ELSEVIER

Fuzzy Sets and Systems 132 (2002) 353–369

FUZZY sets and systems

www.elsevier.com/locate/fss

Rough–fuzzy functions in classification

Manish Sarkar*

Department of Computer Science & Engineering, 191 Auditorium Road, U-155, University of Connecticut, CT 06269-3155 Storrs, USA

Received 21 October 1997; received in revised form 11 January 2002; accepted 24 January 2002

Abstract

This paper generalizes the concept of rough membership functions in pattern classification tasks to rough–fuzzy membership functions and rough–fuzzy ownership functions. Unlike the rough membership value of a pattern, which is sensitive only towards the rough uncertainty associated with the pattern, the rough–fuzzy membership (or ownership) value of the pattern signifies the rough uncertainty as well as the fuzzy uncertainty associated with the pattern. In this paper, various set theoretic properties of the rough–fuzzy functions are exploited to characterize the concept of rough–fuzzy sets. These properties are also used to measure the rough–fuzzy uncertainty associated with the given output class. Finally, a few possible applications of the rough–fuzzy functions are mentioned.

Keywords: Fuzzy sets; Rough sets; Rough–fuzzy sets; Membership functions; Classification and entropy

1. Introduction

In any classification task the aim is to form various classes where each class contains the objects that are not noticeably different [1]. 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, then it is impossible for anyone to correctly classify these two red triangles using the given two attributes. This kind of uncertainty is referred to as *rough uncertainty* [10,19,23,24,26,27,32]. The rough uncertainty is formulated in terms of *rough sets* [20]. 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 since our knowledge about the system generating the data is limited [30]. Therefore, rough sets are essential to deal with a classification system where we do not have complete knowledge of the system.

Fuzzy sets, a generalization of the classical sets, are considered as mathematical tools to model the vagueness present in the human classification mechanism. In the classical set theory, the belongingness of an element to the given universe 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* using *yes*/*no*

* Tel.: +1-860-486-2584; fax: +1-860-486-4817.
E-mail address: manish_sarkar@hotmail.com (M. Sarkar).

0165-0114/02/$ - see front matter
PII: S0165-0114(02)00119-7

category, since there is no 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 [18].

Although probabilistic, fuzzy and rough uncertainties are different facets of uncertainty, they are often confused. The fuzziness deals with the vagueness between the overlapping sets [2,4,11], while the concept of probability concerns the likelihood of randomness of a phenomenon [14]. 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 into a particular class. However, rough sets assert that all the members of the same granule may not have the same membership values into a particular class. Fuzzy sets deal with overlapping classes and *fine* concepts, whereas rough sets deal with nonoverlapping classes and coarse concepts.

In a classification task, both roughness and fuzziness can co-exist [7,8,9]. In some kinds of problems, the indiscernibility relation partitions the input pattern set to form several equivalence classes. These equivalence classes act like granules. In other kinds of problems, the input pattern set is transformed into a set of granules after imposing certain constrains on the interpretation of the structure of the data. In both kinds of problems, the granules try to approximate the given output class. When this approximation is not proper, the roughness appears. In addition, the concepts or the output classes may have ill-defined boundaries. Thus, both roughness and fuzziness appear here due to the indiscernibility relation in the input pattern set and the vagueness in the output class, respectively. To model this type of situation, where both vagueness and approximation are present, the concept of *rough–fuzzy set* [5] is introduced. The resultant model is expected to be more powerful than both rough sets and fuzzy sets.

In the classical set theory, one of the fundamental notions is *characteristic function*. When one considers subsets of a given universe based on the available partial information, it is possible to apply the characteristic function to determine whether a given element belongs to a particular set. Following this direction, the concept of *rough–fuzzy membership function* and *rough–fuzzy ownership function* are proposed in this paper. The rough–fuzzy membership function attempts to quantify the roughness that appears mainly due to the limitation of the representation; in contrast, the rough–fuzzy ownership function measures the roughness that arises primarily due to our interpretation of the structure of the input set. In absence of the fuzziness in the output class, the rough–fuzzy membership function (rough–fuzzy ownership function) reduces to the original rough membership function (rough ownership function). Moreover, when the partition in the input set is *fine*, i.e., each equivalence class contains only one pattern, both rough–fuzzy membership function and rough–fuzzy ownership function reduce to the fuzzy membership function. Similarly, if the partitioning is *fine* and the output classes are crisp simultaneously, the proposed functions reduce to the characteristic function. The concepts of the rough–fuzzy functions become particularly attractive when we do not have complete knowledge of the human classification system, but we attempt to mimic the vagueness present in the human reasoning.

The paper is organized as follows: In Section 2 we discuss the basics of rough sets, fuzzy sets and rough–fuzzy sets. In Section 3, the rough–fuzzy membership function and its properties are described. Section 4 discusses the rough–fuzzy ownership functions. Section 5 outlines how to quantify the rough–fuzzy uncertainty using the proposed functions. Section 6 demonstrates some possible applications of the proposed functions.

2. Background of rough and fuzzy sets

2.1. Rough sets

A binary relation R on a universal set X is an equivalence relation if and only if

(1) R is *reflexive*, i.e., $\boldsymbol{x}$ is related to itself or $\boldsymbol{x}R\boldsymbol{x}$ where $\boldsymbol{x}\in X$,
(2) R is *symmetric*, i.e., $\boldsymbol{x}R\boldsymbol{y} \Rightarrow \boldsymbol{y}R\boldsymbol{x}$ where $\boldsymbol{x},\boldsymbol{y}\in X$,

(3) R is *transitive*, i.e., $\boldsymbol{x}R\boldsymbol{y}$ and $\boldsymbol{y}R\boldsymbol{z} \Rightarrow \boldsymbol{x}R\boldsymbol{z}$ where $\boldsymbol{x},\boldsymbol{y},\boldsymbol{z} \in X$.

For an equivalence relation R on a set X, the set of the elements of X that are related to $\boldsymbol{x} \in X$, is called the equivalence class of $\boldsymbol{x}$, and it is denoted by $[\boldsymbol{x}]_R$. Moreover, let X/R denote the family of all equivalence classes induced on X by R. For any output class $\mathscr{C} \subseteq X$, we can define the lower $\underline{R}(\mathscr{C})$ and upper $\bar{R}(\mathscr{C})$ approximations, which approach $\mathscr{C}$ as closely as possible from the inside and outside, respectively [12]. Here,

$$\underline{R}(\mathscr{C}) = \bigcap \{[\boldsymbol{x}]_R \mid [\boldsymbol{x}]_R \subseteq \mathscr{C} \text{ and } \boldsymbol{x} \in X\} \tag{1}$$

is the union of all equivalence classes in X/R that are contained in $\mathscr{C}$ and

$$\bar{R}(\mathscr{C}) = \bigcup \{[\boldsymbol{x}]_R \mid [\boldsymbol{x}]_R \cap \mathscr{C} \neq \emptyset \text{ and } \boldsymbol{x} \in X\} \tag{2}$$

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

Two examples of rough sets are shown in Fig. 1. In the first example (Fig. 1(a)), X is a closed interval of real numbers, and X/R partitions X into ten semiclosed intervals. The output class $\mathscr{C}_c$, which is to be approximated by the elements of X/R, is shown as the closed interval. We use the subscript c to indicate the cth output class. The rough set approximation of $\mathscr{C}_c$ consists of the two semiclosed intervals $\underline{R}(\mathscr{C}_c)$ and $\bar{R}(\mathscr{C}_c)$. In the second example (Fig. 1(b)), the universal set is $X = X_1 \times X_2$, and the equivalence relation R partitions $X_1 \times X_2$ into 100 small squares.

The *rough membership function* $r_{\mathscr{C}_c}(\boldsymbol{x}) : X \to [0,1]$ of a pattern $\boldsymbol{x} \in X$ in the output class $\mathscr{C}_c$ is defined in [22,31] by

$$r_{\mathscr{C}_c}(\boldsymbol{x}) = \frac{\|[\boldsymbol{x}]_R \cap \mathscr{C}_c\|}{\|[\boldsymbol{x}]_R\|}, \tag{3}$$

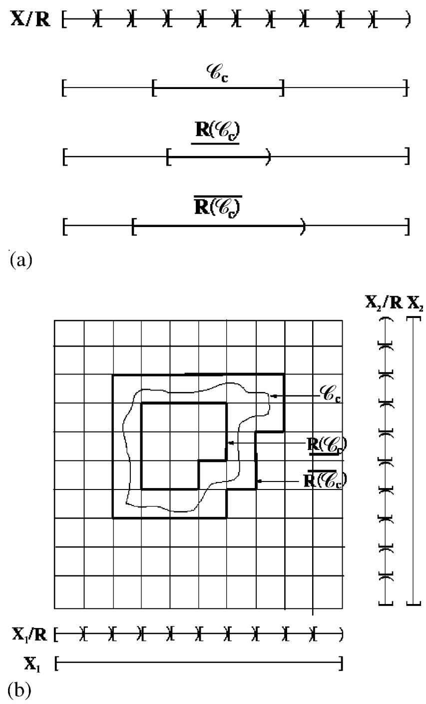

Fig. 1. (a) Rough sets in one-dimensional domain. The equivalence relation R partitions the universal set X into ten intervals. The output class $\mathscr{C}_c$ is approximated by $\underline{R}(\mathscr{C}_c)$ and $\bar{R}(\mathscr{C}_c)$. (b) Rough sets in two-dimensional domain. The equivalence relation R partitions the universal set $X_1 \times X_2$ into 100 small squares. In both the cases, the uncertainty is generated in the intervals/squares that are covered by $\bar{R}(\mathscr{C}_c)$ but not by $\underline{R}(\mathscr{C}_c)$.

where $\|[\boldsymbol{x}]_R\|$ denotes the cardinality of the crisp set $\mathscr{C}_c$.

2.2. Fuzzy sets

In traditional two-state classifiers, where a class $\mathscr{C}$ is defined as a subset of the universal set X, any input pattern $\boldsymbol{x} \in X$ can either be a member or not be a member of the given class $\mathscr{C}$. This property of whether or not a pattern $\boldsymbol{x}$ of the universal set belongs to the class $\mathscr{C}$ can be defined by a characteristic function $\mu_{\mathscr{C}} : X \to \{0,1\}$ as follows [12]: $\mu_{\mathscr{C}}(\boldsymbol{x}) = 1$ *iff* $\boldsymbol{x} \in \mathscr{C}$, $\mu_{\mathscr{C}}(\boldsymbol{x}) = 0$ otherwise. In real life situations, boundaries between the classes may be overlapping (Fig. 2). Hence, it is uncertain whether an input pattern belongs totally to the class $\mathscr{C}$. To consider such situations, in fuzzy sets the concept of the characteristic function

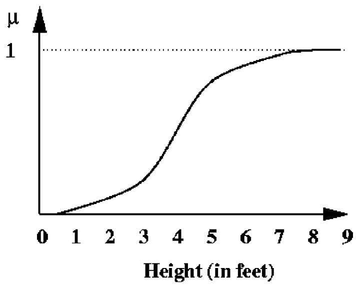

Fig. 2. The fuzzy membership function for the fuzzy set *tall*. There is no single point c on the X-axis such that a person with $height \geqslant c$ can be called *tall* and a person with $height < c$ can be called *not tall*.

has been modified to the *fuzzy membership function* $\mu_{\mathscr{C}} : X \to [0,1]$.

2.3. Rough–fuzzy sets

The *rough–fuzzy set* is the generalization of the rough set in the sense that here the output class is fuzzy. Let X be a set, R be an equivalence relation defined on X, and the output class $\mathscr{C} \subseteq X$ be a fuzzy set. The rough–fuzzy set is a tuple $\langle \underline{R}(\mathscr{C}), \bar{R}(\mathscr{C}) \rangle$, where the lower approximation $\underline{R}(\mathscr{C})$ and the upper approximation $\bar{R}(\mathscr{C})$ are fuzzy sets of X/R, with membership functions defined in [5,6] by

$$\mu_{\underline{R}(\mathscr{C})}([\boldsymbol{x}]_R) = \inf\{\mu_{\mathscr{C}}(\boldsymbol{x}) \mid \boldsymbol{x} \in [\boldsymbol{x}]_R\} \quad \forall \boldsymbol{x} \in X \tag{4}$$

and

$$\mu_{\bar{R}(\mathscr{C})}([\boldsymbol{x}]_R) = \sup\{\mu_{\mathscr{C}}(\boldsymbol{x}) \mid \boldsymbol{x} \in [\boldsymbol{x}]_R\} \quad \forall \boldsymbol{x} \in X. \tag{5}$$

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

3. Rough–fuzzy membership functions

3.1. Definition of rough–fuzzy membership functions

The rough–fuzzy membership function of a pattern $\boldsymbol{x} \in X$ for the fuzzy output class $\mathscr{C}_c \subseteq X$ is defined as

$$\iota_{\mathscr{C}_c}(\boldsymbol{x}) = \frac{\|F \cap \mathscr{C}_c\|}{\|F\|}, \tag{6}$$

where $F = [\boldsymbol{x}]_R$, and $\|\mathscr{C}_c\|$ implies the cardinality of the fuzzy set $\mathscr{C}_c$. One possible way to determine the cardinality [34] is to use $\|\mathscr{C}_c\| = \sum_{\boldsymbol{x} \in X} \mu_{\mathscr{C}_c}(\boldsymbol{x})$. For the '$\cap$' (intersection) operation, we can use $\mu_{A \cap B}(\boldsymbol{x}) = \min\{\mu_A(\boldsymbol{x}), \mu_B(\boldsymbol{x})\}\ \forall \boldsymbol{x} \in X$.

In Fig. 3, the idea of the rough–fuzzy membership function is depicted. The parallelepiped contains all the patterns from the equivalence class $[\boldsymbol{x}_i]$ i.e., the parallelepiped contains the patterns that look like $[\boldsymbol{x}_i]$. But, some of these patterns may have different fuzzy memberships into the class $\mathscr{C}_c$. It creates the rough uncertainty. If the fuzzy membership surface is flat when it passes through the parallelepiped, then the fuzzy memberships into $\mathscr{C}_c$ for all the patterns inside the parallelepiped are same. Hence, in this case, there is no rough uncertainty. However, there will be certain amount of fuzziness if the fuzzy membership values inside the parallelepiped lie in $(0,1)$. Intuitively, the rough–fuzzy membership value of $[\boldsymbol{x}_i]$ is the volume occupied by the overlapped space in the parallelepiped divided by the volume of the complete parallelepiped. The volume of the overlapped space is approximated by the weighted number of patterns in the space, where the weight of each pattern is quantified by its fuzzy membership value.

3.2. Properties of rough–fuzzy membership functions

Following are a few important properties of the rough–fuzzy membership functions that can be exploited in a classification task.

Property 1. $0 \leqslant \mu_{\mathscr{C}_c}(\boldsymbol{x}) \leqslant 1$.

Proof. Since $\emptyset \subseteq F \cap \mathscr{C}_c \subseteq F$, the proof is trivial. □

The above property is obvious from Fig. 3 because the rough–fuzzy membership function achieves the maximum value when the overlapped space covers the whole parallelepiped. Similarly, the minimum value is achieved when the overlapped space does not exist.

Property 2. $\iota_{\mathscr{C}_c}(\boldsymbol{x}) = 1$ and 0 *if and only if no rough–fuzzy uncertainty is associated with the pattern* $\boldsymbol{x}$.

Proof. *If part*: If no rough–fuzzy uncertainty is involved, then either (a) $F \cap \mathscr{C}_c \subseteq F$, i.e., $\iota_{\mathscr{C}_c}(\boldsymbol{x}) = 1$, or (b) $F \cap \mathscr{C}_c = \emptyset$ i.e., $\iota_{\mathscr{C}_c}(\boldsymbol{x}) = 0$.

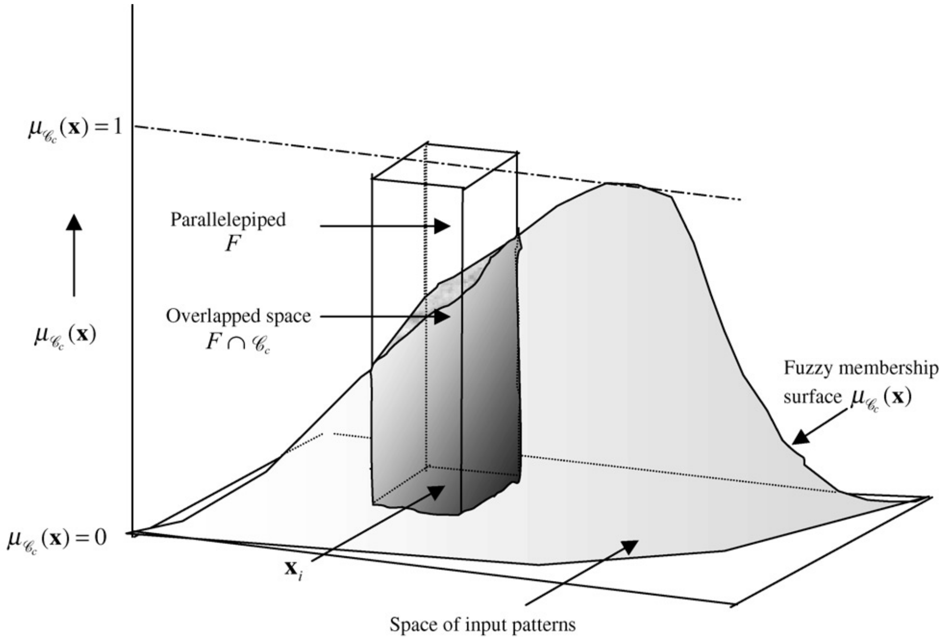

Fig. 3. Intuitive view of the rough–fuzzy membership functions. The parallelepiped contains all (and only) the patterns that have the same input representation of $\boldsymbol{x}_i$. The roughness is created in the parallelepiped when it contains more than one pattern and when the patterns have different fuzzy membership values. The fuzziness is appearing in the parallelepiped when the fuzzy membership values are in $(0,1)$. The presence of both roughness and fuzziness create rough–fuzziness. Intuitively, the rough–fuzzy membership of the pattern $\boldsymbol{x}_i$ is the volume occupied by the overlapped space divided by the volume of the complete parallelepiped.

Only if part: If $\iota_{\mathscr{C}_c}(\boldsymbol{x})=0$, then the numerator of (6) is zero. It implies that $F\cap\mathscr{C}_c=\emptyset$. On the other hand, if $\iota_{\mathscr{C}_c}(\boldsymbol{x})=1$, then the numerator of (6) is equal to the denominator. It means that $F\cap\mathscr{C}_c=\mathscr{C}_c$, i.e., $F\subseteq\mathscr{C}_c$. Both cases imply that no rough–fuzzy uncertainty is involved. □

From Fig. 3, when no roughness is involved, the fuzzy membership surface inside the parallelepiped should be flat. In addition, to avoid the fuzziness, the fuzzy membership values inside the parallelepiped should be either zero or one. In that case, the overlapped space in the parallelepiped must be equal to zero, or the overlapped space must cover the whole parallelepiped. Hence, the rough–fuzzy membership value should be either zero or one. The reverse proof is also intuitive from Fig. 3.

Property 3. *When the output class $\mathscr{C}_c$ is crisp, $\iota_{\mathscr{C}_c}(\boldsymbol{x})=r_{\mathscr{C}_c}(\boldsymbol{x})$.*

Proof. When the output class $\mathscr{C}_c$ is crisp, Eq. (6) reduces to Eq. (3). Hence, the proof follows. □

When the output class is crisp, the fuzzy membership surface inside the parallelepiped can attain only the values zero or one (Fig. 4). Although there is no fuzziness, there may be certain amount of roughness if the fuzzy membership surface inside the parallelepiped attains both the values zero and one. As a result, the rough–fuzzy membership function becomes equivalent to the rough membership function.

Property 4. *When the partitioning is fine, $\iota_{\mathscr{C}_c}(\boldsymbol{x})=\mu_{\mathscr{C}_c}(\boldsymbol{x})$. Moreover, if the partitioning is fine and the output class $\mathscr{C}_c$ is crisp, then $\iota_{\mathscr{C}_c}(\boldsymbol{x})$ is equivalent to the characteristic function.*

Proof. When the partitioning is *fine*, i.e., each F consists of a single pattern, $\iota_{\mathscr{C}_c}(\boldsymbol{x})=\|1\,.\,\mu_{\mathscr{C}_c}(\boldsymbol{x})\|/1=\mu_{\mathscr{C}_c}(\boldsymbol{x})$. If $\mu_{\mathscr{C}_c}(\boldsymbol{x})=0$ or 1, i.e., the output class is crisp, then $\iota_{\mathscr{C}_c}(\boldsymbol{x})$ becomes the characteristic function. □

When the partitioning is *fine*, only one pattern can reside inside the parallelepiped (Fig. 3). Hence, no one-to-many mapping can exist inside the parallelepiped, and therefore, no roughness exists. However, the fuzziness may exist. The fuzziness too does not exist if the fuzzy membership of the pattern inside the parallelepiped is either zero or one.

This property and Property 3 show that both rough and fuzzy membership functions become particular cases of rough–fuzzy membership functions in the absence of the fuzziness and roughness, respectively.

Property 5. $\iota_{X-\mathscr{C}_c}(\boldsymbol{x})=1-\iota_{\mathscr{C}_c}(\boldsymbol{x})$.

Proof.

$$\iota_{X-\mathscr{C}_c}(\boldsymbol{x})=\frac{\|F\cap(X-\mathscr{C}_c)\|}{\|F\|}$$

$$=1-\frac{\|F\cap\mathscr{C}_c\|}{\|F\|}=1-\iota_{\mathscr{C}_c}(\boldsymbol{x}). \quad\square$$

Property 6. *If* $\boldsymbol{x}$ *and* $\boldsymbol{y}$ *are two patterns such that* $\boldsymbol{x}R\boldsymbol{y}$ (*i.e.*, $\boldsymbol{x},\boldsymbol{y}\in F$), *then* $\iota_{\mathscr{C}_c}(\boldsymbol{x})=\iota_{\mathscr{C}_c}(\boldsymbol{y})$.

Proof. It can be derived directly from Eq. (6). □

Property 7. $\iota_{A\cup B}(\boldsymbol{x})\geqslant\max\{\iota_A(\boldsymbol{x}),\iota_B(\boldsymbol{x})\}$ *where* $A,B\subseteq X$.

Proof.

$$\iota_{A\cup B}(\boldsymbol{x})=\frac{\|F\cup(A\cup B)\|}{\|F\|}$$

$$=\frac{\|(F\cup A)\cup(F\cup B)\|}{\|F\|}$$

$$\geqslant\frac{\|F\cup A\|}{\|F\|}=\iota_A(\boldsymbol{x}).$$

Similarly, $\iota_{A\cup B}(\boldsymbol{x})\geqslant\iota_B(\boldsymbol{x})$. □

Property 8. $\iota_{A\cap B}(\boldsymbol{x})\leqslant\min\{\iota_A(\boldsymbol{x}),\iota_B(\boldsymbol{x})\}$ *where* $A,B\subseteq X$.

Proof.

$$\iota_{A\cap B}(\boldsymbol{x})=\frac{\|F\cap(A\cap B)\|}{\|F\|}=\frac{\|(F\cap A)\cap(F\cap B)\|}{\|F\|}$$

$$\leqslant\frac{\|F\cap A\|}{\|F\|}=\iota_A(\boldsymbol{x}).$$

Similarly, $\iota_{A\cap B}(\boldsymbol{x})\leqslant\iota_B(\boldsymbol{x})$. □

Property 9. *If* Z *is a family of pairwise disjoint subsets of* X, *then* $\iota_{\cup Z}(\boldsymbol{x})=\sum_{\mathscr{C}_c\in Z}\iota_{\mathscr{C}_c}(\boldsymbol{x})$.

Proof.

$$\iota_{\cup Z}(\boldsymbol{x})=\frac{\|F\cap(\cup Z)\|}{\|F\|}=\frac{\|\cup(F\cap Z)\|}{\|F\|}$$

$$=\sum_{\mathscr{C}_c\in Z}\iota_{\mathscr{C}_c}(\boldsymbol{x}). \quad\square$$

Property 10. *For a C-class classification problem, the rough–fuzzy membership function of a pattern behaves in a possibilistic way provided the fuzzy membership function of the pattern to the output classes is possibilistic.*

Proof. We show that the sum of the membership values over the output classes may or may not be equal to one. That is

$$\sum_{c=1}^{C}\iota_{\mathscr{C}_c}(\boldsymbol{x})$$

$$=\sum_{c=1}^{C}\frac{\|F\cap\mathscr{C}_c\|}{\|F\|}$$

$$=\sum_{c=1}^{C}\frac{\sum_{\boldsymbol{x}\in X}\min\{\mu_F(\boldsymbol{x}),\mu_{\mathscr{C}_c}(\boldsymbol{x})\}}{\|F\|}$$

$$=\sum_{c=1}^{C}\frac{\sum_{\boldsymbol{x}\in F}\min\{1,\mu_{\mathscr{C}_c}(\boldsymbol{x})\}+\sum_{\boldsymbol{x}\notin F}\min\{0,\mu_{\mathscr{C}_c}(\boldsymbol{x})\}}{\|F\|}$$

$$=\sum_{c=1}^{C}\frac{\sum_{\boldsymbol{x}\in F}\mu_{\mathscr{C}_c}(\boldsymbol{x})}{\|F\|}$$

$$=\sum_{x\in F}\frac{\sum_{c=1}^{C}\mu_{\mathscr{C}_c}(\boldsymbol{x})}{\|F\|}.$$

Therefore, for the crisp and constrained fuzzy classification [3], where $\sum_{c=1}^{C}\mu_{\mathscr{C}_c}(\boldsymbol{x})=1$, the value of

$\sum_{c=1}^{C} \iota_{\mathscr{C}_c}(\boldsymbol{x})$ is equal to one. In case of possibilistic classification [3], $0 \leqslant \sum_{c=1}^{C} \mu_{\mathscr{C}_c}(\boldsymbol{x}) \leqslant 1$ and hence, $0 \leqslant \sum_{c=1}^{C} \iota_{\mathscr{C}_c}(\boldsymbol{x}) \leqslant C$. □

This property is useful while using a classifier. For an input $\boldsymbol{x}$, if $\sum_{c=1}^{C} \iota_{\mathscr{C}_c}(\boldsymbol{x}) \approx 0$, then we are not sure whether the input pattern belongs to any one of the output classes. Surprisingly, the human brain possesses the ability to indicate what it can classify and what it cannot. For instance, if an observer knows only Caucasian and Hispanic people, and if he sees some East Indian person, then he expresses doubt about the belongingness of the East Indian person into any of the two classes. The classifiers, whose outputs can be interpreted as the rough–fuzzy membership values, posses this interesting possibilistic classification ability. In addition, when the rough–fuzzy membership function in more than one class is close to one (say $\iota_{\mathscr{C}_1}(\boldsymbol{x}) \approx 1$ and $\iota_{\mathscr{C}_2}(\boldsymbol{x}) \approx 1$), the input pattern belongs to both the classes with high confidence. It happens typically in the applications where the output classes are overlapping largely. For instance, in the colorectal cancer follow-up program (see Application 2 in Section 6), the patient can be in the following output classes: *Well*, *recurrence*, *metastasis* and *both*. The classes *both* and *recurrence* (similarly *both* and *metastasis*) are highly overlapping. Therefore, $\iota_{\mathscr{C}_c}(\boldsymbol{x})$ can distinguish between equal evidence and ignorance.

Property 11. *For the crisp output classes* $\underline{R}(\mathscr{C}_c) = \{\boldsymbol{x} \in X \mid \iota_{\mathscr{C}_c}(\boldsymbol{x}) = 1\}$, $\bar{R}(\mathscr{C}_c) = \{\boldsymbol{x} \in X \mid \iota_{\mathscr{C}_c}(\boldsymbol{x}) > 0\}$, $BN(\mathscr{C}_c) = \bar{R}(\mathscr{C}_c) - \underline{R}(\mathscr{C}_c) = \{\boldsymbol{x} \in X \mid 1 > \iota_{\mathscr{C}_c}(\boldsymbol{x}) > 0\}$.

Proof. For the crisp output classes, the above results can be derived directly from Eqs. (4) and (5). □

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 $X = \{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n\}$ is an assignment of the rough–fuzzy membership value $\iota_{\mathscr{C}_c}(\boldsymbol{x}_i)$ on each $\boldsymbol{x}_i \in \mathscr{C}_c$ $\forall c = 1, 2, \ldots, C$ and $\forall i = 1, 2, \ldots, n$. In the rough–fuzzy context, C partitions of X are the set of values $\{\iota_{\mathscr{C}_c}(\boldsymbol{x}_i)\}$ that can conveniently be arranged on a $C \times n$ matrix $[\iota_{\mathscr{C}_c}(\boldsymbol{x}_i)]$. Based on the characteristic of $[\iota_{\mathscr{C}_c}(\boldsymbol{x}_i)]$, the classification can be of the following three types [3]:

   (a) *Crisp classification*:

   $$B_{hc} = \left\{ \iota_{\mathscr{C}_c}(\boldsymbol{x}) \in \mathbb{R}^{Cn} \mid \iota_{\mathscr{C}_c}(\boldsymbol{x}_i) \in \{0,1\}\ \forall c \forall i; \quad \sum_{c=1}^{C} \iota_{\mathscr{C}_c}(\boldsymbol{x}_i) = 1; \quad 0 \leqslant \sum_{i=1}^{n} \iota_{\mathscr{C}_c}(\boldsymbol{x}_i) \leqslant n\ \forall c \right\}.$$

   (b) *Constrained rough–fuzzy classification*:

   $$B_{fc} = \left\{ \iota_{\mathscr{C}_c}(\boldsymbol{x}) \in \mathbb{R}^{Cn} \mid \iota_{\mathscr{C}_c}(\boldsymbol{x}_i) \in [0,1]\ \forall c \forall i; \quad \sum_{c=1}^{C} \iota_{\mathscr{C}_c}(\boldsymbol{x}_i) = 1; \quad 0 \leqslant \sum_{i=1}^{n} \iota_{\mathscr{C}_c}(\boldsymbol{x}_i) \leqslant n\ \forall c \right\}.$$

   (c) *Possibilistic rough–fuzzy classification*:

   $$B_{pc} = \left\{ \iota_{\mathscr{C}_c}(\boldsymbol{x}) \in \mathbb{R}^{Cn} \mid \iota_{\mathscr{C}_c}(\boldsymbol{x}_i) \in [0,1]\ \forall c \forall i; \quad 0 \leqslant \sum_{i=1}^{n} \iota_{\mathscr{C}_c}(\boldsymbol{x}_i) \leqslant n\ \forall c \right\}.$$

   It is obvious that $B_{hc} \subseteq B_{fc} \subseteq B_{pc}$.

2. To make the definition of rough approximation loose, the model of variable precision rough set model is proposed by Ziarko [35]. The idea of the variable precision model can further be generalized by defining the following approximation for $\beta \in [0, 0.5]$:

   $$\underline{R}_\beta(\mathscr{C}_c) = \{\boldsymbol{x} \in X \mid \iota_{\mathscr{C}_c}(\boldsymbol{x}) \geqslant 1 - \beta\},$$

   $$\bar{R}_\beta(\mathscr{C}_c) = \{\boldsymbol{x} \in X \mid \iota_{\mathscr{C}_c}(\boldsymbol{x}) > \beta\}.$$

   When $\beta = 0$, we obtain Eqs. (4) and (5). Similarly, for $\beta \in [0.5, 1]$,

   $$\underline{R}_\beta(\mathscr{C}_c) = \{\boldsymbol{x} \in X \mid \iota_{\mathscr{C}_c}(\boldsymbol{x}) \geqslant \beta\},$$

   $$\bar{R}_\beta(\mathscr{C}_c) = \{\boldsymbol{x} \in X \mid \iota_{\mathscr{C}_c}(\boldsymbol{x}) > 1 - \beta\}.$$

   When $\beta = 1$, we obtain Eqs. (4) and (5).

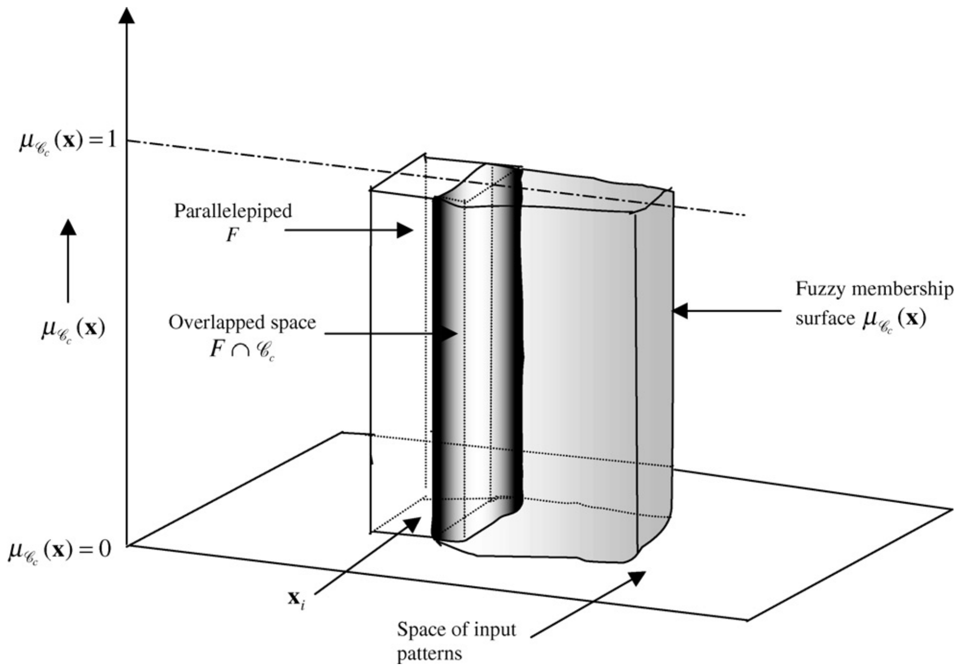

Fig. 4. When the output class is crisp, the fuzzy membership function is either 0 or 1. The roughness exists in the parallelepiped because some patterns from the parallelepiped do or do not belong to the output class $\mathscr{C}_c$. Thus, the input–output relationship becomes one-to-many, and the rough–fuzzy membership function becomes equal to the rough membership function.

3. The ι-rough–fuzzy inclusion of $A\subseteq X$ into $B\subseteq X$ can be defined as $A\subseteq_\iota B$ *iff* $\iota_A(\boldsymbol{x})\leqslant\iota_B(\boldsymbol{x})$ $\forall\boldsymbol{x}\in X$ If the output classes A and B are crisp and $A\subseteq_\iota B$, then from Property 11 it can be shown that $\underline{R}(A)\subseteq\underline{R}(B)$ and $\bar{R}(A)\subseteq\bar{R}(B)$. However, when the output classes are not crisp, the above relationship may or may not hold.
4. From the definition of the rough–fuzzy membership functions, we can group the rough–fuzzy sets into the following four categories:

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

These groups are in fact the generalization of the groups reported in [21].

4. Rough–fuzzy ownership functions

Till now we have discussed the roughness that appears when we ignore some of the features or attributes of the input pattern. Now we will describe the roughness that appears when we impose some restrictions in the structure of the set of input patterns. For instance, while using the conventional K-nearest neighbors algorithm [15], we consider a region or structure W around the test pattern consisting of K-nearest training patterns. The test pattern is classified to class $\mathscr{C}_c$ if most of the training patterns that reside in W belong to $\mathscr{C}_c$. Note that this structure is not naturally owned by the data; rather it is artificial. (In contrast, when we cluster a set of data, we obtain a natural structure. In a cluster, two members are close or similar to each other based on some criterion.) In the K-nearest neighbors algorithm, we assume that the neighbors and the test pattern are *similar*. In fact, here we assume some artificial neighborhood around the test pattern, and hence we expect that the neighbors and the test pattern should have similar class labels. In other words, in this case we call a training pattern neighbor just because it is closer to the test pattern relative to some other training patterns, although from the spatial distance and distribution point of views they may not be close. When two such patterns are artificially treated as similar, although their class assignments are quite different, a one-to-many relationship is created between the spatial similarity and the class labels. Consequently, the roughness emerges.

Similar phenomena are also observed in other fields. For example, we construct a crisp window around each data point of a time series while smoothing the time series. Here also it is assumed that the magnitude of a point in the time series is influenced by its neighbors. This assumption is usually valid for a damped physical system. It is because the response of a physical system does not have sudden jumps as long as (a) the damping is present and (b) the sufficient number of features are used to represent the input of the system. When the sufficient number of features are not used to represent the input–output response in form of a time series, sudden jumps appear in the time series (e.g., spikes in ECG signals). The magnitudes of the time series at these jumps are not similar to that of their neighbors; hence, the use of windows in such positions creates one-to-many relationships, and thus the rough uncertainty appears.

Let us formulate the above situations mathematically. If in the structure or window W, which has been constructed around the test pattern $\boldsymbol{x}$, all the neighbors are from a single class $\mathscr{C}_c$, then there is no uncertainty in the structure. Any pattern that resides in the structure can be assigned to $\mathscr{C}_c$. However, if any pattern of W belongs to another class $\mathscr{C}_j$, $j \neq c$, then the rough uncertainty arises in the structure. Although the patterns in W are similar from the features perspective, they are not similar from the class label perspective. It makes the relationship between the input representation and output class labels one-to-many. This uncertainty can be captured using the *rough ownership function*. The rough ownership function for the test pattern $\boldsymbol{x} \in X$ into the output class $\mathscr{C}_c$ is defined by

$$\rho_{\mathscr{C}_c}(\boldsymbol{x}) = \frac{\|W \cap \mathscr{C}_c\|}{\|W\|}, \tag{7}$$

where W is the neighborhood region around $\boldsymbol{x}$. We do not call the definition (7) rough membership function because it does not signify to what extent the test pattern is a *member* of a *natural* structure; rather it denotes to what extent the test pattern *owns* the *artificial* structure.

The mathematical framework for the rough–fuzzy ownership function is different from that of the rough–fuzzy membership function. We can observe the following:

- When we artificially form a structure W around the pattern $\boldsymbol{x} \in X$, we partition the whole pattern space into two equivalence classes: W and $(X - W)$. Within W, all members are assumed to be similar to $\boldsymbol{x}$; however, there may be other members outside W that are similar to $\boldsymbol{x}$ with respect to the given condition. In the rough–fuzzy ownership function we are only concerned about W, rather than about $(X-W)$. For instance, in the K-nearest neighbors algorithm, all the K-closest neighbors form the structure W around the test pattern $\boldsymbol{x}$. If $\boldsymbol{y} \in W$, $\boldsymbol{y}$ is similar to $\boldsymbol{x}$ with respect to the neighborhood condition. While finding the neighborhood for $\boldsymbol{y}$ in the K-nearest neighbors algorithm, we make another new partitioning. It involves construction of another window W_1 around $\boldsymbol{y}$. Eventually, $\boldsymbol{x}$ may or may not belong to the new structure W_1, i.e., $\boldsymbol{x}$ may or may not be one among the K-closest neighbors of $\boldsymbol{y}$. Thus it is similar to imposing directional symmetry within the structure W: We consider $\boldsymbol{x} \in W$ as similar to $\boldsymbol{y}$, but the reverse may not be true. Therefore here the concept of symmetry is not strictly followed.
- In the K-nearest neighbors algorithm, if $\boldsymbol{y}$ is the one of the K-closest neighbors of $\boldsymbol{x}$, and $\boldsymbol{z}$ is one of the K-closest neighbors of $\boldsymbol{y}$, then $\boldsymbol{z}$ may or may not

be treated as one of the K-closest neighbors of $\boldsymbol{x}$. Thus the transitivity relationship is not maintained always.

- From these angles, the concept of rough ownership function reflects some of the concepts of the generalized definition of rough sets proposed in [28]. In this paper, a generalization of the classical rough set is proposed where the indiscernibility relation may not follow the symmetry and transitivity properties.
- The rough–fuzzy membership function attempts to quantify the roughness that appears mainly due to the limitation of the representation; on the other hand, the rough–fuzzy ownership function measures the roughness that arises primarily due to our interpretation of the structure of the input set.

When the neighbors belong to more than one class, i.e., the class memberships are fuzzy, we need to modify the rough ownership function to the *rough–fuzzy ownership function*. The rough–fuzzy ownership function of a pattern $\boldsymbol{x}\in X$ into the fuzzy output class $\mathscr{C}_c$ is defined by

$$\varsigma_{\mathscr{C}_c}(\boldsymbol{x}) = \frac{\|W \cap \mathscr{C}_c\|}{\|W\|}. \tag{8}$$

Since W is crisp, $\mu_W(\boldsymbol{x})\in\{0,1\}$. Then

$$\varsigma_{\mathscr{C}_c}(\boldsymbol{x}) = \frac{\sum_{\boldsymbol{x}\in X}\min\{\mu_W(\boldsymbol{x}),\mu_{\mathscr{C}_c}(\boldsymbol{x})\}}{\|W\|} = \frac{1}{\|W\|}\sum_{\boldsymbol{x}\in W}\mu_{\mathscr{C}_c}(\boldsymbol{x}), \tag{9}$$

where $\|W\|$ is the cardinality of the window W, i.e., $\|W\|$ is the number of patterns in the window W.

All the properties, except Properties 6 and 11, hold for the rough–fuzzy ownership function when $\imath_{\mathscr{C}_c}(\boldsymbol{x})$ is substituted by $\varsigma_{\mathscr{C}_c}(\boldsymbol{x})$. Since the proofs are similar to that of the rough–fuzzy membership functions, we are not describing them separately. In the following discussion, we represent both rough–fuzzy membership functions and rough–fuzzy ownership functions using the same symbol $\imath$. The context will make it clear which one we are intending to refer.

5. Measures of rough–fuzzy uncertainty

The rough–fuzzy uncertainty plays a critical role in many classification problems since it is capable of modeling non-statistical uncertainty. Consequently, characterization and quantification of the rough–fuzziness are important issues, which influence the management of uncertainty in many classifier designs. Hence, the measures of the rough–fuzziness are essential to estimate the global uncertainty in an output class in some well-defined sense. A measure of the rough–fuzziness for a discrete output class $\mathscr{C}_c\in\mathscr{P}(X)$ is a mapping $H:\mathscr{P}(X)\to[0,1]$ that quantifies the degree of the rough–fuzziness in $\mathscr{C}_c$. Here, $\mathscr{P}(X)$ is the rough–fuzzy power set of X. The rough–fuzzy uncertainty of a set must be zero when there is no uncertainty in deciding whether an input pattern belongs to it or not. If the belongingness is maximally uncertain, i.e. $\imath_{\mathscr{C}_c}(\boldsymbol{x})=0.5\ \forall\boldsymbol{x}$, then the corresponding measure should be maximum. When the rough–fuzzy membership value of the input pattern approaches either to 0 or 1, the uncertainty about the belongingness of the pattern in the output class decreases; hence, the measure of the rough–fuzziness of the class should also decrease. A set $\mathscr{C}_c^*$ is called a *less uncertain* version of $\mathscr{C}_c$ if both the following conditions are satisfied for all $\boldsymbol{x}$:

$$\begin{aligned}&\imath_{\mathscr{C}_c^*}(\boldsymbol{x})\leqslant\imath_{\mathscr{C}_c}(\boldsymbol{x}) \quad \text{if } \imath_{\mathscr{C}_c^*}(\boldsymbol{x})\leqslant 0.5,\\ &\imath_{\mathscr{C}_c^*}(\boldsymbol{x})\geqslant\imath_{\mathscr{C}_c}(\boldsymbol{x}) \quad \text{if } \imath_{\mathscr{C}_c^*}(\boldsymbol{x})\geqslant 0.5.\end{aligned} \tag{10}$$

For a less uncertain version, the measure of the rough–fuzziness should decrease because the above two operations reduce the uncertainty. Another intuitively desirable property is that the measure of the rough–fuzziness of a set and its complement should be equal. Therefore, such a measure (i.e., H) should have the following properties:

*Certainty P*1: $H(\mathscr{C}_c)=0 \Leftrightarrow \imath_{\mathscr{C}_c}(\boldsymbol{x})=0$ or $1\ \forall\boldsymbol{x}\in X$.

*Maximal P*2: $H(\mathscr{C}_c)$ is maximum $\Leftrightarrow \imath_{\mathscr{C}_c}(\boldsymbol{x})=0.5\ \forall\boldsymbol{x}\in X$.

*Resolution P*3: $H(\mathscr{C}_c)\geqslant H(\mathscr{C}_c^*)$, where $H(\mathscr{C}_c^*)$ is a less uncertain version of $H(\mathscr{C}_c)$.

*Symmetry P*4: $H(\mathscr{C}_c)=H(X-\mathscr{C}_c)$, where $\imath_{X-\mathscr{C}_c}(\boldsymbol{x}) = 1-\imath_{\mathscr{C}_c}(\boldsymbol{x})\ \forall\boldsymbol{x}\in X$.

One such possible measure that satisfies all the above properties is *index of rough–fuzziness*. It is defined as

$$H_I(\mathscr{C}_c) = \frac{2}{n_k} d(\mathscr{C}_c, \mathscr{C}_c^{\text{near}}), \tag{11}$$

where d is a distance measure, $k \in \mathbb{R}^+$ depends on d, n is the number of patterns in the set X, and $\mathscr{C}_c^{\text{near}}$ is a set that is without the rough–fuzzy uncertainty and is the nearest to $\mathscr{C}_c$. For $\mathscr{C}_c^{\text{near}}$, the rough–fuzzy membership function is defined as

$$\iota_{\mathscr{C}_c^{\text{near}}}(\boldsymbol{x}) = \begin{cases} 1 & \text{if } \iota_{\mathscr{C}_c}(\boldsymbol{x}) \geqslant 0.5, \\ 0 & \text{otherwise.} \end{cases} \tag{12}$$

For the *Minkowski q norms*, $d(\mathscr{C}_c, \mathscr{C}_c^{\text{near}})$ and $H_I(\mathscr{C}_c)$ appear in the following forms:

$$d(\mathscr{C}_c, \mathscr{C}_c^{\text{near}}) = \left[\sum_{\boldsymbol{x} \in X} |\iota_{\mathscr{C}_c}(\boldsymbol{x}) - \iota_{\mathscr{C}_c}^{\text{near}}(\boldsymbol{x})|^q \right]^{1/q}, \tag{13}$$

$$H_I(\mathscr{C}_c) = \frac{2}{n^{1/q}} \left[\sum_{\boldsymbol{x} \in X} |\iota_{\mathscr{C}_c}(\boldsymbol{x}) - \iota_{\mathscr{C}_c}^{\text{near}}(\boldsymbol{x})|^q \right]^{1/q}, \tag{14}$$

where $|\boldsymbol{x}|$ is the absolute value of $\boldsymbol{x}$ and $q \in [1, \infty)$. For $q=1$, $H_I(\mathscr{C}_c)$ is called *linear index of rough–fuzziness*, and for $q=2$, $H_I(\mathscr{C}_c)$ is known as *quadratic index of rough–fuzziness*.

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

$$H_E(\mathscr{C}_c) = -K \sum_{\boldsymbol{x} \in X} [\iota_{\mathscr{C}_c}(\boldsymbol{x}) \log(\iota_{\mathscr{C}_c}(\boldsymbol{x})) + (1 - \iota_{\mathscr{C}_c}(\boldsymbol{x})) \log(1 - \iota_{\mathscr{C}_c}(\boldsymbol{x}))]. \tag{15}$$

The term log denotes the logarithm to any base $a > 1$, and $K \in \mathbb{R}^+$ is a normalizing constant. Here, H_E can be normalized by adjusting K so that it satisfies $P1$ to $P4$. From Property 4, when the partitioning in X is *fine*, H_E reduces to the fuzzy entropy [17] as follows:

$$H_E(\mathscr{C}_c) = -K \sum_{\boldsymbol{x} \in X} [\mu_{\mathscr{C}_c}(\boldsymbol{x}) \log(\mu_{\mathscr{C}_c}(\boldsymbol{x})) + (1 - \mu_{\mathscr{C}_c}(\boldsymbol{x}))(1 - \mu_{\mathscr{C}_c}(\boldsymbol{x}))]. \tag{16}$$

In the absence of the fuzziness, H_E is reduced to the rough entropy [31] as follows (see Property 3):

$$H_E(\mathscr{C}_c) = -K \sum_{\boldsymbol{x} \in X} [r_{\mathscr{C}_c}(\boldsymbol{x}) \log(r_{\mathscr{C}_c}(\boldsymbol{x})) + (1 - r_{\mathscr{C}_c}(\boldsymbol{x}))(1 - r_{\mathscr{C}_c}(\boldsymbol{x}))]. \tag{17}$$

Similarly there are other measures like *multiplicative* and *additive* measures, which can be directly borrowed from the fuzzy set theory [17]. All these measures satisfy $P1$–$P4$.

Multiplicative measure of rough–fuzziness: Any function $H_M : 2^X \to \mathbb{R}^+$ is a multiplicative rough–fuzzy measure if it can be written as

$$H_M(\mathscr{C}_c) = -K \sum_{\boldsymbol{x} \in X} g(\iota_{\mathscr{C}_c}(\boldsymbol{x})) \quad K \in \mathbb{R}^+, \tag{18}$$

where

$$g(t) = \tilde{g}(t) - \min_{0 \leqslant t \leqslant 1} \{\tilde{g}(t)\},$$

$$\tilde{g}(t) = f(t) f(1-t)$$

and $f : [0,1] \to \mathbb{R}^+$ is a concave increasing function. By taking $f(t) = t\exp(1-t)$, it can be shown that $-K\sum_{\boldsymbol{x} \in X} [\iota_{\mathscr{C}_c}(\boldsymbol{x})(1 - \iota_{\mathscr{C}_c}(\boldsymbol{x}))]$ is an example of this kind of measure [17].

Additive measure of rough–fuzziness: Any function $H_A : 2^X \to \mathbb{R}^+$ is an additive rough–fuzzy measure provided it could be written as

$$H_A(\mathscr{C}_c) = -K \sum_{\boldsymbol{x} \in X} g(\iota_{\mathscr{C}_c}(\boldsymbol{x})) \quad K \in \mathbb{R}^+, \tag{19}$$

where

$$g(t) = \tilde{g}(t) - \min_{0 \leqslant t \leqslant 1} \{\tilde{g}(t)\},$$

$$\tilde{g}(t) = f(t) + f(1-t)$$

and $f : [0,1] \to \mathbb{R}^+$ is an increasing function. If we take $f(t) = t\exp(1-t)$, then one example of this kind of measure is [17] $-K\sum_{\boldsymbol{x} \in X} [\iota_{\mathscr{C}_c}(\boldsymbol{x}) \exp(1 - \iota_{\mathscr{C}_c}(\boldsymbol{x})) + (1 - \iota_{\mathscr{C}_c}(\boldsymbol{x})) \exp(\iota_{\mathscr{C}_c}(\boldsymbol{x})) - 1]$.

6. Applications

The rough–fuzzy membership or ownership functions can be used where class labels are fuzzy, and the class labels in some cases are different from what we would have predicted by observing the features of the pattern. Below we outline some interesting cases although the implementations are generally more complex in practice. Specifically Examples 1–3 show the applications of rough–fuzzy membership functions, and Examples 4 and 5 show the applications of rough–fuzzy ownership functions.

Example 1. *Document classification using natural language processing*: In this problem a set of documents $\boldsymbol{D}=\{D_1,D_2,\ldots,D_m\}$ are to be classified into any one of the output classes $\{\mathscr{C}_1,\mathscr{C}_2,\ldots,\mathscr{C}_C\}$. This kind of problem arises frequently in web page classification. The problem is generally tackled in the following two phases: Keywords generation and classification.

Keywords generation: Without losing the generality, let us consider the document D_1. All words in the document D_1 are extracted in form of a list L_1. From the list L_1 all *stop words* (i.e., words like 'a', 'an', 'the', 'is', 'to', etc., which do not add much value to the meaning of the sentence) are removed. The frequency of occurrence of each member of L_1 is counted. Thus, the words in the list L_1 that are synonymous form separate groups. Thus the set L_1 is partitioned into a set of equivalence classes $\{F_{11},F_{12},\ldots,F_{1K}\}$ such that each equivalence class contains the words with immediate synonyms. To find the synonyms, *WordNet* [1] package is used. Now the strength of each equivalence class is judged based on the number of times their members appear in the document. All the equivalence classes are sorted using their strengths, and the top 5.0% equivalence classes are selected. The members of the selected equivalence classes are treated as the *keywords* of the document. This procedure is similar to crisp clustering. Once the keywords are extracted from a document, the document can be related to different classes.

Classification: In practice, the output classes may be overlapping like "war", "science", "politics", "economy", etc. Some of the documents may be related partially to "war" and partially to "science" since they report some scientific discoveries related to war [29]. It creates fuzziness. In addition, some members of an equivalence class are related to "war" and some others are related to "science". It creates one-to-many relationship between the equivalence class and the output class. There is a set of terms $t=\{t_1,t_2,\ldots,t_l\}$ available from the experts, where the fuzzy class labels for each term are obtained from the subjective judgments of the experts. Now we consider only the keywords that are present in the set t, i.e., we remove the members of the equivalence classes $F_{11},F_{12},\ldots,F_{1K}$ that are not in t. Thus each element of the equivalence class F_{11} (say) has the fuzzy class labels. The rough–fuzzy membership function can be used here to find the class label of the equivalence class F_{11}. One way to find the fuzzy class label of the complete document D_1 is $conf_c(D_1)=\sum_{i=1}^{K}\iota_{\mathscr{C}_c}(F_{1i})\ \forall c$. The crisp class label of the document is $\arg\max_c\{conf_c(D_1)\}$.

Example 2. *Importance of features*: The rough–fuzzy membership functions in conjugation with the rough–fuzzy entropy can be used to measure the importance of a feature in a given classification task. Here we illustrate this problem in the context of *colorectal cancer* diagnosis. In this example, we study the association between the key prognostic factors and the outcomes of the patients who are undergoing the follow-up program of the colorectal cancer.

The colorectal cancer occurs frequently in the developed countries. The colorectal cancer forms initially in the mucosa lining of the bowel. In most cases, the first step in the formation of a colorectal cancer is the appearance of polyps. When the abnormal cells within the polyps begin to spread and invade through normal tissue, polyps become cancer growths. If no proper treatment is adopted, then the cancer can spread beyond the skin and the underlying tissues of the bowel wall, and eventually the cancer may spread to the distant sites like liver. To express the condition of the patient, the following four possible outcomes are used: *Well*, *recurrence*, *metastasis* and *both* (i.e., recurrence and metastasis simultaneously). The main treatment for the colorectal cancer is the surgical removal of the tumor, while the survival of a patient with the colorectal cancer is dependent on four fundamental factors: (a) The biology of that individual's malignancy, (b) the immune response to the tumor, (c) the time in

[1] WordNet, a lexical database for English, is publicly available from http://www.cogsci.princeton.edu/∼wn. English nouns, verbs, adjectives and adverbs are organized into synonym sets, each representing one underlying lexical concept.

the cancer patient's life history when the diagnosis is made, and (d) the adequacy of the treatment. About 50% patients eventually die from the local recurrence and/or distant metastasis within 5 years after the curative resection [16]. Therefore, it is important to detect or predict the recurrent or metastasis tumor in the follow-up so that the appropriate therapy is prescribed to increase the chance of survival.

The following 16 attributes are considered: Liver metastasis, peritoneal metastasis, CT scan for liver, regional lymph node, apical node status, apical node number, adjacent structure invasion, venous invasion, perineaural invasion, differentiation, ascites, carcinoembryonic antigen, surgeon rank, site of tumor, perioperative blood transfusion, and fix to adjacent structure. The patient, who is in the follow up program, may fall into any of the following states: Metastasis, recurrence, bad and well. If the state of a particular patient can be correctly decided, then the state information can be utilized to choose an appropriate treatment. Here a patient can belong to more than one class. A physician can subjectively judge the belongingness of each patient in the output classes.

Each feature partitions the set of patients into some equivalence classes. Let us consider the sth feature. The sth feature is considered important if the compactness and interclass distance of all the classes along the sth axis is high. We attempt to exploit this criterion to measure the importance of each feature. The compactness of the classes is affected when the classes are overlapping and the patterns with the same sth feature have different class labels. The lower the value of the corresponding entropy $H_E^s(\mathscr{C}_c)$ is, the greater is the number of patterns having $\iota_{\mathscr{C}_c}(\boldsymbol{x}) \approx 1$ or $\iota_{\mathscr{C}_c}(\boldsymbol{x}) \approx 0$, i.e., less is the difficulty in deciding whether the pattern can be considered as a member of $\mathscr{C}_c$ or not. In particular, when $\iota_{\mathscr{C}_c}(\boldsymbol{x}) \approx 1$, greater is the tendency of $\boldsymbol{x}$ to form a compact class $\mathscr{C}_c$ along the sth axis, resulting in less internal scatter along the sth axis. Moreover, when $\iota_{\mathscr{C}_c}(\boldsymbol{x}) \approx 0$, along the sth axis $\boldsymbol{x}$ is far away from the cth class, and hence, the interclass distance increases along the sth axis. On the other hand, when $\iota_{\mathscr{C}_c}(\boldsymbol{x}) \approx 0.5$, $\boldsymbol{x}$ lies in between $\mathscr{C}_c$ and the other classes along the sth axis. Hence, both compactness and interclass distance along the sth axis decrease. The reliability of a feature s, in characterizing the class $\mathscr{C}_c$, increases as the corresponding H_c^s value decreases. Therefore, $H_E^s(\mathscr{C}_c)$ quantifies the importance of the sth input feature for the output class $\mathscr{C}_c$. We introduce *total rough–fuzzy entropy* to quantify the importance of the sth input feature for all the classes. It is defined as

$$\mathscr{H}_E^s = \sum_{c=1}^{C} P_c H_E^s(\mathscr{C}_c). \tag{20}$$

Here P_c is the weight that has to be assigned to the cth class. One possible choice for P_c is the a priori probability of the cth class. Note that $\mathscr{H}_E^s$ lies in $[0,1]$. Evidently, the more the value of $\mathscr{H}_E^s$ is, the less is the importance of the sth feature. Using the above technique (with some bias corrections as given in [13]), we have attempted to quantify the importance of the prognostic factors.

Example 3. *Rule generation*: In classification problems, rough sets are used to extract the rules present in the given input data set [20]. Suppose the input data set has only one feature, and there exist two equivalence classes $[\boldsymbol{x}]_R$ and $[\boldsymbol{y}]_R$ generated by the partition of the input set. If it is a 2-class problem, then the set of extracted rules is

R_1: If the *input* is x, then the *output class* is $\mathscr{C}_1$ with the *confidence factor* $\gamma_{\mathscr{C}_1}(x)$,
R_2: If the *input* is x, then the *output class* is $\mathscr{C}_2$ with the *confidence factor* $\gamma_{\mathscr{C}_2}(x)$,
R_3: If the *input* is y, then the *output class* is $\mathscr{C}_1$ with the *confidence factor* $\gamma_{\mathscr{C}_1}(y)$,
R_4: If the *input* is y, then the *output class* is $\mathscr{C}_2$ with the *confidence factor* $\gamma_{\mathscr{C}_2}(y)$.

One problem in constructing the above rule base is that we do not know the values of the confidence factors. However, the rough–fuzzy membership functions can be helpful here. The confidence factors $\gamma_{\mathscr{C}_c}(z)$ for $z = x, y$ and $c = 1, 2$ can be made equal to $\iota_{\mathscr{C}_c}(z)$. We can observe that two rules are assigned to the class 1. Now we can find which rule is more useful between the two. It may seem that the rule with the highest *true positive rate* for the class 1 is the most useful one. The true positive rate for the class 1 can be computed from $tp_1 = \|F \cap \mathscr{C}_1\| / \|\mathscr{C}_1\|$. However, it may happen that this particular rule causes many class 2 patterns to be classified as class 1. Therefore, the most useful rule is the one that produces the least *false positive rate* and the largest true positive rate.

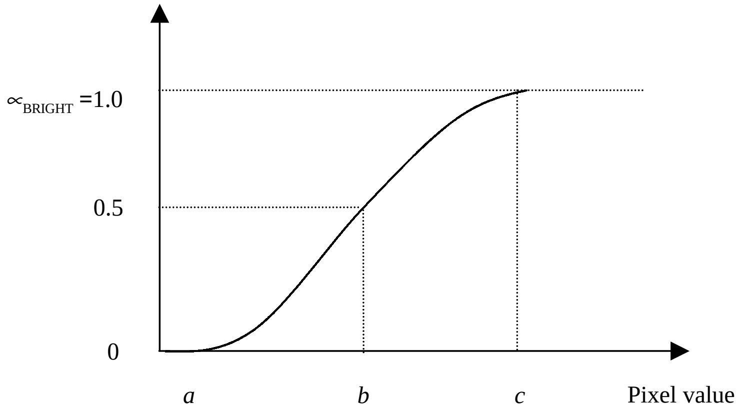

Fig. 5. An example of a fuzzy function is showing the relationship between the pixel value and the brightness. The points *a* and *c* denote the minimum and maximum intensity values in the given image. The point *b* on the abscissa, at which the membership value is 0.5, is called the *crossover point*.

The false positive rate for the class 1 can be computed from $fp_1 = \|F \cap (X - \mathscr{C}_1)\| / \|(X - \mathscr{C}_1)\|$. However, from Property 5, we know $fp_1 = 1 - \iota_{\mathscr{C}_1}(\boldsymbol{x})$. Thus, the true positive rate and the rough–fuzzy membership function can be used together to select the best fuzzy rules in the rule base. Moreover, if the confidence factor of any rule is equal to zero or very close to zero, then we can directly remove this particular rule from the rule base since it is useless from the classification point of view. Thus, some sort of optimization of the rule base can be achieved by using the rough–fuzzy membership functions.

Example 4. *Image segmentation*: In the image thresholding problems [18], the objective is to transform a gray level image into a binary image. In the gray level image, each pixel can be in between $[0, 255]$ where pixel value 0 represents the background or dark, and the pixel value 255 indicates foreground or bright intensity. In contrast, in the thresholded image or segmented image each pixel can have values only in $\{0, 1\}$, where the values 0 and 1 represent the background and foreground, respectively. The set of intensity values of the pixels in the gray-leveled image can be treated as a fuzzy set BRIGHT, where the maximum pixel value of the image indicates complete membership into the set BRIGHT, and the minimum pixel value indicates zero membership into the set BRIGHT. A fuzzy membership function can be constructed by employing the intensity value of each pixel and by using the following membership function (Fig. 5):

$$
\begin{aligned}
\mu_{\mathrm{BRIGHT}}(\boldsymbol{x}) &= 0 && \text{for } x \leqslant a,\\
&= 0.5\left(\frac{x-a}{b-a}\right)^2 && \text{for } a \leqslant x \leqslant b,\\
&= 0.5 + 0.5\left(\frac{x-b}{c-b}\right)^2 && \text{for } b \leqslant x \leqslant c,\\
&= 1 && \text{for } x \geqslant c.
\end{aligned}
$$

The points *a* and *c* denote the minimum and maximum intensity values in the given image. The point *b* on the abscissa, at which the membership value is 0.5, is called the *crossover point*.

One simple approach of segmentation is to transform any pixel into 1 (or 0) if its intensity value is more (or less) than the *crossover* point *b*. This simple scheme does not work well because many pixels are affected by noise, and hence after thresholding the image contains numerous disjoints regions of background or foreground. The quality of the thresholded image can be improved by considering a pixel bright (dark) if its neighbors are also bright (or dark). It makes the foreground or background continuous. Now a window W of a certain size (say 3×3) is considered around each pixel (say x) of the image. The average brightness within the window, i.e., $\tilde{\mu}_{\mathrm{BRIGHT}}(x) = \frac{1}{9}\sum_{y \in W} \mu_{\mathrm{BRIGHT}}(y)$, is used as the fuzzy membership of the pixel x in the class BRIGHT

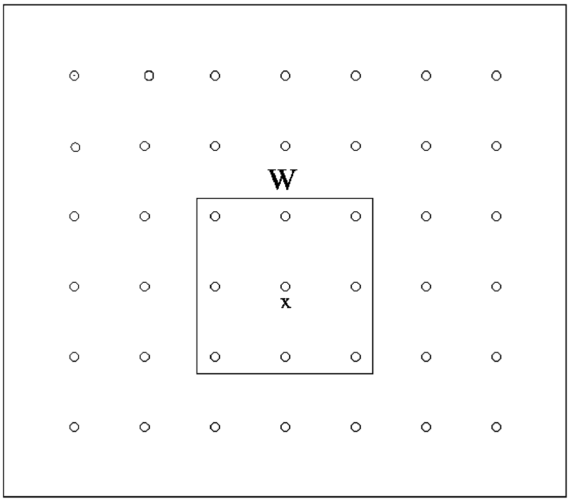

Fig. 6. For segmenting the 6×7 image, a 3×3 window W is considered around the pixel x. The brightness of the pixel x is expressed as $\mu_{\mathrm{BRIGHT}}(x)$. For better segmentation results, the influence of the neighboring pixels is also considered [18]. Hence, the brightness of the pixel x is assumed to be $\tilde{\mu}_{\mathrm{BRIGHT}}(x) = \frac{1}{9}\sum_{y \in W} \mu_{\mathrm{BRIGHT}}(y)$. $\tilde{\mu}_{\mathrm{BRIGHT}}(x)$ is in fact the rough–fuzzy ownership function.

(Fig. 6). The derived membership values $\tilde{\mu}_{\mathrm{BRIGHT}}(x)$ for all the pixels of the image are used to calculate the fuzzy entropy corresponding to BRIGHT. The segmented image can be highly homogeneous when the fuzzy uncertainty of the image with respect to BRIGHT is the least. Hence, the fuzzy entropy, i.e., $\sum_{x \in \mathrm{image}}[\tilde{\mu}_{\mathrm{BRIGHT}}(x)\log(\tilde{\mu}_{\mathrm{BRIGHT}}(x)) + (1 - \tilde{\mu}_{\mathrm{BRIGHT}}(x))\log(1 - \tilde{\mu}_{\mathrm{BRIGHT}}(x))]$, is computed by fixing the crossover point at all possible pixel values in between a and c of the fuzzy membership functions. The threshold point is chosen as the crossover point for which the entropy is minimum. This procedure can be looked at from the rough–fuzzy ownership function angle. The window around x is a crisp structure, and $\tilde{\mu}_{\mathrm{BRIGHT}}(x) = \frac{1}{9}\sum_{y \in W} \mu_{\mathrm{BRIGHT}}(y) = \|W \cap \mathscr{C}_{\mathrm{BRIGHT}}\| / \|W\|$ is precisely the expression for the rough–fuzzy ownership function of x. Then we calculate the rough–fuzzy entropy of the output class BRIGHT, and we seek to fix the threshold so that the rough–fuzzy entropy becomes the least. Here our understanding in the rough–fuzziness sheds light on the validity of taking $\tilde{\mu}_{\mathrm{BRIGHT}}(x)$ instead of $\mu_{\mathrm{BRIGHT}}(x)$ as the membership value of x.

Example 5. *K-nearest neighbors algorithm*: In classification problems, one well-know non-parametric technique is K-nearest neighbors algorithm [15]. It assigns the class label to the test pattern based on the class labels of the K-closest (in some distance sense) neighbors of the input. All the K-neighbors are from the training set. In a dense region, this neighborhood region is small since the K-nearest neighbors will be found within a close distance. Similarly, in a sparse region, this neighborhood occupies a larger space. It implies that we are trying to construct a structure W around the test pattern $\boldsymbol{y}$. If all the neighbors are from a single class $\mathscr{C}_c$, then there is no uncertainty in the structure. Any test pattern that resides in the structure W can be assigned to $\mathscr{C}_c$. However, if any neighbor belongs to another class $\mathscr{C}_j$, $j \neq c$, then the rough uncertainty arises in the structure. Although the neighbors are similar from the features perspective, they are not similar from the class label perspective. It makes the relationship between the input representation and the output class labels one-to-many. This uncertainty can be captured using the rough ownership function. This is precisely used in the K-nearest neighbors algorithm. Moreover, each neighbor can belong to more than one class, i.e., the class memberships are fuzzy. To accommodate both fuzzy and rough uncertainties in the K-nearest neighbors setting, the K-nearest neighbors algorithm is slightly modified such that its outputs can be interpreted as the rough–fuzzy ownership values. The resultant algorithm is more powerful than the simple K-nearest neighbors counterpart [25].

7. Summary and conclusions

We have discussed some pattern classification problems where the rough and fuzzy uncertainties together arise when some of the features are ignored and the output classes are overlapping. Specifically we have illustrated how the rough–fuzzy uncertainty of a pattern could be quantified using the rough–fuzzy membership function. In addition, we have shown that the rough–fuzzy uncertainty can also appear when we artificially call two patterns similar although from the pattern distribution angle they are not similar. To characterize the rough–fuzzy uncertainty of a set of patterns, different measures are discussed. In the application fields, we observe that the concepts of rough–fuzzy membership function, rough–fuzzy ownership function and rough–fuzzy entropy have already been used unknowingly in many classification problems.

In many practical problems, both the equivalence class and the artificial structure around the test pattern can be fuzzy. Thus, the pattern set is partitioned into fuzzy equivalence classes by the induced fuzzy equivalence relation. Specifically, in this case the crisp transitivity relationship is weakened by the fuzzy transitivity relationship. To frame this type of model, we can borrow some of the concepts used in [28]. We have kept this problem as a future work.

Acknowledgements

The constructive comments from the reviewers' helped us to enhance the quality of the paper.

References

[1] J.C. Bezdek, Pattern Recognition with Fuzzy Objective Function Algorithms, Plenum Press, New York, 1981.
[2] J.C. Bezdek, The thirsty traveler visits gamont: a rejoinder to "comments on fuzzy set—what are they and why?", IEEE Trans. Fuzzy Systems 2 (1) (1994) 43–45.
[3] J.C. Bezdek, J. Keller, R. Krishnapuram, N.R. Pal, Fuzzy Models and Algorithms for Pattern Recognition and Image Processing, Kluwer Academic Publisher, Boston, 1999.
[4] D. Dubois, H. Prade, Fuzzy Sets and Systems, Academic Press, New York, 1980.
[5] D. Dubois, H. Prade, Rough fuzzy sets and fuzzy rough sets, Internat. J. General Systems 17 (2–3) (1990) 191–209.
[6] D. Dubois, H. Prade, Putting rough sets and fuzzy sets together, in: R. Slowinski (Ed.), Intelligent Decision Support, Handbook of Applications and Advances of the Rough Set Theory, Kluwer Academic Publishers, Dordrecht, 1992.
[7] S. Greco, B. Matarazzo, R. Slowinski, Fuzzy similarity relation as a basis for rough approximations, in: L. Polkowski, A. Skowron (Eds.), Rough Sets and Current Trends in Computing, Lecture Notes in Artificial Intelligence, Vol. 1424, Springer, Berlin, 1998, pp. 283–289.
[8] S. Greco, B. Matarazzo, R. Slowinski, The use of rough sets and fuzzy sets in MCDM, in: T. Gal, T. Stewart, T. Hanne (Eds.), Advances in Multiple Criteria Decision Making, Kluwer Academic Publishers, Boston/Dordrecht/London, 1999 (Chapter 14).
[9] S. Greco, B. Matarazzo, R. Slowinski, Rough set processing of vague information using fuzzy similarity relations, in: C.S. Calude, G. Paun (Eds.), Finite versus Infinite Contributions to an Eternal Dilemma, Springer, London, 2000, pp. 149–173.
[10] S. Greco, B. Matarazzo, R. Slowinski, Rough sets theory for multicriteria decision analysis, European J. Oper. Res. 129 (1) (2001) 1–47.
[11] A. Kandle, Fuzzy Mathematical Techniques with Applications, Addison-Wesley, Reading, MA, 1986.
[12] G.S. Klir, B. Yuan, Fuzzy sets and Fuzzy Logic—Theory and Applications, Prentice-Hall, Englewood Cliffs, NJ, 1995.
[13] I. Kononenko, On biases in evaluating multivalued attributes, Proc. Internat. Joint Conf. on Artificial Intelligence, Montreal, August 20–25, 1995, pp. 1034–1040.
[14] C.T. Lin, C.S.G. Lee, Neural Fuzzy Systems, Prentice-Hall, Englewood Cliffs, NJ, 1996.
[15] T.M. Mitchell, Machine Learning, McGraw-Hill, New York, 1997.
[16] L.S. Ong, B. Shepherd, L.C. Tong, F.S. Choen, Y.H. Ho, C.L. Tang, Y.S. Ho, K. Tan, The colorectal cancer recurrence support (CARES) system, Artificial Intelligence Medicine 11 (1997) 175–188.
[17] N.R. Pal, J.C. Bezdek, Measuring fuzzy entropy, IEEE Trans. Fuzzy Systems 2 (2) (1994) 107–118.
[18] S.K. Pal, D.D. Majumder, Fuzzy Mathematical Approach to Pattern Recognition, Wiley (Halsted Press), New York, 1986.
[19] Z. Pawlak, Rough sets, Internat. J. Comput. Inform. Sci. 11 (1982) 341–356.
[20] Z. Pawlak, Rough Sets: Theoretical Aspects of Reasoning about Data, Kluwer Academic Publishers, Dordrecht, 1991.
[21] 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.
[22] Z. Pawlak, Rough sets present state and further prospects, Technical Report ICS Research Report 32/94, Institute of Computer Science, Warsaw University of Technology, Warsaw, Poland, 1995.
[23] Z. Pawlak, J.G. Busse, R. Slowinski, W. Ziarko, Rough sets, Commun. ACM 38 (11) (1995) 89–95.
[24] M. Quafafou, [alpha]-RST: a generalization of rough set theory, Inform. Sci. 124 (1–5) (2000) 301–316.
[25] M. Sarkar, Fuzzy–rough nearest neighbors algorithm, Proc. IEEE Internat. Conf. on Systems, Man and Cybernetics, Nashville, Tennessee, USA, October 8–11, 2000, pp. 3556–3601.
[26] R. Slowniski, J. Stefanowski, Rough set reasoning about uncertain data, Fundam. Inform. 27 (1996) 229–243.
[27] R. Slowinski, J. Stefanowski, S. Greco, B. Matarazo, Rough set based processing of inconsistent information in decision analysis, Control Cybernet. 29 (1) (2000) 379–404.
[28] R. Slowinski, D. Vanderpooten, A generalized definition of rough approximations based on similarity, IEEE Trans. Knowledge Data Eng. 12 (2) (2000) 331–336.
[29] P. Srinivasan, M.E. Ruiz, D.H. Kraft, J. Chen, Vocabulary mining for information retrieval: rough sets and fuzzy sets, Inform. Process. Manage. 37 (1) (2001) 15–38.
[30] J. Stefanowski, Rough sets theory and discriminant methods in an analysis of information systems, Foundam. Comput. Decision Sci. 17 (1992) 81–89.

[31] S.K.M. Wong, W. Ziarko, Comparison of the probabilistic approximate classification and fuzzy set model, Fuzzy Sets and Systems 21 (1987) 357–362.
[32] Y.Y. Yao, Information granulation and rough set approximation, Internat. J. Intell. System 16 (1) (2001) 87–104.
[33] L.A. Zadeh, Fuzzy sets, Inform. Control 8 (1965) 338–353.
[34] L.A. Zadeh, Fuzzy sets as a basis for a theory of possibility, Fuzzy Sets and Systems 1 (1978) 3–28.
[35] W. Ziarko, Variable rough set model, J. Comput. System Sci. 40 (1) (1993) 39–59.