

# Fuzzy-Rough Membership Functions

Manish Sarkar and B. Yegnanarayana  
Department of Computer Science & Engineering  
Indian Institute of Technology, Madras - 600 036, INDIA  
E-mail: manish@bronto.iitm.ernet.in

## ABSTRACT

This paper generalizes the concepts of rough membership functions in pattern classification tasks to fuzzy-rough membership functions. Unlike the rough membership value of a pattern, which is sensitive only towards the rough uncertainty associated with the pattern, the fuzzy-rough membership value of the pattern signifies the rough uncertainty as well as the fuzzy uncertainty associated with it. In absence of fuzziness, the fuzzy-rough membership functions reduce to the existing rough membership functions. Moreover, under certain conditions the fuzzy-rough membership functions are equivalent to fuzzy membership functions or characteristic functions. In this paper, various set theoretic properties of the fuzzy-rough membership functions are exploited to characterize the concept of fuzzy-rough sets. Some measures of the fuzzy-rough ambiguity associated with a given output class are also discussed.

## I. INTRODUCTION

In any classification task the aim is to form various classes where each class contains objects that are not noticeably different. These *indiscernible* or indistinguishable objects can be viewed as basic building blocks (concepts) used to build up a knowledge base about the real world. For example, if the objects are classified according to color (red, black) and shape (triangle, square and circle), then the classes are: red triangles, black squares, red circles, etc. Thus, these two attributes make a *partition* in the set of objects and the universe becomes coarse. Now, if two red triangles with different areas belong to different classes, it is impossible for anyone to correctly classify these two red triangles based on the given two attributes. This kind of uncertainty is referred to as *rough uncertainty* [14]. The rough uncertainty is formulated in terms of *rough sets* [15]. Obviously, the rough uncertainty can be completely avoided if we can successfully extract the essential features so that distinct feature vectors are used to represent different objects. But, it may not be possible to guarantee as our knowledge about the system generating the data is limited. Therefore, rough sets are essential to deal with a classification system, where we do not have complete knowledge about the system.

*Fuzzy sets*, a generalization of classical sets, is a mathematical tool to model the vagueness present in the human classification mechanism. In the classical set theory, an element of the universe either belongs to or does not belong to a set. That is, the belongingness of the element is crisp: it is either *yes* (in the set) or *no* (not in the set). In fuzzy sets, the belongingness of the element can be anything in between *yes* or *no*; for example, a set of *tall* persons. We cannot classify a person as tall into *yes/no* category, as there does not exist any well-defined boundary for the set *tall*. The concept of fuzzy sets is important in pattern classification, because the psycho-physiological process involved in the human reasoning does not employ any precise mathematical formulation [13].

However, probabilistic, fuzzy and rough uncertainties are often confused. In fact, they are different facets of uncertainty. Fuzziness deals with vagueness between the overlapping sets [2] [3], while probability concerns the likelihood of randomness of a phenomenon [9]; on the other hand, rough sets deal with coarse non-overlapping concepts [5] [6]. Both roughness and fuzziness do not depend on the occurrence of the

event; whereas probability does. Fuzziness lies in the subsets defined by the linguistic variables, like *tall*, *big*, whereas indiscernibility is a property of the referential itself, as perceived by some observers, not of its subsets [6]. In fuzzy sets, each granule of knowledge can have only one membership value to a particular class. However, rough sets assert that each granule may have different membership values to the same class. Fuzzy sets deal with overlapping classes and *fine* concepts; whereas rough sets deal with nonoverlapping classes and coarse concepts.

In a classification task, the indiscernibility relation, based on the equivalence of the features of the patterns, partitions the input pattern set into several equivalence classes. These equivalence classes try to approximate the given output class. When the approximation is not proper, the roughness is generated. In most of the real life cases, the features of the two patterns may not be exact, but similar. Therefore, the indiscernibility relation formulated based on the features do not obey the law of equivalence, and is a matter of degree. Hence, the equivalence relation takes the form of a similarity relation, and the equivalence classes form fuzzy clusters. The whole situation becomes more complicated because the output classes can be fuzzy too. Thus, the roughness appears here due to the indiscernibility relation in the input pattern set, and the fuzziness is generated due to the vagueness present in the output class and the clusters. To model this type of situation, where both vagueness and approximation are present, the concept of *fuzzy-rough sets* [5] can be employed. The resultant model is expected to be more powerful than either of rough sets or fuzzy sets.

In the classical set theory, one of the fundamental notion is the *characteristic function* or *membership function* [7]. When one considers subsets of a given universe based on the available partial information, it is possible to apply the characteristic functions to determine whether or not a given element belongs to a particular set. Following this direction, the concept of *fuzzy-rough membership function* is proposed in this paper. In absence of the fuzziness, the proposed membership function is reduced to the rough membership function [22]. In addition, when the clusters and the output classes are crisp, the proposed membership function becomes equivalent to the rough-fuzzy membership function [19], [20]. If the clusters are crisp and each cluster contains only one pattern, then the fuzzy-rough membership function of the pattern for a given output class is further reduced to the fuzzy membership function of the pattern for the output class. In this paper, we discuss some set theoretic properties of the proposed membership function, and subsequently we use it to characterize the concepts of the fuzzy-rough sets. Finally, these properties are used to measure the fuzzy-rough ambiguity associated with a given output class.

## II. BACKGROUND

### A. Rough Sets

Let  $R$  be an equivalence relation on a universal set  $X$ . Moreover, let  $X/R$  denote the family of all equivalence classes induced on  $X$  by  $R$ . One such equivalence class in  $X/R$ , that contains  $x \in X$ , is designated by  $[x]_R$ . For any output class  $A \subseteq X$ , we can define the lower  $\underline{R}(A)$  and upper  $\overline{R}(A)$  approximations, which approach  $A$  as closely as possible from inside and outside, respectively [7]. Here,

$$\underline{R}(A) = \cup\{[x]_R \mid [x]_R \subseteq A, x \in X\} \quad (1-a)$$

This work was carried out while Manish Sarkar held a Dr. K. S. Krishnan Research Fellowship from Department of Atomic Energy, Government of India.

is the union of all equivalence classes in  $X/R$  that are contained in  $A$ , and

$$\overline{R}(A) = \cup\{[x]_R \mid [x]_R \cap A \neq \emptyset, x \in X\} \quad (1-b)$$

is the union of all equivalence classes in  $X/R$  that overlap with  $A$ . A rough set  $R(A) = \langle \overline{R}(A), \underline{R}(A) \rangle$  is a representation of the given set  $A$  by  $\underline{R}(A)$  and  $\overline{R}(A)$  [15]. The set  $BN(A) = \overline{R}(A) - \underline{R}(A)$  is a rough description of the boundary of  $A$  by the equivalence classes of  $X/R$ . The approximation is rough uncertainty free if  $\overline{R}(A) = \underline{R}(A)$ . Thus, when all the patterns from an equivalence class do not carry the same output class label, rough ambiguity is generated as a manifestation of the one-to-many relationship between that equivalence class and the output class labels.

The *rough membership function*  $r_A(x) : A \rightarrow [0, 1]$  of a pattern  $x \in X$  for the output class  $A$  is defined by

$$r_A(x) = \frac{\| [x]_R \cap A \|}{\| [x]_R \|} \quad (2)$$

where  $\|A\|$  denotes the cardinality of the set  $A$ .

### B. Fuzzy Sets

In traditional two-state classifiers, where a class  $A$  is defined as a subset of a universal set  $X$ , any input pattern  $x \in X$  can either be a member or not be a member of the given class  $A$ . This property of whether or not a pattern  $x$  of the universal set belongs to the class  $A$  can be defined by a *characteristic function*  $\mu_A : X \rightarrow \{0, 1\}$  as follows:

$$\mu_A(x) = \begin{cases} 1 & \text{if and only if } x \in A \\ 0 & \text{if and only if } x \notin A \end{cases} \quad (3)$$

In real life situations, however, boundaries between the classes may be overlapping. Hence, it is uncertain whether an input pattern belongs totally to the class  $A$ . To take care of such situations, in fuzzy sets [1], the concept of characteristic function has been modified to *fuzzy membership function*  $\mu_A : X \rightarrow [0, 1]$ . This function is called membership function, because larger value of the function denotes more membership of the element to the set under consideration.

### C. Rough-Fuzzy Sets

Let  $X$  be a set,  $R$  be an equivalence relation defined on  $X$  and the output class  $A \subseteq X$  be a fuzzy set. A rough-fuzzy set is a tuple  $\langle \overline{R}(A), \underline{R}(A) \rangle$ , where the lower approximation  $\underline{R}(A)$  and the upper approximation  $\overline{R}(A)$  of  $A$  are fuzzy sets of  $X/R$ , with membership functions defined by [6]

$$\mu_{\underline{R}(A)}([x]_R) = \inf \{ \mu_A(x) \mid x \in [x]_R \} \quad (4-a)$$

$$\mu_{\overline{R}(A)}([x]_R) = \sup \{ \mu_A(x) \mid x \in [x]_R \} \quad (4-b)$$

Here,  $\mu_{\underline{R}(A)}([x]_R)$  and  $\mu_{\overline{R}(A)}([x]_R)$  are the membership values of  $[x]_R$  in  $\underline{R}(A)$  and  $\overline{R}(A)$ , respectively.

The rough-fuzzy membership function of a pattern  $x \in X$  for the fuzzy output class  $C_c \subseteq X$  is defined by [19] [20]

$$\iota_{C_c}(x) = \frac{\| F \cap C_c \|}{\| F \|} \quad (5)$$

where  $F = [x]_R$ , and  $\|C_c\|$  is equal to the cardinality of the fuzzy set  $C_c$ . One possible way to determine the cardinality is to use [23]  $\|C_c\| \stackrel{\text{def}}{=} \sum_{x \in X} \mu_{C_c}(x)$ . For the '∩' (intersection)

operation, we can use [23]  $\mu_{A \cap B}(x) \stackrel{\text{def}}{=} \min\{\mu_A(x), \mu_B(x)\} \forall x \in X$ . When the output class is crisp, the definition (5) reduces to the definition (2).

### D. Fuzzy-Rough Sets

When the equivalence classes are not crisp, they are in form of fuzzy clusters  $\{F_1, F_2, \dots, F_H\}$  generated by a *fuzzy weak partition* [6] of the input set  $X$ . The term fuzzy weak partition means that each  $F_j, j \in \{1, 2, \dots, H\}$  is a normal fuzzy set, i.e.,  $\max_x \mu_{F_j}(x) = 1$  and  $\inf_x \max_j \mu_{F_j}(x) > 0$  while

$$\sup_x \min_i \{ \mu_{F_i}(x), \mu_{F_j}(x) \} < 1 \quad \forall i, j \in \{1, 2, \dots, H\} \quad (6)$$

Here  $\mu_{F_j}(x)$  is the fuzzy membership function of the pattern  $x$  in the cluster  $F_j$ . In addition, the output classes  $C_c, c = \{1, 2, \dots, C\}$  may be fuzzy too. Given a weak fuzzy partition  $\{F_1, F_2, \dots, F_H\}$  on  $X$ , the description of any fuzzy set  $C_c$  by means of the fuzzy partitions under the form of an upper and a lower approximation  $\overline{C}_c$  and  $\underline{C}_c$  is as follows:

$$\mu_{\underline{C}_c}(F_j) = \inf_{x \in C_c} \max\{1 - \mu_{F_j}(x), \mu_{C_c}(x)\} \forall x \quad (7-a)$$

$$\mu_{\overline{C}_c}(F_j) = \sup_{x \in C_c} \min\{\mu_{F_j}(x), \mu_{C_c}(x)\} \forall x \quad (7-b)$$

The tuple  $\langle \underline{C}_c, \overline{C}_c \rangle$  is called a fuzzy-rough set. Here,  $\mu_{C_c}(x) = \{0, 1\}$  is the fuzzy membership of the input  $x$  to the class  $C_c$ . Fuzzy-roughness appears when a fuzzy cluster contains patterns that belong to different classes.

## III. FUZZY-ROUGH MEMBERSHIP FUNCTIONS

### A. Definition

If the equivalent classes form the fuzzy clusters  $\{F_1, F_2, \dots, F_H\}$ , then each fuzzy cluster can be considered as a fuzzy linguistic variable. Using this fact, we generalize the definition (5) to the following definition of the fuzzy-rough membership function:

$$\tau_{C_c}^f(x) = \begin{cases} \frac{1}{\sum_j \mu_{F_j}(x)} \sum_{j=1}^H \mu_{F_j}(x) \iota_{C_c}^j & \text{if } \exists j \text{ with } \mu_{F_j}(x) > 0 \\ 0 & \text{otherwise} \end{cases} \quad (8)$$

Here,  $\iota_{C_c}^j = \frac{\| F_j \cap C_c \|}{\| F_j \|}$ , which can be viewed as a modified form of (2). The term,  $\frac{1}{\sum_j \mu_{F_j}(x)}$ , in the above definition normalises

the fuzzy membership  $\mu_{F_j}(x)$ , and therefore, we call the above function *constrained fuzzy-rough membership function*. We can further modify the above definition to the following:

$$\tau_{C_c}^p(x) = \begin{cases} \frac{1}{\hat{H}} \sum_{j=1}^{\hat{H}} \mu_{F_j}(x) \iota_{C_c}^j & \text{if } \exists j \text{ with } \mu_{F_j}(x) > 0 \\ 0 & \text{otherwise} \end{cases} \quad (9)$$

where  $\hat{H}$  is number of the clusters in which  $x$  has nonzero memberships. When  $x$  does not belong to any cluster, then  $\hat{H} = 0$ , and hence,  $\frac{1}{\hat{H}} \sum_{j=1}^{\hat{H}} \mu_{F_j}(x) \iota_{C_c}^j$  becomes undefined. In order to avoid this problem,  $\tau_{C_c}^p(x)$  is made equal to zero when  $x$  does not belong to any cluster. Since this definition does not normalize the fuzzy membership values, we call it *possibilistic fuzzy-rough membership function*.

The above two definitions, characterize the fuzzy-roughness of a pattern in different manners. The first definition expresses that if an input pattern belongs only to the clusters (all of which belong to only one class) with nonzero memberships, then no fuzzy-roughness is involved. The second definition implies that if an input pattern *completely* belongs to the clusters (all of which belong to only one class), then no fuzzy-roughness is involved. Unlike the first definition, in the second definition, it matters to what extent the pattern belongs to the clusters. This difference will be clearer while discussing the property 11 in the section iii-B.

## B. Properties of Fuzzy-Rough Membership Functions

Most of the properties discussed below are applicable to both the definitions (8) and (9). In what follows, we will use  $\tau_{C_c}(x)$  to denote both  $\tau_{C_c}^f(x)$  and  $\tau_{C_c}^p(x)$ . In case, certain property does not hold for both the definitions, we state that explicitly.

**Property 1:**  $0 \leq \tau_{C_c}(x) \leq 1$

*Proof:* Since  $\phi \subseteq F_j \cap C_c \subseteq F_j$ ,  $0 \leq \nu_{C_c}^j \leq 1$ . Moreover,  $0 \leq \mu_{F_j}(x) \leq 1$ . Hence, the proof follows. ■

**Property 2:**  $\tau_{C_c}(x) = 1$  or  $0$  if and only if no fuzzy-rough uncertainty is associated with the pattern  $x$ .

*Proof:* We show the proof for the possibilistic fuzzy-rough membership function, but the same method can be followed for the constrained fuzzy-rough membership function as well.

*If part:* If no fuzzy-rough uncertainty is involved, then  $x$  must belong completely to all the clusters in which it has non-zero belongingness. It implies  $\mu_{F_j}(x) = 1$  for which  $\mu_{F_j}(x) > 0$ . Moreover, all the clusters in which  $x$  has non-zero belongingness either (a) must be the subsets of the class  $C_c$ , or (b) must not share any pattern with the class  $C_c$ . In other words, the condition (a) implies that  $F_j \subseteq C_c \forall j$  for which  $\mu_{F_j}(x) > 0$ .

Hence,  $\tau_{C_c}^p(x) = \frac{1}{H} \sum_{j=1}^H 1.1 = 1$ . Similarly the condition (b) expresses that  $F_j \cap C_c = \phi \forall j$  for which  $\mu_{F_j}(x) > 0$ . Hence,  $\tau_{C_c}^p(x) = \frac{1}{H} \sum_{j=1}^H \mu_{C_c}(x).0 = 0$

*Only if part:*  $\tau_{C_c}^p(x) = 0$  implies that either  $x$  does not belong to any cluster, or each term under the summation symbol, i.e.,  $\mu_{F_j}(x)\nu_{C_c}^j$  is separately zero. In the first case, there is no fuzzy-roughness associated with  $x$ . The second case implies that either  $\mu_{F_j}(x)$  or  $\nu_{C_c}^j$ , or both  $\mu_{F_j}(x)$  and  $\nu_{C_c}^j$  are zero. If  $\mu_{F_j}(x) = 0$ , then the pattern  $x$  does not belong to the cluster  $F_j$ , and hence, no fuzzy-rough uncertainty is associated with  $x$ . If  $\nu_{C_c}^j = 0$ , then  $F_j$  and  $C_c$  do not have any pattern common, and therefore, no fuzzy-rough uncertainty exists with  $x$ . Thus,  $\tau_{C_c}^p(x) = 0$  implies that fuzzy-roughness is not associated with the pattern  $x$ . If  $\tau_{C_c}^p(x) = 1$ , then  $\mu_{F_j}(x) = 1$  and  $\nu_{C_c}^j = 1$ ,  $\forall j = 1, 2, \dots, H$ . It also indicates the absence of fuzzy-roughness.

It is to be noted here that if fuzzy-rough uncertainty is absent,  $H > 1$  and  $\tau_{C_c}^p(x) \neq 0$ , then  $\tau_{C_c}^p(x)$  never becomes one, rather it approaches towards one. It is because, the condition expressed in (6) does not allow  $\mu_{F_j}(x) = 1$  to be true for more than one cluster. However, it hardly happens in practice as it needs two cluster centers to be same. ■

**Property 3:** If no fuzzy linguistic uncertainty is associated with the pattern  $x$ , then  $\tau_{C_c}(x) = \nu_{C_c}^j$  for some  $j \in \{1, 2, \dots, H\}$ .

*Proof:* If no fuzzy linguistic uncertainty is involved, then  $\mu_{F_j}(x) = 1$  for some  $j \in \{1, 2, \dots, H\}$ , and  $\mu_{F_k}(x) = 0$ , for  $k \in \{1, 2, \dots, H\}$ ,  $j \neq k$ . Hence,  $\tau_{C_c}(x) = \nu_{C_c}^j$ ,  $j \in \{1, 2, \dots, H\}$ . ■

**Property 4:** If no fuzzy linguistic and fuzzy classification uncertainties are associated with the pattern  $x$ , then  $\tau_{C_c}(x) = r_{C_c}(x)$ .

*Proof:* If no fuzzy linguistic uncertainty is involved, then each cluster is crisp. Consequently, the input pattern belongs to only one cluster. Let it be the  $j$ th cluster. So,  $\mu_{F_j}(x) = 1$  and  $\mu_{F_k}(x) = 0 \forall k \neq j$ . Since the classification is crisp,  $\tau_{C_c}(x) = \frac{\|F_j \cap C_c\|}{\|F_j\|} = r_{C_c}(x)$  (see the definition (2)). ■

**Property 5:** When each cluster is crisp and fine, that is, each cluster consists of a single pattern and the associated cluster

memberships are crisp,  $\tau_{C_c}(x)$  is equivalent to the fuzzy membership of  $x$  in the class  $C_c$ . If the output class is also crisp, then  $\tau_{C_c}(x)$  is equivalent to the characteristic function of  $x$  in the class  $C_c$ .

*Proof:* Since each cluster is crisp and fine,  $\tau_{C_c}(x) = 1. \mu_{C_c}(x) = \mu_{C_c}(x)$ . In addition, if the output class is crisp, then  $\tau_{C_c}(x)$  lies in  $\{0, 1\}$ , and thus, it becomes the characteristic function. ■

**Property 6:** If  $x$  and  $y$  are the two input patterns with  $\mu_{F_j}(x) = \mu_{F_j}(y) \forall j$  and  $\mu_{F_{C_c}}(x) = \mu_{F_{C_c}}(y)$ , then  $\tau_{C_c}(x) = \tau_{C_c}(y)$ .

*Proof:* Directly comes from the definitions (8) and (9). ■

**Property 7:**  $\tau_{X-C_c}^p(x) = \frac{1}{H} \sum_j \mu_{F_j}(x) - \tau_{C_c}^p(x)$  and  $\tau_{X-C_c}^f(x) = 1 - \tau_{C_c}^f(x)$ .

*Proof:*  $\tau_{X-C_c}^p(x) = \frac{1}{H} \sum_j \mu_{F_j}(x) \frac{\|F_j \cap (X - C_c)\|}{\|F_j\|} = \frac{1}{H} \sum_j \mu_{F_j}(x) \left(1 - \frac{\|F_j \cap (C_c)\|}{\|F_j\|}\right) = \frac{1}{H} \sum_j \mu_{F_j}(x) - \tau_{C_c}^p(x)$ . Similarly, in case of constrained fuzzy-rough membership, it can be shown that  $\tau_{X-C_c}^f(x) = 1 - \tau_{C_c}^f(x)$ . ■

**Property 8:**  $\tau_{C_c \cup B}(x) \geq \max\{\tau_{C_c}(x), \tau_B(x)\}$

*Proof:*  $\tau_{C_c \cup B}^p(x) = \frac{1}{H} \sum_j \mu_{F_j}(x) \frac{\|F_j \cap (C_c \cup B)\|}{\|F_j\|} \geq \frac{1}{H} \sum_j \mu_{F_j}(x) \frac{\|F_j \cap C_c\|}{\|F_j\|} = \tau_{C_c}^p(x)$ . Similarly,  $\tau_{C_c \cup B}^p(x) \geq \tau_B^p(x)$ . Therefore,  $\tau_{C_c \cup B}^p(x) \geq \max\{\tau_{C_c}^p(x), \tau_B^p(x)\}$ .

Likewise,  $\tau_{C_c \cup B}^f(x) \geq \max\{\tau_{C_c}^f(x), \tau_B^f(x)\}$  ■

**Property 9:**  $\tau_{C_c \cap B}(x) \leq \min\{\tau_{C_c}(x), \tau_B(x)\}$

*Proof:*  $\tau_{C_c \cap B}^p(x) = \frac{1}{H} \sum_j \mu_{F_j}(x) \frac{\|F_j \cap (C_c \cap B)\|}{\|F_j\|} \leq \frac{1}{H} \sum_j \mu_{F_j}(x) \frac{\|F_j \cap C_c\|}{\|F_j\|} = \tau_{C_c}^p(x)$ . Similarly,  $\tau_{C_c \cap B}^p(x) \leq \tau_B^p(x)$ . Therefore,  $\tau_{C_c \cap B}^p(x) \leq \min\{\tau_{C_c}^p(x), \tau_B^p(x)\}$ .

Likewise,  $\tau_{C_c \cap B}^f(x) \leq \min\{\tau_{C_c}^f(x), \tau_B^f(x)\}$  ■

**Property 10:** If  $U$  is a family of pairwise disjoint crisp subsets of  $X$ , then  $\tau_{\cup U}(x) = \sum_{C_c \in U} \tau_{C_c}(x)$ .

*Proof:*  $\tau_{\cup U}^p(x) = \frac{1}{H} \sum_j \mu_{F_j}(x) \frac{\|F_j \cap (\cup U)\|}{\|F_j\|} = \frac{1}{H} \sum_j \mu_{F_j}(x) \frac{\|\cup (F_j \cap U)\|}{\|F_j\|} = \sum_{C_c \in U} \tau_{C_c}^p(x)$ .

Similarly,  $\tau_{\cup U}^f(x) = \sum_{C_c \in U} \tau_{C_c}^f(x)$ . ■

**Property 11:** For a  $C$ -class classification problem with crisp output classes, the possibilistic fuzzy-rough membership functions behave in a possibilistic manner, whereas the constrained fuzzy-rough membership functions lack this property.

*Proof:*

$$\begin{aligned} \sum_{c=1}^C \tau_{C_c}^p(x) &= \frac{1}{H} \sum_{c=1}^C \sum_{j=1}^H \mu_{F_j}(x) \frac{\|F_j \cap C_c\|}{\|F_j\|} \\ &= \frac{1}{H} \sum_{j=1}^H \mu_{F_j}(x) \frac{\sum_{c=1}^C \sum_y \min\{\mu_{F_j}(y), \mu_{C_c}(y)\}}{\sum_y \mu_{F_j}(y)} \end{aligned}$$

$$\begin{aligned}
&= \frac{1}{\hat{H}} \sum_{j=1}^{\hat{H}} \mu_{F_j}(x) \frac{\sum_{c=1}^C \sum_{y \in C_c} \mu_{F_j}(y)}{\sum_y \mu_{F_j}(y)} \\
&= \frac{1}{\hat{H}} \sum_{j=1}^{\hat{H}} \mu_{F_j}(x) \frac{\sum_y \mu_{F_j}(y)}{\sum_y \mu_{F_j}(y)} \\
&= \frac{1}{\hat{H}} \sum_{j=1}^{\hat{H}} \mu_{F_j}(x) \quad (10)
\end{aligned}$$

Since  $\sum_{c=1}^C \tau_{C_c}^p(x)$  needs not to be equal to a constant, the resultant classification procedure is *possibilistic* [7] [11]. Hence,  $\tau_{C_c}^p(x)$  can distinguish between *equal evidence* and *ignorance*, which are well discussed in belief theory [21] and possibility theory [7].

Similarly, for the constrained fuzzy-rough membership function with crisp output classes, it can be shown that  $\sum_{c=1}^C \tau_{C_c}^f(x)$  is equal to one (except the case when  $\tau_{C_c}^f(x) = 0 \forall c$ ), which is a constant. Therefore, it does not have the possibilistic classification ability, i.e., this kind of membership assignment cannot indicate how *typical* an input pattern is.

This property explains why (8) and (9) are called constrained fuzzy-rough membership function and possibilistic fuzzy-rough membership function, respectively. ■

*Property 12:* If the output class is fuzzy, then  $0 \leq \sum_{c=1}^C \tau_{C_c}(x) \leq C$ .

*Proof:* We show the result for possibilistic fuzzy-rough membership function. Likewise, it can be shown for constrained fuzzy-rough membership function also.

If the input pattern does not belong to any cluster, then from the definition (9),  $\tau_{C_c}(x) = 0 \forall c$ . Therefore,  $\sum_{c=1}^C \tau_{C_c}(x) = 0$ . In pattern classification it can happen when the input pattern is not from any of the existing classes. On the otherhand, when the input pattern belongs to all the classes with fuzzy membership value 1,

$$\begin{aligned}
&\sum_{c=1}^C \tau_{C_c}^p(x) \\
&= \frac{1}{\hat{H}} \sum_{j=1}^{\hat{H}} \mu_{F_j}(x) \frac{\sum_y \sum_{c=1}^C \min\{\mu_{F_j}(y), \mu_{C_c}(y)\}}{\sum_y \mu_{F_j}(y)} \\
&= \frac{1}{\hat{H}} \sum_{j=1}^{\hat{H}} \mu_{F_j}(x) \frac{\sum_y \sum_{c=1}^C \min\{\mu_{F_j}(y), 1\}}{\sum_y \mu_{F_j}(y)} \\
&= \frac{1}{\hat{H}} \sum_{j=1}^{\hat{H}} \mu_{F_j}(x) \frac{C \sum_y \mu_{F_j}(y)}{\sum_y \mu_{F_j}(y)} \\
&= \frac{C}{\hat{H}} \sum_{j=1}^{\hat{H}} \mu_{F_j}(x) \quad (11)
\end{aligned}$$

Therefore, if the input pattern belongs to all the clusters completely, then  $\sum_{c=1}^C \tau_{C_c}^p(x)$  attains the maximum value  $C$ . ■

*Property 13:* When the clusters and the output classes are crisp,

$$\underline{R}(C_c) = \{x \in X \mid \tau_{C_c}(x) = 1\} \quad (12-a)$$

$$\bar{R}(C_c) = \{x \in X \mid \tau_{C_c}(x) > 0\} \quad (12-b)$$

$$\begin{aligned}
BN(C_c) &= \bar{R}(C_c) - \underline{R}(C_c) \\
&= \{x \in X \mid 0 < \tau_{C_c}(x) < 1\} \quad (12-c)
\end{aligned}$$

*Proof:* For the crisp clusters and the crisp output classes, the above results come directly from the definitions (1-a) and (1-b). ■

Following are a few trivial but interesting definitions based on the above properties.

1. A  $C$ -class classification problem for a set of input patterns  $\{x_1, x_2, \dots, x_n\}$  can be looked at as an assignment of the fuzzy-rough membership value  $\tau_{C_c}(x_i)$  on each  $x_i \in X$ ,  $\forall c = 1, 2, \dots, C$ ,  $\forall i = 1, 2, \dots, n$ . In fuzzy-rough context,  $C$  partitions of  $X$  are the set of values  $\{\tau_{C_c}(x_i)\}$ , that can be conveniently arranged on a  $C \times n$  matrix  $S = [\mu_{C_c}(x_i)]$ . Based on the characteristic of  $S$ , classification can be of the following three types [11]:

(a) *Crisp classification:*

$$M_{hc} = \left\{ S \in \mathfrak{R}^{Cn} \mid \tau_{C_c}(x_i) \in \{0, 1\} \forall c, \forall i; \right.$$

$$\left. \sum_{c=1}^C \tau_{C_c}(x_i) = 1; 0 < \sum_{i=1}^n \tau_{C_c}(x_i) < n \forall c \right\} \quad (13-a)$$

(b) *Constrained fuzzy-rough classification:*

$$M_{fc} = \left\{ S \in \mathfrak{R}^{Cn} \mid \tau_{C_c}(x_i) \in [0, 1] \forall c, \forall i; \right.$$

$$\left. \sum_{c=1}^C \tau_{C_c}(x_i) = 1; 0 < \sum_{i=1}^n \tau_{C_c}(x_i) < n \forall c \right\} \quad (13-b)$$

(c) *Possibilistic fuzzy-rough classification:*

$$M_{pc} = \left\{ S \in \mathfrak{R}^{Cn} \mid \tau_{C_c}(x_i) \in [0, 1] \forall c, \forall i; \right.$$

$$\left. 0 < \sum_{i=1}^n \tau_{C_c}(x_i) < n \forall c \right\} \quad (13-c)$$

From the relations (13-a), (13-b) and (13-c), it is obvious that  $M_{hc} \subset M_{fc} \subset M_{pc}$ .

2. To make the definition of rough approximation loose, the model of variable precision rough set model is proposed by Ziarko [24]. The idea of the variable precision model can further be generalized by the following way: The following approximation can be defined for  $\beta \in [0, 0.5]$

$$\underline{R}_\beta(C_c) = \{x \in X \mid \tau_{C_c}(x) \geq 1 - \beta\} \quad (14-a)$$

$$\bar{R}_\beta(C_c) = \{x \in X \mid \tau_{C_c}(x) > \beta\} \quad (14-b)$$

When  $\beta = 0$ , we obtain the equations (12-a) and (12-b). Similarly, for  $\beta \in [0.5, 1]$

$$\underline{R}_\beta(C_c) = \{x \in X \mid \tau_{C_c}(x) \geq \beta\} \quad (15-a)$$

$$\bar{R}_\beta(C_c) = \{x \in X \mid \tau_{C_c}(x) > 1 - \beta\} \quad (15-b)$$

- When  $\beta = 1$ , we obtain the equations (12-a) and (12-b).
3. The  $\tau$ -fuzzy-rough inclusion of  $A \subseteq X$  into  $B \subseteq X$  can be defined as

$$A \subseteq_\tau B \text{ if and only if } \tau_A(x) \leq \tau_B(x) \quad \forall x \in X \quad (16)$$

If the clusters and the output classes  $A$  and  $B$  are crisp, and  $A \subseteq_\tau B$ , then it can be shown that  $\underline{R}(A) \subseteq \underline{R}(B)$  and

$\bar{R}(A) \subseteq \bar{R}(B)$ . However, when the clusters or the output classes are not crisp, the above relationship may or may not hold.

4. From the definition of the fuzzy-rough membership functions, we can group the fuzzy-rough sets into the following four categories:

- (a)  $X$  is *partially  $R$ -observable in a fuzzy-rough manner*, if  $C_c \subseteq X$ ,  $\exists x \in X | \tau_{C_c}(x) = 0$  and  $\exists y \in X | \tau_{C_c}(y) = 1$ . It means that for some element of  $X$  we can decide whether it belongs to  $C_c$  or not.
- (b)  $X$  is *internally  $R$ -unobservable in a fuzzy-rough manner*, if  $C_c \subseteq X$ ,  $\exists x \in X | \tau_{C_c}(x) = 0$  and  $\exists y \in X | \tau_{C_c}(y) = 1$ . It implies that for certain element of  $X$  we can decide that it does not belong to  $C_c$ ; however, for all elements of  $X$ , we cannot decide whether they belong to  $C_c$ .
- (c)  $X$  is *externally  $R$ -observable in a fuzzy-rough manner*, if  $C_c \subseteq X$ ,  $\exists x \in X | \tau_{C_c}(x) = 0$  and  $\exists y \in X | \tau_{C_c}(y) = 1$ . It states that for all elements of  $X$  we cannot decide that they do not belong to  $C_c$ ; however, for certain element of  $X$ , we can decide whether it belongs to  $C_c$ .
- (d)  $X$  is *totally  $R$ -unobservable in a fuzzy-rough manner*, if  $C_c \subseteq X$ ,  $\exists x | \tau_{C_c}(x) = 0$  and  $\exists y | \tau_{C_c}(y) = 1$ . It means that for all elements of  $X$  we cannot decide whether they belong to  $C_c$  or not.

These groups are, in fact, the generalization of the groups reported in [16] [19].

### C. How to Determine the Fuzzy-Rough Membership Values?

Determination of the fuzzy-rough membership values depend on the fuzzy clustering of the input data set. The fuzziness in the clusters represent the fuzzy linguistic uncertainty present in the input data set. The clustering can be performed by

1. *Unsupervised Clustering*: It involves collecting the data from all the classes, and cluster them subsequently without considering the associated class labels with the data.
2. *Supervised Clustering*: Separate data sets are formed for each class, and clustering is performed on each such data set to find the subgroups present in the data from the same class.

Both the clustering can be done by fuzzy  $K$ -means clustering algorithm [1]. But, the problems with it are: a) the number of clusters have to be fixed *a priori*, which may not be known, b) it will not work in case the number of cluster is one, and c) generated fuzzy memberships are not possibilistic. To overcome the first problem, the evolutionary programming-based method used in [18] can be used. For various clustering problems, this method can automatically determine the number of clusters. Other cluster validity measures [4] can also be used along with the fuzzy  $K$ -means to determine the number of clusters. It is to be noted that the number of the clusters should be determined as best as we can. Otherwise, the calculation of the fuzzy linguistic variables will be different, and as a result, the fuzzy-rough membership values may also vary. For the second problem, if we know *a priori* that only one cluster is present for a class, then we can find the mean and standard deviation from the input data set, and we can fit a ' $\pi$ ' fuzzy membership curve [13]. But, while doing so we must take care of the possible presence of the outliers in the input data set. To overcome the third problem, the possibilistic fuzzy  $K$ -means clustering algorithm [8] or mixed  $K$ -means clustering algorithm [12] can be used. As of now, there is no single clustering algorithm which can solve all the above problems.

If the output class is fuzzy, then it may be possible to assign the fuzzy memberships for the output class subjectively. For instance, if the output class for an image processing application is represented by BRIGHT, then it is possible to fit an 'S' fuzzy membership curve [13], and assign the fuzzy membership values for  $\mu_{\text{BRIGHT}}(x)$ . However, if the domain specific knowledge is absent, then we have to be satisfied with the given crisp membership values.

### D. Measures of Fuzzy-Rough Ambiguity

The fuzzy-rough ambiguity plays a critical role in many classification problems as it is capable of modeling nonstatistical uncertainty. Consequently, characterisation and quantification of fuzzy-roughness are important issues, that affect the management of uncertainty in many classifier design. Hence, measures of fuzzy-roughness are essential to estimate the *average* ambiguity in an output class in some well-defined sense. A measure of fuzzy-roughness for a discrete output class  $C_c \subseteq X$  is a mapping  $\mathcal{P}(X) \rightarrow \mathbb{R}^+$  that quantifies the degree of fuzzy-roughness present in  $C_c$ . Here  $\mathcal{P}(X)$  is the set of all fuzzy-rough power sets that can be defined within the universal set  $X$  (i.e.,  $\mathcal{P}(X)$  can be termed as *fuzzy-rough power set* of  $X$ ). The fuzzy-rough ambiguity of a set must be zero when there is no ambiguity in deciding whether an input pattern belongs to it or not. If the output class is maximally ambiguous, i.e.,  $\tau_{C_c}(x) = 0.5 \forall x$ , then the corresponding measure should be maximum. When the fuzzy-rough membership value of an input pattern approaches either 0 or 1, the uncertainty about the belongingness of the pattern in the output class decreases; hence the measure of fuzzy-roughness of the class should also decrease. A set  $C_c^*$  is called a *less ambiguous* version of  $C_c$  if the following conditions are satisfied:

$$\tau_{C_c^*}(x) \leq \tau_{C_c}(x) \quad \text{if} \quad \tau_{C_c}(x) \leq 0.5 \quad (17-a)$$

and

$$\tau_{C_c^*}(x) \geq \tau_{C_c}(x) \quad \text{if} \quad \tau_{C_c}(x) \geq 0.5 \quad (17-b)$$

For a less ambiguous version, the measure of fuzzy-roughness should decrease because the above two operations reduce the ambiguity. Therefore, such a measure (say  $\mathcal{H}$ ) should obey the following axioms:

- A1. *Certainty*:  $\mathcal{H}(C_c) = 0 \Leftrightarrow \tau_{C_c}(x) = 0$  or  $1 \forall x \in X$ .
- A2. *Maximality*:  $\mathcal{H}(C_c)$  is maximum  $\Leftrightarrow \tau_{C_c}(x) = 0.5 \forall x \in X$ .
- A3. *Resolution*:  $\mathcal{H}(C_c) \geq \mathcal{H}(C_c^*)$ , where  $C_c^*$  is a less ambiguous version of  $C_c$ .

In case of the constrained fuzzy-rough membership, another desirable property is that the fuzzy-roughness measure of a set and its compliment should be equal. Hence, the fourth axiom may be

$$A4. \textit{Symmetry}: \mathcal{H}(C_c) = \mathcal{H}(1 - C_c), \text{ where } \tau_{1-C_c}^f(x) = 1 - \tau_{C_c}^f(x) \forall x \in X.$$

Since  $\tau_{1-C_c}^p(x) = \frac{1}{H} \sum_{j=1}^H \mu_{F_j}(x) - \tau_{C_c}^p(x) \forall x \in X$ , the above property does not necessarily hold for the possibilistic fuzzy-rough membership function.

One such possible measure, which satisfies all the above axioms, is *index of fuzzy-roughness*. It is defined as

$$\mathcal{H}_I(C_c) = \frac{2}{n^k} d(C_c, C_c^{\text{near}}) \quad (18)$$

where  $n$  is the number of input patterns,  $k \in \mathbb{R}^+$ ,  $d$  is a distance measure and  $C_c^{\text{near}}$  is a set without fuzzy-rough ambiguity that is the nearest one to  $C_c$ . For  $C_c^{\text{near}}$ , the fuzzy-rough membership function is defined as

$$\tau_{C_c^{\text{near}}}(x) = \begin{cases} 1 & \text{if } \tau_{C_c}(x) \geq 0.5 \\ 0 & \text{otherwise} \end{cases}$$

The value of  $k$  depends on  $d$ . For the (Minkowski)  $q$  norms,  $d(C_c, C_c^{\text{near}})$  and  $\mathcal{H}_I(C_c)$  take the following forms:

$$d(C_c, C_c^{\text{near}}) = \left[ \sum_{x \in C_c} |\tau_{C_c}(x) - \tau_{C_c^{\text{near}}}(x)|^q \right]^{1/q} \quad (19)$$

and

$$\mathcal{H}_I(C_c) = \frac{2}{n^{1/q}} \left[ \sum_{x \in C_c} |\tau_{C_c}(x) - \tau_{C_c^{\text{near}}}(x)|^q \right]^{1/q} \quad (20)$$

where  $q \in [1, \infty)$ . For  $q = 1$ ,  $\mathcal{H}_I(C_c)$  is called *linear index of fuzzy-roughness*, and for  $q = 2$ ,  $\mathcal{H}_I(C_c)$  is known as *quadratic index of fuzzy-roughness*.

Another possible measure is *fuzzy-rough entropy*, which can be defined as follows:

$$\mathcal{H}_E = -K \sum_{x \in X} \left[ \tau_{C_c}(x) \log(\tau_{C_c}(x)) + (1 - \tau_{C_c}(x)) \log(1 - \tau_{C_c}(x)) \right] \quad (21)$$

The term  $\log$  denotes the logarithm to any base  $a > 1$ . It can be normalized by adjusting the value of  $K$ , so that it satisfies  $A1$  to  $A4$ . From the property 3, when the clusters are crisp,  $\mathcal{H}_E$  simplifies to the following rough-fuzzy entropy [19]:

$$\mathcal{H}_E = -K \sum_{x \in X} \left[ \iota_{C_c}(x) \log(\iota_{C_c}(x)) + (1 - \iota_{C_c}(x)) \log(1 - \iota_{C_c}(x)) \right] \quad (22)$$

From the property 5, when the clusters are crisp and *fine*,  $\mathcal{H}_E$  is reduced to the fuzzy entropy [10] as follows:

$$\mathcal{H}_E = -K \sum_{x \in X} \left[ \mu_{C_c}(x) \log(\mu_{C_c}(x)) + (1 - \mu_{C_c}(x)) \log(1 - \mu_{C_c}(x)) \right] \quad (23)$$

In absence of fuzziness (property 4),  $\mathcal{H}_E$  is simplified to the rough entropy as follows [17]:

$$\mathcal{H}_E = -K \sum_{x \in X} \left[ r_{C_c}(x) \log(r_{C_c}(x)) + (1 - r_{C_c}(x)) \log(1 - r_{C_c}(x)) \right] \quad (24)$$

Similarly, there are following measures like *multiplicative* and *additive* measures, which can be directly borrowed from the fuzzy set theory [10]. All these measures satisfy  $A1$ - $A4$ .

*Multiplicative measure of fuzzy-roughness*: Any function  $\mathcal{H}_M : \mathcal{P}(X) \rightarrow \mathbb{R}^+$  is a multiplicative fuzzy-rough measure if it can be written as

$$\mathcal{H}_M(C_c) = K \sum_{x \in X} g(\tau_{C_c}(x)) \quad K \in \mathbb{R}^+ \quad (25)$$

where

$$g(t) = \hat{g}(t) - \min_{0 \leq t \leq 1} \{\hat{g}(t)\} \quad (26)$$

$$\hat{g}(t) = f(t)f(1-t) \quad (27)$$

and  $f : [0, 1] \rightarrow \mathbb{R}^+$  is a concave increasing function. By taking  $f(t) = t \exp(1-t)$ , it can be shown that

$K \sum_{x \in X} [\tau_{C_c}(x)(1 - \tau_{C_c}(x))]$  is an examples of this kind of measure [10].

*Additive measure of fuzzy-roughness*: Any function  $\mathcal{H}_A : \mathcal{P}(X) \rightarrow \mathbb{R}^+$  is an additive fuzzy-rough measure, provided that it can be written as

$$\mathcal{H}_A(C_c) = K \sum_{x \in X} g(\tau_{C_c}(x)) \quad K \in \mathbb{R}^+ \quad (28)$$

where

$$g(t) = \hat{g}(t) - \min_{0 \leq t \leq 1} \{\hat{g}(t)\} \quad (29)$$

$$\hat{g}(t) = f(t) + f(1-t) \quad (30)$$

and  $f : [0, 1] \rightarrow \mathbb{R}^+$  is an increasing function. If we take  $f(t) = t \exp(1-t)$ , then it is possible to show that one example of this kind of measure is [10]

$$K \sum_{x \in X} [\tau_{C_c}(x) \exp(1 - \tau_{C_c}(x)) + (1 - \tau_{C_c}(x)) \exp(\tau_{C_c}(x)) - 1].$$

## REFERENCES

- [1] J. C. Besdek. *Pattern Recognition with Fuzzy Objective Function Algorithms*. Plenum Press, New York, 1981.
- [2] J. C. Besdek. The thirsty traveller visits gamont: A rejoinder to "comments on fuzzy set - what are they and why?". *IEEE Transactions on Fuzzy Systems*, 2(1):43-45, February 1994.
- [3] J. C. Besdek and S. K. Pal. *Fuzzy Models for Pattern Recognition*, Eds. IEEE Press, New York, 1992.
- [4] R. Dubes and A. Jain. *Algorithms that Cluster Data*. Prentice Hall, Englewood Cliffs, NJ, 1987.
- [5] D. Dubois and H. Prade. Rough-fuzzy sets and fuzzy-rough sets. *International Journal of General Systems*, 17(2-3):191-209, 1990.
- [6] D. Dubois and H. Prade. Putting rough sets and fuzzy sets together. In R. Slowinski, editor, *Intelligent Decision Support. Handbook of Applications and Advances of the Rough Set Theory*. Kluwer Academic Publishers, Dordrecht, 1992.
- [7] G. S. Klir and B. Yuan. *Fuzzy sets and Fuzzy Logic - Theory and Applications*. Prentice-Hall, Englewood Cliffs, NJ, 1995.
- [8] R. Krishnapuram and J. M. Keller. A possibilistic approach to clustering. *IEEE Transactions on Fuzzy Systems*, 1(2):98-110, May 1993.
- [9] C. T. Lin and C. S. G. Lee. *Neural Fuzzy Systems*. Prentice Hall, Englewood Cliffs, New Jersey, 1996.
- [10] N. R. Pal and J. C. Besdek. Measuring fuzzy entropy. *IEEE Transactions on Fuzzy Systems*, 2(2):107-118, May 1994.
- [11] N. R. Pal and J. C. Besdek. On cluster validity for the fuzzy C-means model. *IEEE Transactions on Fuzzy Systems*, 3(3):330-379, August 1995.
- [12] N. R. Pal, K. Pal, and J. C. Besdek. A mixed c-means clustering model. In *Proceedings of IEEE International Conference on Fuzzy Systems (Barcelona, Spain)*, pages 11-21, July 1997.
- [13] S. K. Pal and D. Dutta Majumder. *Fuzzy Mathematical Approach to Pattern Recognition*. Wiley (Halsted Press), New York, 1986.
- [14] Z. Pawlak. Rough sets. *International Journal of Computer and Information Science*, 11:341-356, 1982.
- [15] Z. Pawlak. *Rough Sets: Theoretical Aspects of Reasoning About Data*. Kluwer, Dordrecht, 1991.
- [16] Z. Pawlak. Vagueness and uncertainty: A rough set perspective. Technical Report ICS Research Report 19/94, Institute of Computer Science, Warsaw University of Technology, Warsaw, Poland, March 1994.
- [17] Z. Pawlak, S. K. M. Wong, and W. Ziarko. Rough sets: Probabilistic versus deterministic approach. *International Journal of Man-Machine Studies*, 29:81-95, 1988.
- [18] M. Sarkar. Evolutionary programming based fuzzy clustering. In *Proceedings of Fifth Annual Conference on Evolutionary Programming (San Diego)*, pages 247-256, 1996.
- [19] M. Sarkar and B. Yegnanarayana. Rough-fuzzy membership functions in classification. Accepted in *Fuzzy Sets and Systems*.
- [20] M. Sarkar and B. Yegnanarayana. Rough-fuzzy membership functions. In *Proceedings of IEEE International Conference on Fuzzy Systems (Anchorage, Alaska, USA)*, May 4-9 1998.
- [21] G. Shafer. *A Mathematical Theory of Evidence*. Princeton University Press, Princeton, 1976.
- [22] S. K. M. Wong and W. Ziarko. Comparison of the probabilistic approximate classification and fuzzy set model. *Fuzzy Sets and Systems*, 21:357-362, 1987.
- [23] L. A. Zadeh. Fuzzy sets as a basis for a theory of possibility. *Fuzzy Sets and Systems*, 1:3-28, 1978.
- [24] W. Ziarko. Variable rough set model. *Journal of Computer and Systems Sciences*, 40(1):39-59, 1993.