hence, the final cluster is as follows:

$$C = \{H_1, H_2, H_3, H_4, H_5\}$$

The dendrograms of the clustering obtained using distance measures  $d_{hsnh}$  and  $d_{nhsnh}$  is shown in Fig. 1(a) and Fig. 1(b), respectively.

**Comparison of the results:**

It can be seen that using the new distance measure  $d_{nhsnh}$ , the problem existing when using the distance measures  $d_{hsnh}$ ,  $d_{hsne}$ , and  $d_{hshm}$  will be resolved. Since the proposed distance measures consider the hesitations of sets as well, and since the hesitation in set  $H_4$  is closer to the hesitation of the sets  $H_1$  and  $H_3$ , it seems logical that, considering

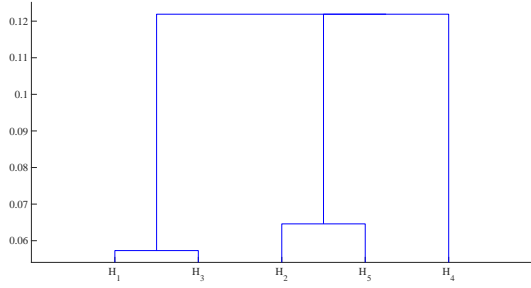
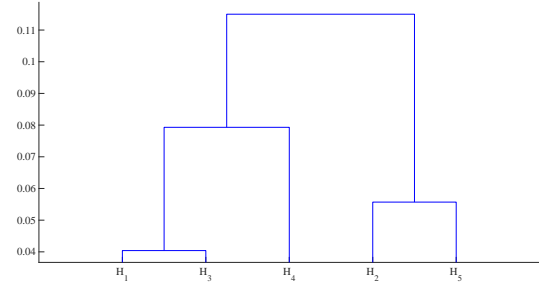
(a) The dendrogram obtained using  $d_{hsnh}$  distance measure(b) The dendrogram obtained using  $d_{nhsnh}$  distance measure

Figure 1: The clustering dendrograms

two clusters, these three sets to lay in one cluster. In addition, in case of using the measures  $d_{nhsne}$  and  $d_{nhsnm}$ , the clustering result will be similar to the result obtained using the  $d_{nhsnh}$  measure. Table 15 shows the obtained results.

Number of clusters	Clustering using $d_{hsnh}$ , $d_{hsne}$ and $d_{hsnm}$	Clustering using $d_{nhsnh}$ , $d_{nhsne}$ and $d_{nhsnm}$
5	$\{H_1\}, \{H_2\}, \{H_3\}, \{H_4\}, \{H_5\}$	$\{H_1\}, \{H_2\}, \{H_3\}, \{H_4\}, \{H_5\}$
4	$\{H_1, H_3\}, \{H_2\}, \{H_4\}, \{H_5\}$	$\{H_1, H_3\}, \{H_2\}, \{H_4\}, \{H_5\}$
3	$\{H_1, H_3\}, \{H_2, H_5\}, \{H_4\}$	$\{H_1, H_3\}, \{H_2, H_5\}, \{H_4\}$
2	$\{H_1, H_3\}, \{H_2, H_4, H_5\}$ or $\{H_1, H_3, H_4\}, \{H_2, H_5\}$	$\{H_1, H_3, H_4\}, \{H_2, H_5\}$
1	$\{H_1, H_2, H_3, H_4, H_5\}$	$\{H_1, H_2, H_3, H_4, H_5\}$

Table 15: Clustering results

## 5.2 Application of the proposed hesitance index in decision-making

In [5] the application of HFSS is investigated in group decision-making problems and an algorithm is also proposed for solving these problems using the HFSS. In [2], this algorithm is also used for handling multicriteria group decision-making under hesitant fuzzy soft numbers based on hesitant fuzzy soft set. Assuming that  $U = \{u_1, u_2, \dots, u_n\}$  be a universal set including the alternatives and  $E = \{e_1, e_2, \dots, e_m\}$  be a set of criteria and  $A \subseteq E$ , this algorithm is presented in [2] in the form of Algorithm 1.

### Algorithm 1. [2]

- 1) Input the hesitant fuzzy soft set  $(\tilde{F}, A)$ .
- 2) Compute the induced fuzzy soft set  $\Delta_{\tilde{F}} = (\tilde{G}, A)$ .
- 3) Calculate the average of  $\tilde{G}(e_i)$  for each  $u_j$  and let it denoted as  $a_j$ , this is the decision table.
- 4) Select the optimal alternative  $u_k$  if  $a_k = \max_j a_j$ .
- 5) If there are more than one  $u_k$  then any one of  $u_k$  may be chosen.

In step 2 of algorithm 1, each entry of the  $(\tilde{G}, A)$  is the score of the corresponding element in the  $(\tilde{F}, A)$ . But, the use of this method to calculate the induced fuzzy soft set may cause undesirable result for the decision-making problem. For example, the score of the two elements  $\{0.1, 0.9\}$  and  $\{0.45, 0.55\}$  is equal to 0.5, but with respect to the hesitance degree in the decision makers' comments for the element  $\{0.1, 0.9\}$ , Equivalence of the membership degrees of the two elements in the induced fuzzy soft set does not seem logical. Therefore, it is suggested that the Eq. (18) be used to calculate each entry of the induced fuzzy soft set.

$$\tilde{G}(e_i)(u_j) = \frac{s(h(e_i)(u_j))}{1 + \sigma(h(e_i)(u_j))} \quad (18)$$

which,  $\tilde{G}(e_i)(u_j)$  is the membership degree of  $j$ st member of universal set in fuzzy set  $\tilde{G}(e_i)$ .  $\sigma(h(e_i)(u_j))$  is the hesitance index of the element  $h(e_i)(u_j)$  in hesitant fuzzy soft set  $(\tilde{F}, A)$  and  $s(h(e_i)(u_j))$  is the score of this element.

**Example 5.2.** Suppose that five persons are applying for a job in a company. Each of these persons is considered as a member of universal set  $U = \{u_1, u_2, u_3, u_4, u_5\}$ . two decision makers have evaluated the candidates by considering a set of criteria  $E = \{e_1, e_2, e_3, e_4, e_5\}$ , which are “responsibility”, “appropriate knowledge”, “reliability”, “experience” and “creativity”, respectively. The comments of decision makers are gathered in the form of hesitant fuzzy soft set  $(\tilde{F}, E)$  shown in Table 16. The problem is choosing the best person for the desired job.

Using the algorithm 1, and without considering the Eq. (18), the induced fuzzy soft set and the decision table are shown in Table 17 and Table 18, respectively. As shown in Table 18, the  $a_2 = a_4$ , therefore, one of the candidates 2 or 4 should be selected randomly, but, the comments of decision makers about candidate 4 are more hesitancy and it is not logical to consider these two candidates equally. However, if the Eq. (18) is used to compute  $(\tilde{G}, A)$ , then the new induced fuzzy soft set and the new decision table will be obtained as Table 19 and Table 20, respectively. According to Table 20, it is clear that candidate 2 is selected for this job.

$U$	$e_1$	$e_2$	$e_3$	$e_4$	$e_5$
$u_1$	{0.45, 0.5}	{0.6}	{0.3, 0.4}	{0.5}	{0.2, 0.35}
$u_2$	{0.7, 0.8}	{0.5, 0.7}	{0.5, 0.55}	{0.6}	{0.6, 0.8}
$u_3$	{0.3, 0.45}	{0.2}	{0.4, 0.5}	{0.25}	{0.3, 0.4}
$u_4$	{0.5, 1}	{0.3, 0.75}	{0.4, 0.8}	{0.5, 0.9}	{0.6}
$u_5$	{0.5, 0.6}	{0.3}	{0.4, 0.55}	{0.7}	{0.2, 0.35}

Table 16: The comments of decision makers

$U$	$e_1$	$e_2$	$e_3$	$e_4$	$e_5$
$u_1$	0.475	0.6	0.35	0.5	0.275
$u_2$	0.75	0.6	0.525	0.6	0.7
$u_3$	0.375	0.2	0.45	0.25	0.35
$u_4$	0.75	0.525	0.6	0.7	0.6
$u_5$	0.55	0.3	0.475	0.7	0.275

Table 17: The induced fuzzy soft set

$a_j$	Values
$a_1$	0.44
$a_2$	0.635
$a_3$	0.325
$a_4$	0.635
$a_5$	0.46

Table 18: The decision table

$U$	$e_1$	$e_2$	$e_3$	$e_4$	$e_5$
$u_1$	0.463	0.6	0.333	0.5	0.256
$u_2$	0.714	0.545	0.512	0.6	0.636
$u_3$	0.349	0.2	0.429	0.25	0.333
$u_4$	0.6	0.429	0.5	0.583	0.6
$u_5$	0.524	0.3	0.442	0.7	0.256

Table 19: The new induced fuzzy soft

$a_j$	Values
$a_1$	0.4304
$a_2$	0.6014
$a_3$	0.3122
$a_4$	0.5424
$a_5$	0.4444

Table 20: The new decision table

## 6 Conclusion

In the present study new distance and similarity measures were proposed for HFSSs by evaluating the standard deviation as a hesitance index for these sets. Based on the experiments in the areas of pattern recognition and clustering, the new measures could yield more logical results, compared to the distance and similarity measures which disregarded the concept of standard deviation. For example, the correct clustering may not be detected in some steps while using the distance and similarity measures without a hesitance index in clustering the hesitant fuzzy soft data. However, a more logical clustering can be obtained for these data when implementing the proposed measures. In addition, the proposed hesitance index was utilized to improve a decision-making algorithm and the improved algorithm provided

a more logical result based on a decision-making example. Future studies can be performed by using the proposed distance and similarity measures in more complex cases of decision-making, pattern recognition, clustering, and the like. Finally, the proposed measures can be extended to other hybridizations in soft sets.

## Acknowledgement

The authors wish to express their appreciation for several excellent suggestions for improvements in this paper made by the referees.

## References

- [1] L. Aliahmadipour, E. Eslami, *GHFHC: Generalized Hesitant Fuzzy Hierarchical Clustering Algorithm*, International Journal of Intelligent Systems, **31** (9) (2016), 855-871.
- [2] H. A. Alshehri, H. A. Abujabal, N. O. Alshehri, *New types of hesitant fuzzy soft ideals in BCK-algebras*, Soft Computing, **22**(11) (2018), 3675-3683.
- [3] M. Arefi, S. M. Taheri, *Weighted similarity measure on interval-valued fuzzy sets and its application to pattern recognition*, Iranian Journal of Fuzzy Systems, **11**(5) (2014), 67-79.
- [4] K. T. Atanassov, *Intuitionistic fuzzy sets*, Fuzzy Sets and Systems, **20**(1) (1986), 87-96.
- [5] K. V. Babitha, S. J. John, *Hesitant fuzzy soft sets*, Journal of New Results in Science, **2**(3) (2013), 98-107.
- [6] V. Balopoulos, A. G. Hatzimichailidis, B. K. Papadopoulos, *Distance and similarity measures for fuzzy operators*, Information Sciences, **177**(11) (2007), 2336-2348.
- [7] I. Beg, T. Rashid, *Ideal solutions for hesitant fuzzy soft sets*, Journal of Intelligent & Fuzzy Systems, **30** (1) (2016), 143-150.
- [8] C. Bin, *Generalized hesitant fuzzy soft sets*, Italian journal of pure and applied mathematics, **36** (2016), 35-54.
- [9] F. E. Boran, D. Akay, *A biparametric similarity measure on intuitionistic fuzzy sets with applications to pattern recognition*, Information Sciences, **255** (2014), 45-57.
- [10] N. Chen, Z. Xu, M. Xia, *Correlation coefficients of hesitant fuzzy sets and their applications to clustering analysis*, Applied Mathematical Modelling, **37**(4) (2013), 2197-2211.
- [11] N. Chen, Z. Xu, M. Xia, *The ELECTRE I Multi-Criteria Decision-Making Method Based on Hesitant Fuzzy Sets*, International Journal of Information Technology & Decision Making, **14**(3) (2015), 621-657.
- [12] S. Das, S. Kar, *The Hesitant Fuzzy Soft Set and Its Application in Decision-Making*, Facets of Uncertainties and Applications, Springer Proceedings in Mathematics & Statistics, (2015), 235-247.
- [13] S. Das, D. Malakar, S. Kar, T. Pal, *Correlation measure of hesitant fuzzy soft sets and their application in decision making*, Neural Computing and Applications, (2017) 1-17.
- [14] J. Y. Dong, F. F. Yuan, S. P. Wan, *Extended VIKOR method for multiple criteria decision-making with linguistic hesitant fuzzy information*, Computers & Industrial Engineering, **112** (2017), 305-319.
- [15] D. Dubois, H. Prade, *Fuzzy Sets and Systems: Theory and Applications*, Academic Press, New York, 1980.
- [16] B. Farhadinia, *Information measures for hesitant fuzzy sets and interval-valued hesitant fuzzy sets*, Information Sciences, **240** (2013), 129-144.
- [17] F. Feng, X. Liu, V. Leoreanu-Fotea, Y. B. Jun, *Soft sets and soft rough sets*, Information Sciences, **181**(6) (2011), 1125-1137.
- [18] P. Grzegorzewski, *Distances between intuitionistic fuzzy sets and/or interval-valued fuzzy sets based on the Hausdorff metric*, Fuzzy Sets and Systems, **148**(2) (2004), 319-328.



- [19] X. Guan, G. Sun, X. Yi, J. Zhao, *Grey Relational Analysis for Hesitant Fuzzy Sets and Its Applications to Multiattribute Decision-Making*, Mathematical Problems in Engineering, **2018**, 1-11.
- [20] X. Guan, G. Sun, X. Yi, Z. Zhou, *Synthetic correlation coefficient between hesitant fuzzy sets with applications*, International Journal of Fuzzy Systems, **20**(6) (2018), 1968-1985.
- [21] W. L. Hung, M. S. Yang, *Similarity Measures between Type-2 Fuzzy Sets*, International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, **12**(6) (2004), 827-841.
- [22] Y. Jiang, Y. Tang, Q. Chen, H. Liu, J. Tang, *Interval-valued intuitionistic fuzzy soft sets and their properties*, Computers & Mathematics with Applications, **60**(3) (2010), 906-918.
- [23] J. Kacprzyk, *Multistage Fuzzy Control: A Model-Based Approach to Fuzzy Control and Decision Making*, Wiley, Chichester 1997.
- [24] C. Kahraman, İ. Kaya, *A fuzzy multicriteria methodology for selection among energy alternatives*, Expert Systems with Applications, **37**(9) (2010), 6270-6281.
- [25] P. K. Maji, R. Biswas, A. R. Roy, *Fuzzy soft sets*, Journal of Fuzzy Mathematics, **9** (2001), 589-602.
- [26] P. K. Maji, R. Biswas, A. R. Roy, *Intuitionistic Fuzzy Soft Sets*, Journal of Fuzzy Mathematics, **9** (2001), 677-692.
- [27] P. Majumdar, S. K. Samanta, *Similarity Measure of Soft Sets*, New Mathematics and Natural Computation, **4**(1) (2008), 1-12.
- [28] D. Molodtsov, *Soft set theory-First results*, Computers & Mathematics with Applications, **37**(4) (1999), 19-31.
- [29] Z. Pawlak, *Rough sets*, International Journal of Computer & Information Sciences, **11**(5) (1982), 341-356.
- [30] G. Qian, H. Wang, X. Feng, *Generalized hesitant fuzzy sets and their application in decision support system*, Knowledge-Based Systems, **37** (2013), 357-365.
- [31] P. H. A. Sneath, R. R. Sokal *Numerical taxonomy. The principles and practice of numerical classification*, Freeman, San Francisco, 1973.
- [32] G. Sun, X. Guan, X. Yi, Z. Zhou, *An innovative TOPSIS approach based on hesitant fuzzy correlation coefficient and its applications*, Applied Soft Computing, **68** (2018), 249-267.
- [33] G. Sun, X. Guan, X. Yi, Z. Zhou, *Grey relational analysis between hesitant fuzzy sets with applications to pattern recognition*, Expert Systems with Applications, **92** (2018), 521-532.
- [34] G. Sun, X. Guan, X. Yi, Z. Zhou, *Multi-Attribute Decision Making with Interval-Valued Hesitant Fuzzy Information, a Novel Synthetic Grey Relational Degree Method*, INFORMATICA, **29**(3) (2018), 517-537.
- [35] V. Torra, *Hesitant fuzzy sets*, International Journal of Intelligent Systems, **25**(6) (2010), 529-539.
- [36] V. Torra, Y. Narukawa, *On hesitant fuzzy sets and decision*, 2009 IEEE International Conference on Fuzzy Systems, (2009), 1378-1382.
- [37] S. P. Wan, Y. L. Qin, J. Y. Dong, *A hesitant fuzzy mathematical programming method for hybrid multi-criteria group decision making with hesitant fuzzy truth degrees*, Knowledge-Based Systems, **138** (2017), 232-248.
- [38] F. Wang, X. Li, X. Chen, *Hesitant Fuzzy Soft Set and Its Applications in Multicriteria Decision Making*, Journal of Applied Mathematics **2014** (2014), 1-10.
- [39] W. J. Wang, *New similarity measures on fuzzy sets and on elements*, Fuzzy Sets and Systems, **85**(3) (1997), 305-309.
- [40] M. Xia, Z. Xu, *Hesitant fuzzy information aggregation in decision making*, International Journal of Approximate Reasoning, **52**(3) (2011), 395-407.
- [41] G. L. Xu, S. P. Wan, J. Y. Dong, *A hesitant fuzzy programming method for hybrid MADM with incomplete attribute weight information*, Informatica, **27** (4) (2016), 863-892.

- [42] Z. Xu, M. Xia, *Distance and similarity measures for hesitant fuzzy sets*, Information Sciences, **181**(11) (2011), 2128-2138.
- [43] X. Yang, T. Y. Lin, J. Yang, Y. Li, D. Yu, *Combination of interval-valued fuzzy set and soft set*, Computers & Mathematics with Applications, **58**(3) (2009), 521-527.
- [44] L. A. Zadeh, *Fuzzy sets*, Information and Control, **8**(3) (1965), 338-353.
- [45] L. A. Zadeh, *The concept of a linguistic variable and its application to approximate reasoning-I*, Information Sciences, **8**(3) (1975), 199-249.
- [46] W. Zeng, D. Li, Q. Yin, *Distance and similarity measures between hesitant fuzzy sets and their application in pattern recognition*, Pattern Recognition Letters, **84** (2016), 267-271.
- [47] X. Zeshui, *Intuitionistic fuzzy hierarchical clustering algorithms*, Journal of Systems Engineering and Electronics, **20**(1) (2009), 90-97.
- [48] X. Zhang, Z. Xu, *Novel distance and similarity measures on hesitant fuzzy sets with applications to clustering analysis*, Journal of Intelligent & Fuzzy Systems, **28**(5) (2015), 2279-2296.