¹All India Shri Shivaji Memorial Society's Institute of Information Technology (AISSMS IOIT), Savitribai Phule Pune University, Pune, Maharashtra, India
²All India Shri Shivaji Memorial Society's College of Engineering (AISSM-SCOE), Savitribai Phule Pune University, Pune, Maharashtra, India
³Bharati Vidyapeeth College of Engineering (BVCOE), Pune

A Framework for Optimal Attribute Evaluation and Selection in Hesitant Fuzzy Environment Based on Enhanced Ordered Weighted Entropy Approach for Medical Dataset

Dikshit-Ratnaparkhi A.[1]*, Bormane D.[2], Ghongade R.[3]

ABSTRACT

Background: In this paper, a generic hesitant fuzzy set (HFS) model for clustering various ECG beats according to weights of attributes is proposed. A comprehensive review of the electrocardiogram signal classification and segmentation methodologies indicates that algorithms which are able to effectively handle the nonstationary and uncertainty of the signals should be used for ECG analysis. Extensive research that focuses on incorporating vagueness in the form of fuzzy sets, fuzzy rough sets and hesitant fuzzy sets (HFS) has been in past decades.

Objective: The paper aims to develop an enhanced entropy based on the clustering technique for calculating the weights of the attributes to finally generate appropriately clustered attributes.

Material and Methods: Finding optimal attributes to make a decision has always been a matter of concern for the researchers. Metrics used for optimal attribute generation can be broadly classified into mutual dependency, similarity, correlation and entropy based metrics in fuzzy domain .The experimentation has been carried out on ECG dataset in a hesitant fuzzy framework with four attributes.

Results: We propose a novel correlation based on an algorithm that takes entropy based weighted attributes as input which effectively generates a relevant and non-redundant set of attributes. We have also derived correlation coefficient formulas for HFSs and applied them to clustering analysis under framework of hesitant fuzzy sets. The results show the comparison of the proposed mathematical model with the existing similarity based on algorithms.

Conclusion: The selection of optimal relevant attributes certainly highlights the robustness and efficacy of the proposed approach. The entire experimentation and comparative results help us conclude that selection of optimal attributes in hesitant fuzzy domain certainly prove to be a powerful tool in order to express uncertainty in the process of data acquisition and classification.

Introduction

Electrocardiogram classification and segmentation have been subjects of research for a prolonged period since the enormous deaths occur worldwide due to uncertainty and variability in the diagnosis of heart diseases. Health care systems have undergone revolutionary changes to provide timely and relevant information about the patients'

*Corresponding author: A. Dikshit-Ratnaparkhi All India Shri Shivaji Memorial Society's Institute of Information Technology (AISSMS IOIT), Savitribai Phule Pune University, Pune, Maharashtra, India E-mail: archana.ratnaparkhi@gmail.com

health periodically, thereby providing a cost-effective way for prediction of heart condition. The signal acquired from 12 lead data in real time ECG acquisition systems is contaminated by artifacts due to power line noise, respiration, muscular movements and electrode impedance. This leads to changing in the electrical characteristics of ECG which are seen as baseline drifts and motion artifacts. Generally, amplitude and duration of major features in the ECG segmentation process which have undergone considerable changes. Hence fuzzy domain and framework of hesitant fuzzy sets are the solutions which are able to effectively deal with such variations and uncertainty.

Introduction and comprehensive mathematical modelling for hesitant fuzzy sets and intuitionistic fuzzy sets have been carried out as an extension to Zadeh's fuzzy set theory involving the creation of hesitant sets allowing the membership degree of an element of a set to have several possible values and be able to explain the hesitant information more comprehensively than any other fuzzy sets [1]. Multicriteria decision in accordance with intuitionistic fuzzy sets using optimal weights for the attributes is made [2-3]. Clustering methods for optimal criteria selection have been discussed at length in [4-6]. Significance of hesitant fuzzy sets (HFS) on the basis of the similarities between the fuzzy sets and HFS has been assessed [1, 5]. However, it presents limitations when various sources of vagueness appear simultaneously. Comparison of type 1 and type 2 of fuzzy sets with hesitant fuzzy sets so as to provide a clear perspective on the different concepts, tools and trends related to the extension of fuzzy sets is provided [7]. ECG segmentation and k nearest neighbour based on classification approaches for dataset reduction have been discussed [8-10]. Type 1 of fuzzy sets used widely and applied successfully in different data to handle uncertainties is discussed [11]. Rough sets theory developed by Pawlak deals with crisp upper and lower sets to handle imprecision, vagueness and un-

certainty in the data analysis [12]. This theory was further developed in fuzzy framework to incorporate the imprecise information to improve the accuracy in classification problems [13-15].The problems associated with the formation of indiscernibility equations or partitions as far as rough sets concerned are addressed. Fusion of rough and fuzzy sets are explored through axiomatic and constructive approach [16].With reference to the practical problems associated with ECG signal acquisition, hesitant fuzzy sets theory has been developed while taking into account the fact that patients might exhibit multiple symptoms at the same time. With reference to ECG analysis, this could include variation in ST segment, QRS duration or any other attributes which could create a complex decision system. Such uncertainty can be resolved with hesitant fuzzy sets. More recently, HFS can be effectively if it is applied to medical data diagnosis [17, 18]. Thus the aforementioned literature applying the concepts of hesitant fuzzy sets certainly prove ECG analysis. In the next section, preliminaries of hesitant fuzzy sets have been discussed.

PRELIMINARIES

The concept of fuzziness is further extended in the hesitant fuzzy sets allowing the membership degree of an element to a set to have several possible values [19, 20]. We propose a hesitant fuzzy set model to attribute reduction of ECG dataset. In real world scenario, it is possible that a patient's medical report shows a variation in number of parameters causing complications in decision made for a physician. In such cases, a physician might assign variable probable membership values to the attributes or the classes.

a. Properties of Hesitant Fuzzy sets (HFS)

Definition 1. Let X be a reference set, a hesitant fuzzy set (HFS) A on X is defined in terms of function $h_A(x)$ that after applying to X returns a subset A which is given by $A = \{\langle x, h_A(x) \rangle \mid x \in X\}$ Where $h_A(x)$ is a set

of possible membership degrees of the element $x \in X$ to A. Let $h_A(x)$ be called the hesitant fuzzy element (HFS) [1-2, 21].

Definition 2. Given a HFS (h), its lower and upper bounds are defined respectively as below

$$h^-(x) = \min h(x)$$

$$h^+(x) = \max h(x)$$

Definition 3: For a HFS, the score function $s(h_A(x))$ is defined as follows:

Where $s(h_A(x)) \in [(0,1)]$

$$s(h_A(x)) = \frac{\sum_{i=1}^{l(h_A(x))} h_A^{\sigma(j)}(x)}{l(h_A(x))}$$

Definition 4: Correlation coefficients of HFS Let $x = \{x_1, x_2, x_3 \ldots \ldots x_n\}$ be universe of discourse in discrete form, and A and B be two HFS on X denoted by $A = \{\langle x, h_A(x) \rangle \mid x \in X\}$ and $B = \{\langle x, h_B(x) \rangle \mid x \in X\}$; the informational energy of the set A is defined as

$$E_{HFS}(A) = \sum_{i=1}^{n} \frac{1}{l_i} \left(\sum_{j=1}^{l_i} h_{A\sigma(j)}^2(x_i) \right)$$

For two HFS A and B, their correlation is defined by

$$C_{HFS}(A,B) = \sum_{i=1}^{n} \frac{1}{l_i} \left(\sum_{j=1}^{l_i} h'_{A\sigma(j)}(x_i) * h'_{B\sigma(j)}(x_i) \right)$$

The above equation satisfies the following properties

1) $C_{HFS}(A,B) = C_{HFS}(B,A)$
2) $C_{HFS}(A,A) = E_{HFS}(A)$

Then the correlation coefficient between two HFS A and B given by [21], and it can be followed as:

$$\rho_{HFS}(A,B) = \frac{C_{HFS}(A,B)}{[C_{HFS}(A,A)]^{1/2}[C_{HFS}(B,B)]^{1/2}}$$

$$= \frac{\sum_{i=1}^{n} \frac{1}{l_i} \left(\sum_{j=1}^{l_i} h'_{A\sigma(j)}(x_i) h'_{B\sigma(j)}(x_i) \right)}{\left[\sum_{i=1}^{n} \frac{1}{l_i} \left(\sum_{j=1}^{l_i} h'_{A\sigma(j)}(x_i) h'_{A\sigma(j)}(x_i) \right) \right]^{1/2} \left[\sum_{i=1}^{n} \frac{1}{l_i} \left(\sum_{j=1}^{l_i} h'_{B\sigma(j)}(x_i) h'_{B\sigma(j)}(x_i) \right) \right]^{1/2}}$$

If the length of $h_A(x_i)$ is less than $h_B(x_i)$, then $h_A(x_i)$ should be extended by adding the minimum value in it until having the length

same as $h_B(x_i)$. This concept has also been applied to hesitant fuzzy sets [1-2]. The Correlation coefficient has been further modified by [21] as follows

$$\rho_{HFS}(A,B) = \frac{C_{HFS}(A,B)}{[C_{HFS}(A,A)]^{1/2}[C_{HFS}(B,B)]^{1/2}}$$

$$= \frac{\sum_{i=1}^{n} w_i \left(\sum_{j=1}^{l_i} h'_{A\sigma(j)}(x_i) h'_{B\sigma(j)}(x_i) \right)}{\left[\sum_{i=1}^{n} \frac{1}{l_i} \left(\sum_{j=1}^{l_i} h'_{A\sigma(j)}(x_i) h'_{A\sigma(j)}(x_i) \right) \right]^{1/2} \left[\sum_{i=1}^{n} \frac{1}{l_i} \left(\sum_{j=1}^{l_i} h'_{B\sigma(j)}(x_i) h'_{B\sigma(j)}(x_i) \right) \right]^{1/2}}$$

Where w_i is the weight vector.

Definition 5: The novel correlation measure is an enhanced version of the correlation measure put forth [17, 21]. A systematic mathematical approach for calculation of attribute weights is provided in the following section. The attribute weights are calculated on the basis of the entropy of the attributes rather than assuming random values [21]. Two entropy based ordered weighted attribute selection (EBOW-AS) approaches for hesitant fuzzy sets have been provided to calculate correlation measure.

Method 1: For two hesitant fuzzy sets, A and B on $X = \{x_1, x_2 \ldots \ldots x_n\}$ have weights associated with them and given by weight vector $w = \{w_1, w_2 \ldots \ldots w_n\}$ with $w_i \geq 1$ and $\sum_{i=1}^{n} w_i = 1$ then the proposed correlation based EBOW-AS for HFS can be given as

$$\delta_m(A,B) = \frac{C_{HFS}(A,B)}{[C_{HFS}(A,A)]^{1/2}[C_{HFS}(B,B)]^{1/2}}$$

$$= \frac{\left[\sum_{i=1}^{n} w_{mi} \left(\frac{1}{l_i} \left(\sum_{j=1}^{l_i} h'_{A\sigma(j)}(x_i) * h'_{B\sigma(j)}(x_i) \right) \right) \right]^{1/2}}{\left[\sum_{i=1}^{n} w_{mi} \left(\frac{1}{l_i} \left(\sum_{j=1}^{l_i} h'_{A\sigma(j)}(x_i) * h'_{A\sigma(j)}(x_i) \right) \right) \right]^{1/2} \left[\sum_{i=1}^{n} w_{mi} \left(\frac{1}{l_i} \left(\sum_{j=1}^{l_i} h'_{B\sigma(j)}(x_i) * h'_{B\sigma(j)}(x_i) \right) \right) \right]^{1/2}}$$

Where $\delta_m(A,B)$ satisfies the properties

(1) $\delta_m(A,B) = \delta_m(B,A)$

(2) $0 \leq \delta_m(A,B) \leq 1$;

(3) $\delta_m(A,B) = 1$ if A=B

and attribute weight $w_{mi}(i \quad 1 \ldots \ldots n)$ can be defined as

$$w_{mi} = \frac{1 - E_i}{\sum_{i=1}^{n} (1 - E_i)}$$

E_i is the Shannon's entropy that indicates the importance of attributes Hence E_i can be given as

$$E_i = -\frac{1}{\ln m}\sum_{i=1}^{m} s_{ij} \ln s_{ij}$$

Where S_{ij} normalized score matrix as describe [17].

Method 2: The correlation measure considering the attribute weights can also be calculated using the formula as given below which is motivated by Zhang et. al.

$$\varepsilon(A,B) = \frac{1}{2}(\varepsilon_1(A,B) + \varepsilon_2(A,B))$$

$$\varepsilon_1(A,B) = 1 - \left[\sum_{i=1}^{n}\Delta(\alpha,\beta)^{\lambda}\right]^{1/\lambda}$$

Where

$$\alpha = \frac{w_i \bar{h}_A(x_i)}{\left(\sum_{i=1}^{n} w_i\left(\bar{h}_A(x_i)^{\lambda}\right)\right)^{1/\lambda}}$$

$$\beta = \frac{\bar{h}_B(x_i)}{\left(\sum_{i=1}^{n}\left(w_i \bar{h}_B(x_i)\right)^{\lambda}\right)^{1/\lambda}}$$

$$\varepsilon_2(A,B) = 1 - \left[\sum_{i=1}^{n}\Delta(\alpha',\beta')^{\lambda}\right]^{1/\lambda}$$

Where

$$\alpha' = \frac{w_i \tilde{\bar{h}}_A(x_i)}{\left(\sum_{i=1}^{n}\left(w_i \tilde{\bar{h}}_A(x_i)\right)^{\lambda}\right)^{1/\lambda}}$$

$$\beta' = \frac{w_i \tilde{\bar{h}}_B(x_i)}{\left(\sum_{i=1}^{n}\left(w_i \tilde{\bar{h}}_B(x_i)\right)^{\lambda}\right)^{1/\lambda}}$$

Where

$$\bar{h}_A(x_i) = \sum_{j=1}^{l(\bar{h}_A(x_i))}\bar{h}_A^{\sigma(j)}(x_i)/l(\bar{h}_A(x_i))$$

$$\bar{h}_B(x_i) = \sum_{j=1}^{l(\bar{h}_B(x_i))}\bar{h}_B^{\sigma(j)}(x_i)/l(\bar{h}_B(x_i))$$

$$\bar{h}_A(x_i) = \sum_{j=1}^{l(\bar{h}_A(x_i))}\bar{h}_A^{\sigma(j)}(x_i)/\sum_{i=1}^{n}\sum_{j=1}^{l(\bar{h}_A(x_i))}\bar{h}_A^{\sigma(j)}(x_i)$$

$$\bar{h}_B(x_i) = \sum_{j=1}^{l(\bar{h}_B(x_i))}\bar{h}_B^{\sigma(j)}(x_i)/\sum_{i=1}^{n}\sum_{j=1}^{l(\bar{h}_B(x_i))}\bar{h}_B^{\sigma(j)}(x_i)$$

For A_j (j=1,2……m) and hesitant fuzzy set then C=$(\rho_{ij})_{m\times m}$ will be correlation matrix as

ststed in definition 4 given by

$$C = \begin{pmatrix} \rho_{11} & \cdots & \rho_{1n} \\ \vdots & \ddots & \vdots \\ \rho_{m1} & \cdots & \rho_{mn} \end{pmatrix}$$

Material and Methods

In this section, we present the enhanced clustering algorithm pertaining to Entropy Based Ordered Weighted Attribute Selection (EBOW-AS) for hesitant fuzzy sets. A detailed description is also given below.

Step 1: Let { $A_1, A_2, A_3 \ldots\ldots A_n$ } be the set of attributes of a Hesitant fuzzy set (HFS) in $X = \{x_1, x_2, x_3 \ldots\ldots x_n\}$. The hesitant fuzzy decision matrix can be given as follows

$$D = \begin{bmatrix} h_{11} & h_{12} & \ldots & h_{1n} \\ h_{21} & h_{22} & \ldots & \ldots \\ . & \ldots & \ldots & \ldots \\ h_{m1} & \ldots & \ldots & h_{mn} \end{bmatrix}$$

Step 2: Next step is to determine the normalized score matrix S_n where S is the score matrix according to Definition 3.

Step 3: Entropy which is based on the determination of attributes weights as given in Definition 4.

$$E_j = -\frac{1}{\ln m}\sum_{i=1}^{m}\bar{s}_{ij}\ln \bar{s}_{ij} \quad \text{where j=1.......n}$$

Step 4: calculate the correlation coefficients and correlation matrix as discussed in definition 5.

Step 5: Calculate the resultant equivalent correlation matrix C_λ for a deviation threshold of λ. Elements in Correlation matrix which have value greater than or equal to threshold are replaced by 1 and elements having value less than threshold are replaced by 0 i.e

$$\rho_{ij} = 0 \text{ if } \rho_{ij} < \lambda$$

$$= 1 \text{ if } \rho_{ij} \geq \lambda$$

The resultant thresholded matrix shall indicate the required informative attributes and also show the redundant attributes.

Results and Discussion

An illustration of proposed Entropy Based Ordered Weighted Attribute Selection approach which is an algorithm inspired by clustering algorithm given by [1-2, 21] has been discussed below. HFS clustering algorithm for clustering different objects or instances is performed [21]. The idea here is to apply novel entropy based on the attribute weighing algorithm for attribute reduction of hesitant fuzzy sets. A detailed explanation of the algorithm providing an example for attribute reduction of ECG dataset is considered below.

Let x_1, x_2, x_3, x_4 be 4 examples for ECG classification and A_1, A_2, A_3, A_4 be the attributes for feature selection namely

A_1 is QRS_width
A_2 is Rpeak
A_3 is ST_segment duration
A_4 is Pwave onset

Initially, the weights of the attributes are not known, then the hesitant fuzzy decision matrix can be given as shown in the Table 1.

For attribute reduction, correlation coefficients for the attributes need to be calculated. Considering the weight vector given by w=[0.12 0.56 0.08 0.24], the calculations for correlation matrix are shown according to algorithm proposed [18]

$$C_{HFS}(A_1,A_1) = E_{HFS}(A_1) = w_1 \sum_{i=1}^{4} \frac{1}{l_i} \sum_{j=1}^{l_i} h_{A_1(j)}^2(x_i) =$$

$$= 0.12 \left[\frac{1}{2}\left(0.8^2 + 0.82^2\right) + \frac{1}{2}\left(0.4^2 + 0.43^2\right) + \frac{1}{4}\left(0.81^2 + 0.83^2\right) + \frac{1}{4}\left(0.43^2 + 0.44^2\right) \right]$$

$$= 0.246$$

$$C_{HFS}(A_2,A_2) = E_{HFS}(A_2) = w_2 \sum_{i=1}^{4} \frac{1}{l_i} \sum_{j=1}^{l_i} h_{A_2(j)}^2(x_i) =$$

$$= 0.56 \left[\frac{1}{2}\left(0.81^2 + 0.84^2\right) + \frac{1}{2}\left(0.85^2 + 0.86^2\right) + \frac{1}{2}\left(0.66^2 + 0.67^2\right) + \frac{1}{2}\left(0.61^2 + 0.64^2\right) \right]$$

$$= 1.25$$

$$C_{HFS}(A_3,A_3) = E_{HFS}(A_3) = w_3 \sum_{i=1}^{4} \frac{1}{l_i} \sum_{j=1}^{l_i} h_{A_3(j)}^2(x_i) =$$

$$= 0.08 \left[\frac{1}{2}\left(0.51^2 + 0.54^2\right) + \frac{1}{2}\left(0.45^2 + 0.46^2\right) + \frac{1}{2}\left(0.56^2 + 0.57^2\right) + \frac{1}{2}\left(0.41^2 + 0.44^2\right) \right]$$

$$= 0.07$$

$$C_{HFS}(A_4,A_4) = E_{HFS}(A_4) = w_4 \sum_{i=1}^{4} \frac{1}{l_i} \sum_{j=1}^{l_i} h_{A_4(j)}^2(x_i) =$$

$$= 0.24 \left[\frac{1}{2}\left(0.71^2 + 0.74^2\right) + \frac{1}{2}\left(0.35^2 + 0.36^2\right) + \frac{1}{2}\left(0.76^2 + 0.77^2\right) + \frac{1}{2}\left(0.31^2 + 0.34^2\right) \right]$$

$$= 0.318$$

$$\delta_m(A_1,A_2) \le \frac{C_{HFS}(A_1,A_2)}{\left[C_{HFS}(A_1,A_1)\right]^{1/2}\left[C_{HFS}(A_2,A_2)\right]^{1/2}} = \frac{0.554}{\left(\sqrt{0.246}\right)\left(\sqrt{1.25}\right)} \le 1$$

Similarly $\delta_m(A_1,A_3) = 0.9$, $\delta_m(A_1,A_4) = 0.805$, $\delta_m(A_2,A_3) = 0.988$, $\delta_m(A_3,A_4) = 0.9885$

For the example considered above, correlation matrix can be then written as

$$C = \begin{pmatrix} \delta_{m\,A_1A_1} & \delta_{m\,A_1A_2} & \delta_{m\,A_1A_3} & \delta_{m\,A_1A_4} \\ \delta_{m\,A_2A_1} & \delta_{m\,A_2A_2} & \delta_{m\,A_2A_3} & \delta_{m\,A_2A_4} \\ \delta_{m\,A_3A_1} & \delta_{m\,A_3A_2} & \delta_{m\,A_3A_3} & \delta_{m\,A_3A_4} \\ \delta_{m\,A_4A_1} & \delta_{m\,A_4A_2} & \delta_{m\,A_4A_3} & \delta_{m\,A_4A_4} \end{pmatrix}$$

Calculation of correlation matrix is followed by calculating equivalent correlation matrix which further helps in determination of optimal attributes according to value of threshold λ. The above algorithm as given by Chen has been enhanced to provide a mathematical approach for calculation of weight vector which has not been given by [21]. Accordingly the steps to be followed in the proposed Entropy Based Ordered Weighted Attribute Selection approach(EBOW-AS) for hesitant fuzzy sets are as follows.

Step 1: Determination of hesitant fuzzy decision matrix D from given data is shown in Table 1.

Table 1: Hesitant Fuzzy Decision Matrix

D	x_1	x_2	x_3	x_4
A_1	{0.8,0.82}	{0.4,0.43}	{0.81,0.83}	{0.43,0.44}
A_2	{0.81,0.84}	{0.85,0.86}	{0.66,0.67}	{0.61,0.64}
A_3	{0.51,.54}	{0.45,0.46}	{0.56,0.57}	{0.41,0.44}
A_4	{0.71,.74}	{0.35,0.36}	{0.76,0.77}	{0.31,0.34}

Step 2: Next step is to determine the normalized score matrix S_n where S is the score matrix according to Definition

$$S = \begin{bmatrix} 0.81 & 0.415 & 0.82 & 0.435 \\ 0.82 & 0.855 & 0.665 & 0.625 \\ 0.525 & 0.455 & 0.565 & 0.425 \\ 0.725 & 0.355 & 0.765 & 0.325 \end{bmatrix}$$

$$S_n = \begin{bmatrix} 0.28 & 0.20 & 0.29 & 0.24 \\ 0.28 & 0.41 & 0.24 & 0.35 \\ 0.18 & 0.22 & 0.20 & 0.23 \\ 0.25 & 0.17 & 0.27 & 0.18 \end{bmatrix}$$

Step 3: Entropy based determination of attributes weights is a significant step that finally determines the significance of attributes [17, 19].

$$E_j = -\frac{1}{\ln m} \sum_{i=1}^{m} \bar{s}_{ij} \ln \bar{s}_{ij}$$

Where j=1.......n

$$E_j = [0.99 \quad 0.95 \quad 0.99 \quad 0.98]$$

$$w_{mj} = [0.12 \quad 0.56 \quad 0.08 \quad 0.24]$$

Using the above weight vector, the calculation of correlation matrix has already been shown in the above section.

Step 4: Calculate correlation coefficients to generate the correlation matrix using method 2 which is a modified approach [17]. The modified approach involves the clustering of the attributes by calculating the resultant equivalent matrix to generate the non-redundant informative attributes. The correlation coefficients as given by [17] are as follows

$$\varepsilon_{HFS}(A_1, A_2) = 0.84, \varepsilon_{HFS}(A_1, A_3) = 0.96$$

$$\varepsilon_{HFS}(A_1, A_4) = 0.86, \varepsilon_{HFS}(A_2, A_3) = 0.942$$

$$\varepsilon_{HFS}(A_2, A_4) = 0.78, \varepsilon_{HFS}(A_3, A_4) = 0.61$$

Let Aj (j= (1,2…..m) be m HFS and C=(ε_{ij}) $_{m \times m}$ be correlation matrix.

Step5: Calculate the resultant equivalent correlation matrix for a deviation threshold of λ, and find the C_λ matrix. If the elements of i -th column in are same as the elements of the j-th line in C_λ matrix, the corresponding HFS will be the same type. Elements in correlation matrix which have values greater than or equal to threshold and elements which have value less than threshold are replaced by 1, 0, respectively as discussed in Step 5.

Correlation matrixes generated for the weighted attribute reduction for the ECG example under discussion are nas follows

$$C = \begin{pmatrix} 1 & 0.84 & 0.96 & 0.86 \\ 0.84 & 1 & 0.94 & 0.78 \\ 0.96 & 0.94 & 1 & 0.61 \\ 0.86 & 0.78 & 0.61 & 1 \end{pmatrix}$$

$$C^2 = C \circ C = \max(\min(\rho_{ij}, \rho_{jk}))$$

$$C^2 = \begin{pmatrix} 1 & 0.94 & 0.96 & 0.86 \\ 0.94 & 1.0 & 0.94 & 0.84 \\ 0.96 & 0.94 & 1.0 & 0.86 \\ 0.86 & 0.84 & 0.86 & 1 \end{pmatrix}$$

$$C^3 = \begin{pmatrix} 1 & 0.94 & 0.96 & 0.86 \\ 0.94 & 1.0 & 0.94 & 0.84 \\ 0.96 & 0.94 & 1.0 & 0.86 \\ 0.86 & 0.84 & 0.86 & 1 \end{pmatrix} = C^2$$

C^3 is an equivalent matrix.

Step 6: For a deviation threshold of λ =0.9, calculated C_λ matrix would be as follows

$$C_{0.9} = \begin{pmatrix} 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

For a deviation threshold of λ =0.8, calculate C_λ matrix would be as follows

$$C_{0.8} = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{pmatrix}$$

Hence attribute similarity cluster would all look alike i.e. redundant attributes would be large in number. Selection of proper deviation threshold would change as per the application. Selecting the deviation threshold of 0.9 has shown that attributes A_1 and A_3 are redundant while attribute A_4 and A_2 are non-redundant or essential attributes. The above approach provides an appropriate model for attribute reduction and clustering on the basis of entropy. This approach has been compared with classical approaches involving the distance based similarity measure to rank the significance of

attributes. The results for the same are tabulated below in Table 2.

The tabulated results indicate the order of significance for the attributes; however, the redundant attributes cannot be found. The proposed entropy based on the model for hesitant fuzzy sets is able to effectively determine the optimal informative attributes. Moreover, the mathematical model proposed for calculation of weight vector for hesitant fuzzy sets is able to provide an exact evaluation of significant instances.

Table 2: Comparative Analysis

Sr.No	Similarity Measure Applied to ECG da-taset	Significance values for attributes (A_1....A_4)	Rankings ob-tained for the attributes (A_1....A_4)
1.	Ordered Weighted hamming Distance meas-ure(OWD) [1-2]	0.81>0.78>0.77>0.63	$A_4 \succ A_3 \succ A_1 \succ A_2$
2.	Relative closeness measure [22] based on ordered weighted square root distance measure	0.51>0.49>0.48>0.38	$A_4 \succ A_1 \succ A_3 \succ A_2$
3.	Ordered weighted geometric method [3]	0.80>0.78>0.77>0.63	$A_4 \succ A_3 \succ A_1 \succ A_2$

Conclusion

In this paper, we have proposed a novel Entropy Based Ordered Weighted Attribute Selection approach for hesitant fuzzy sets to select optimal attributes. Development of a comprehensive mathematical model to calculate the weight vector for the hesitant fuzzy sets is a novel feature of the proposed work. The comparison with the existing distance based on measures has been extensively carried out to prove the efficacy of the proposed model. The illustrations in the examples indicate the efficiency and simplicity of the developed HFS application. The future directions will include development of a generic mathematical model for weight calculation to have an appropriate selection of attributes to find the most informative ones. This work can also be extended for development of an appropriate classification model for hesitant fuzzy sets.

Acknowledgment

Conflict of Interest

None

References

1. Xia M, Xu Z, Chen N. Some hesitant fuzzy aggregation operators with their application in group decision making. *Group Decision and Negotiation*. 2013;**22**:259-79.

2. Chen N, Xu Z, Xia M. Correlation coefficients of hesitant fuzzy sets and their applications to clustering analysis. *Applied Mathematical Modelling*. 2013;**37**:2197-211.doi: 10.1016/j.apm.2012.04.031.

3. Wang J-q, Wu J-t, Wang J, Zhang H-y, Chen X-h. Interval-valued hesitant fuzzy linguistic sets and their applications in multi-criteria decision-making problems. *Information Sciences*. 2014;**288**:55-72. doi: 10.1016/j.ins.2014.07.034.

4. Peng J-j, Wang J-q, Wang J, Yang L-J, Chen X-h. An extension of ELECTRE to multi-criteria decision-making problems with multi-hesitant fuzzy sets. *Information Sciences*. 2015;**307**:113-26.

5. Torra V. Hesitant fuzzy sets. *International Journal of Intelligent Systems*. 2010;**25**:529-39.

6. Torra V, Narukawa Y, editors. On hesitant fuzzy sets and decision. 20-24 Aug. 2009. Jeju Island:

IEEE International Conference on Fuzzy Systems. 2009. p. 1378-82.

7. Rodriguez RM, Martinez L, Herrera F. Hesitant fuzzy linguistic term sets for decision making. *IEEE Transactions on Fuzzy Systems*. 2012;**20**:109-19.

8. Castillo O, Melin P, Ramírez E, Soria J. Hybrid intelligent system for cardiac arrhythmia classification with Fuzzy K-Nearest Neighbors and neural networks combined with a fuzzy system. *Expert Systems with Applications*. 2012;**39**:2947-55.

9. Abenstein JP, Tompkins WJ. A new data-reduction algorithm for real-time ECG analysis. *IEEE Trans Biomed Eng*. 1982;**29**:43-8. doi: 10.1109/TBME.1982.324962. PubMed PMID: 7076268.

10. Sufi F, Khalil I. Diagnosis of cardiovascular abnormalities from compressed ECG: a data mining-based approach. *IEEE Trans Inf Technol Biomed*. 2011;**15**:33-9. doi: 10.1109/TITB.2010.2094197. PubMed PMID: 21097383.

11. Dubois D, Prade H. Fuzzy sets in approximate reasoning, Part 1: Inference with possibility distributions. *Fuzzy sets and systems*. 1991;**40**:143-202doi: 10.1016/0165-0114(91)90050-z.

12. Pawlak Z. Rough sets. *International journal of computer & information sciences*. 1982;**11**:341-56.

13. Jensen R, Shen Q. New approaches to fuzzy-rough feature selection. *IEEE Transactions on Fuzzy Systems*. 2009;**17**:824.doi: 10.1109/tfuzz.2008.924209.

14. Bhatt RB, Gopal M. On fuzzy-rough sets approach to feature selection. *Pattern recognition letters*. 2005;**26**:965-75.doi: 10.1016/j.patrec.2004.09.044.

15. Shen Q, Jensen R. Selecting informative features with fuzzy-rough sets and its application for complex systems monitoring. *Pattern recognition*. 2004;**37**:1351-63.doi: 10.1016/j.patcog.2003.10.016.

16. Wang G-y, Yu H, Yang D. Decision table reduction based on conditional information entropy. *CHINESE JOURNAL OF COMPUTERS-CHINESE EDITION-*. 2002;**25**:759-66.

17. Wang J, Zhang R, Buchmeister B, Wang R. Generalized Dual Hesitant Fuzzy Bonferroni Mean with Its Application to Supplier Selection. Tokyo: IEEE International Conference on Logistics, Informatics and Service Sciences; 2017.

18. Farhadinia B. Information measures for hesitant fuzzy sets and interval-valued hesitant fuzzy sets. *Information Sciences*. 2013;**240**:129-44.doi: /10.1016/j.ins.2013.03.034.

19. Yu D, Zhang W, Huang G. Dual hesitant fuzzy aggregation operators. *Technological and Economic Development of Economy*. 2016;**22**:194-209.

20. Zadeh LA. Probability measures of fuzzy events. *Journal of mathematical analysis and applications*. 1968;**23**:421-7.doi: 10.1016/0022-247x(68)90078-4.

21. Chen N, Xu Z, Xia M. Correlation coefficients of hesitant fuzzy sets and their applications to clustering analysis. *Applied Mathematical Modelling*. 2013;**37**:2197-211.doi: 10.1016/j.apm.2012.04.031.

22. Bai C, Zhang R, Qian L, Wu Y. Comparisons of probabilistic linguistic term sets for multi-criteria decision making. *Knowledge-Based Systems*. 2017;**119**:284-91.doi: 10.1016/j.knosys.2016.12.020.