Abstract

Fuzzy sets and rough sets are known as uncertainty models. They are proposed to treat different aspects of uncertainty. Therefore, it is natural to combine them to build more powerful mathematical tools for treating problems under uncertainty. In this chapter, we describe the state of the art in the combinations of fuzzy and rough sets dividing into three parts.

In the first part, we first describe two kinds of models of fuzzy rough sets: one is classification-oriented model and the other is approximation-oriented model. We describe the fundamental properties and show the relations of those models. Moreover, because those models use logical connectives such as conjunction and implication functions, the selection of logical connectives can sometimes be a question. Then we propose a logical connective-free model of fuzzy rough sets.

In the second part, we develop a generalized fuzzy rough set model. We first introduce general types of belief structures and their induced dual pairs of belief and plausibility functions in the fuzzy environment. We then build relationships between belief and plausibility functions in the Dempster-Shafer theory of evidence and the lower and upper approximations in rough set theory in various situations. We also provide the potential applications of the main results to intelligent information systems.

In the third part, we give an overview of the practical applications of fuzzy rough sets. The main focus will be on the machine learning domain. In particular, we review fuzzy-rough approaches for attribute selection, instance selection, classification and prediction.

Fuzzy-rough Hybridization

Masahiro Inuiguchi[1)], Wei-Zhi Wu[2)], Chris Cornelis[3)4)]
and Nele Verbiest[4)]

1) Graduate School of Engineering Science,
Osaka University, Toyonaka, Osaka 560-8531, Japan
2) School of Mathematics, Physics and Information Science,
Zhejiang Ocean University,
Zhoushan, Zhejiang, 316022, PR China
3) Department of Computer Science and Artificial Intelligence,
University of Granada, 18071 Granada, Spain
4) Department of Applied Mathematics, Computer Science
and Statistics, Ghent University, B-9000 Gent, Belgium

November 11, 2013

1 Introduction

Rough set approaches [1, 2] have been successfully applied to various fields related to data analysis, knowledge discovery, decision analysis and so on. In order to expand the application area and to develop its theory further, rough sets have been generalized under various settings. There are two different generalizations. One relaxed the precision so that the sizes of lower and upper approximations are controlled by a precision parameter. This generalized rough set is called a variable precision rough set. The other generalizes the approximation space, i.e., the structure of background knowledge. Many researchers generalized an equivalence relation which is often referred to as an indiscernibility relation to a general binary relation or a family. The other many researchers [3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23] generalized an equivalence relation to a fuzzy binary relation or a family of fuzzy sets.

In this chapter, we describe the generalizations of rough sets in the latter sense. More precisely, we concentrate on the fuzzy generalizations of rough set approaches called fuzzy rough hybridizations. Fuzzy rough sets were originally proposed by Nakamura [3] and by Dubois and Prade [4, 5]. The fundamental properties of fuzzy rough sets have been investigated by Dubois and Prade [4, 5] and Radzikowska and Kerre [9]. In those studies, an equivalence relation of approximation space in the original rough sets is generalized to a fuzzy equivalence relation. Greco et al. [7] proposed fuzzy rough sets under a fuzzy dominance relation. Those fuzzy rough sets are based on possibility and necessity measures directly. Moreover, this type of fuzzy rough sets is defined under more generalized settings [11, 15] and different types of fuzzy rough sets were proposed based on certainty qualifications by Inuiguchi and Tanino [10, 12] and based on modifier functions by Greco et al. [24, 25]. The rough fuzzy set model can be used to deal with attribute reduction in information systems with fuzzy decision while the fuzzy rough set model can be employed in reasoning and knowledge acquisition with decision tables with real valued conditional attributes or quantitative data (see, for example, [39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49]).

In the first part of this paper, we introduce three models of fuzzy rough sets. Those fuzzy sets are classified into two groups, i.e., classification-oriented fuzzy rough set models and approximation-oriented fuzzy rough set models proposed by Inuiguchi [26] originally in the crisp settings. In the classification-oriented models, we are interested in a set to which objects belong. We evaluate each object whether its membership to a set X is consistent with all information we have at hand or not. The positive region of X is defined by collecting all objects whose memberships to X are consistent with whole information. The possible region of X is defined by collecting all objects whose memberships to X are conceivable from some part of information but not consistent with all information. Then the fuzzy rough set of X is defined by a pair of the positive and possible regions of X. On the contrary, in approximation-oriented models, we are interested in the approximations of a set by using elementary sets of a family. We approximate a set X by unions of the elementary sets and by intersections of the complementary sets of the elementary sets. The lower and upper approximations are defined by the inner and outer approximations of X, respectively. A rough set of X is defined by a pair of the lower and upper approximations. We describe that one of the three models belongs to the group of classification-oriented models and the remaining two models belong to the group of approximation-oriented models.

Another important method used to deal with uncertainty in intelligent systems is the Dempster-Shafer theory of evidence [27]. Shafer's belief and plausibility functions are constructed under the assumption that the focal elements in the belief structure are all crisp. In some situations, it seems to be quite natural that the evidence mass may be assigned to a fuzzy subset of the universe of discourse. In fact, combining the Dempster-Shafer theory and fuzzy set theory has been suggested to be a way to deal with different kinds of uncertain information in intelligent systems in a number of studies. It is demonstrated that the lower and upper approximation operators in rough set theory have strong relationship with the belief and plausibility functions in the Dempster-Shafer theory of evidence [21, 23, 28, 29, 30, 31, 32, 33]. The Dempster-Shafer theory of evidence may be used to analyze knowledge acquisition in information systems (see, for example, [34, 35, 36, 37, 38]).

In the second part of this chapter, we will explore the relationships between belief and plausibility functions in the Dempster-Shafer theory of evidence and the lower and upper approximations in rough set theory with their potential applications to intelligent information systems.

Both fuzzy set and rough set theories have fostered broad research communities and have been applied in a wide range of settings. More recently, this has extended also to the hybrid fuzzy rough set models. The third part of this chapter tries to give a sample of those applications, which are in particular numerous for machine learning but which also cover many other fields, like image processing, decision making and information retrieval.

Note that we do not consider applications that simply involve a joint application of fuzzy sets and rough sets, like for instance a rough classifier that induces fuzzy rules. Rather, we focus on applications that specifically involve one of the fuzzy rough set models discussed in the previous sections.

This chapter is organized as follows. In the next section, three models of fuzzy rough sets are explained dividing into two groups. In Section 3, we introduce generalized fuzzy belief structures with application in fuzzy information systems. In Section 4, we give an overview of the practical applications of fuzzy rough sets focusing on the machine learning domain.

2 Classification- versus Approximation-oriented Fuzzy Rough Set Models

2.1 Two views of the classical rough set model

In this section, we review three kinds of fuzzy rough sets from classification-oriented and approximation-oriented points of view. Focusing on the membership of an object to a set X under the indiscernibility relation, the classical rough set defined by a pair of lower and upper approximations of a set X can be seen as a classifier of objects into three disjoint regions: positive, negative and boundary regions of a set X. Namely, the lower approximation defines the positive region, the complement of the upper approximation defines the negative region and the difference between upper and lower approximations defines the boundary region. On the other hand, focusing on the approximations of X by means of elementary sets of the partition, the rough set of X defines the inner and outer approximations of X. Namely, the lower approximation defines the inner approximation and the upper approximation defines the outer approximation. Those two different views of rough sets give the different definitions of rough sets in the generalized settings (see Inuiguchi [50]). In this section, we describe fuzzy rough sets in a generalized setting from those points of view and show the fundamental properties, differences and similarities.

Throughout this chapter, U will be a nonempty set called the universe of discourse. The class of all subsets (respectively, fuzzy subsets) of U will be denoted by $\mathcal{P}(U)$ (respectively, by $\mathcal{F}(U)$). For any

$A \in \mathcal{F}(U)$, the complement of A will be denoted by $\sim A$, i.e. $(\sim A)(x) = 1 - A(x)$ for all $x \in U$.

2.2 Classification-oriented fuzzy rough sets

2.2.1 Definitions in crisp setting

In this subsection, we define fuzzy rough sets under the interpretation of rough sets as classification of objects into positive, negative and boundary regions of a set and describe their properties. As the introduction, we first describe the definitions of positive and possible regions of a set in the crisp setting. Let U be a set of all objects. Assume that we do not know objects which fit with a particular concept $\mathcal{C}$ but we have pieces of information that tell some objects fit with $\mathcal{C}$ and that the other objects do not fit. Let $X \subseteq U$ be the set of objects which are supposed to fit with $\mathcal{C}$ in the information and $U - X$ the set of objects which are supposed not to fit with $\mathcal{C}$ in the information. On the other hand, there is knowledge about $\mathcal{C}$ expressed by a binary relation $P \subseteq U \times U$. Under the binary relation P, we presume 'y fits with $\mathcal{C}$' from facts $(y, x) \in P$ and 'x fits with $\mathcal{C}$'.

Under this circumstance, we investigate credible members of X and plausible members of X. Objects whose membership to X is consistent with the knowledge can be understood as credible members of X, while objects whose membership to X is presumable from the information and the knowledge can be understood as plausible members. For convenience, we define $P(x) = \{y \in U \mid (y, x) \in P\}$ which is the set of objects whose membership to X is presumed from the fact $x \in X$. Therefore, if $x \in X$ satisfies $\forall y \in P(x),\ y \in X$ or simply, $P(x) \subseteq X$, x can be considered a credible member of X. Thus, the set of credible members of X is defined by

$$P_*(X) = \{x \in X \mid P(x) \subseteq X\} = X \cap \{x \in U \mid P(x) \subseteq X\}. \tag{1}$$

On the other hand, we may presume $x \in X$ if $x \in X$ or $\exists y \in X,\ x \in P(y)$ under the information and the knowledge. Then the set of plausible members of X can be defined by

$$P^*(X) = X \cup \{x \in U \mid \exists y \in X,\ x \in P(y) \neq \emptyset\}. \tag{2}$$

$P_*(X)$ is called the positive region of X and $P^*(X)$ is called the possible region of X. Moreover, we do not assume the reflexivity of P, i.e., $\forall x \in U, (x, x) \in P$. This is why we take the intersection with X in the definition of $P_*(X)$ and the union with X in the definition of $P^*(X)$. Those intersection and union can be dropped when P is reflexive.

When there is knowledge about $\mathcal{C}$ expressed by a binary relation $Q \subseteq U \times U$ instead of P. Under the binary relation Q, we presume 'y does not fit with $\mathcal{C}$' from facts $(y, x) \in Q$ and 'x does not fit with $\mathcal{C}$'. In this case, we directly obtain positive and possible regions of $U - X$ respectively by

$$\begin{aligned} Q_*(U - X) &= \{x \in U - X \mid Q(x) \subseteq U - X\} \\ &= (U - X) \cap \{x \in U \mid Q(x) \subseteq U - X\}, \qquad (3) \\ Q^*(U - X) &= (U - X) \cup \{x \in U \mid \exists y \in U - X,\ x \in Q(y) \neq \emptyset\}. \qquad (4) \end{aligned}$$

Because an object which is not a member of $Q_*(U - X)$ can be seen as a plausible member of X and an object which is not a member of $Q^*(U - X)$ can be seen as a credible member of X, we may define positive and possible regions of X by

$$\bar{Q}_*(X) = U - Q^*(U - X), \quad \bar{Q}^*(X) = U - Q_*(U - X). \tag{5}$$

Inuiguchi [50] investigated the properties of those positive and possible regions.

2.2.2 Definitions in fuzzy setting and their properties

We now extend those definitions of positive and possible regions into the fuzzy setting. First, we assume a fuzzy set $X \subseteq U$ and a fuzzy binary relation $P \in \mathcal{F}(U \times U)$ are given. Their membership functions $\mu_X(x)$ and $\mu_P(y, x)$ shows the membership degree of $x \in U$ to a fuzzy set X and the degree to what extent we presume that y is a member of X from the fact x is a member of a fuzzy set X, where $\mu_X : U \to [0, 1]$ and $\mu_P : U \times U \to [0, 1]$. We define $P(x)$ by its membership function $\mu_{P(x)}(y) = \mu_P(y, x)$.

To define the positive region under this circumstance, we should consider the consistency degree of the information that x is a member of X to membership degree $\mu_X(x)$ with the knowledge P. This can be measured by the truth value of statement '$y \in P(x)$ implies $y \in X$' under fuzzy sets $P(x)$ and X. The truth value of this statement can be defined by a necessity measure $\inf_{y \in U} I(\mu_{P(x)}(y), \mu_X(y))$ with an implication function $I : [0,1] \times [0,1] \to [0,1]$ such that $I(0,0) = I(0,1) = I(1,1) = 1$, $I(1,0) = 0$, $I(\cdot, a)$ is decreasing for any $a \in [0,1]$ and $I(a, \cdot)$ is increasing for any $a \in [0,1]$. Therefore, in the analogy to Eq. (1), the membership function of the positive region $P_*(X)$ of X can be defined by

$$\begin{aligned}\mu_{P_*(X)}(x) &= \min\left(\mu_X(x), \inf_{y \in U} I(\mu_{P(x)}(y), \mu_X(y))\right) \\ &= \min\left(\mu_X(x), \inf_{y \in U} I(\mu_P(y,x), \mu_X(y))\right),\end{aligned} \tag{6}$$

where we note the intersection $C \cap D$ of two fuzzy sets $C, D \in \mathcal{F}(U)$ is normally defined by $\mu_{C \cap D}(x) = \min(\mu_C(x), \mu_D(x))$, $\forall x \in U$. $\mu_{C \cap D}$, μ_C and μ_D are membership functions of $C \cap D$, C and D, respectively. However some researchers use t-norms [51] instead of the min operation. A t-norm t is a conjunction function $t : [0,1] \times [0,1] \to [0,1]$ such that (t1) $\forall a \in [0,1],\ t(a,1) = t(1,a) = a$ (boundary condition), (t2) $\forall a, b \in [0,1],\ t(a,b) = t(b,a)$ (commutativity) and (t3) $\forall a, b, c \in [0,1],\ t(a, t(b,c)) = t(t(a,b), c)$ (associativity).

Now let us define the possible region when X and P are a fuzzy set and a fuzzy binary relation, respectively. To do this, we should define the truth value of statement 'there exists $y \in X$ such that $x \in P(y)$' under fuzzy sets X and $P(x)$. The truth value of this statement can be obtained by a possibility measure $\sup_{y \in U} T(\mu_{P(y)}(x), \mu_X(y))$ with a conjunction function $T : [0,1] \times [0,1] \to [0,1]$ such that $T(1,1) = 1$, $T(0,0) = T(0,1) = T(1,0) = 0$ and T is increasing in both arguments. Therefore, in the analogy to Eq. (2), the membership function of the possible region $P^*(X)$ of X can be defined by

$$\begin{aligned}\mu_{P^*(X)}(x) &= \max\left(\mu_X(x), \sup_{y \in U} T(\mu_{P(y)}(x), \mu_X(y))\right) \\ &= \max\left(\mu_X(x), \sup_{y \in U} T(\mu_P(x,y), \mu_X(y))\right),\end{aligned} \tag{7}$$

where we note the union $C \cup D$ of two fuzzy sets $C, D \subseteq U$ is normally defined by $\mu_{C \cup D}(x) = \max(\mu_C(x), \mu_D(x))$, $\forall x \in U$. $\mu_{C \cup D}$ is a membership functions of $C \cup D$. However some researchers use t-conorms [51] instead of the max operation. A t-conorm s is a function $s : [0,1] \times [0,1] \to [0,1]$ such that (s1) $\forall a \in [0,1],\ t(a,0) = t(0,a) = a$ (boundary condition), (s2) $\forall a, b \in [0,1],\ s(a,b) = s(b,a)$ (commutativity), (s3) $\forall a, b, c \in [0,1],\ s(a, s(b,c)) = s(s(a,b), c)$ (associativity), and (s4) $\forall a, b, c, d$ such that $a \geq c$ and $b \geq d$; $s(a,b) \geq s(c,d)$ (monotonicity).

Note that we do not assume the reflexivity of P, i.e., $\mu_P(x,x) = 1$, $\forall x \in U$, so that we take the minimum between μ_X and $\inf_{y \in U} I(\mu_{P(x)}(y), \mu_X(y))$ in Eq. (6) and the maximum between μ_X and $\sup_{y \in U} T(\mu_{P(y)}(x), \mu_X(y))$ in Eq. (7). When P is reflexive, $I(1,a) \leq a$, and $T(1,a) \geq a$ for all $a \in [0,1]$, we have

$$\mu_{P_*(X)}(x) = \inf_{y \in U} I(\mu_{P(x)}(y), \mu_X(y)), \quad \mu_{P^*(X)}(x) = \sup_{y \in U} T(\mu_{P(y)}(x), \mu_X(y)). \tag{8}$$

Those definitions of lower and upper approximations have been proposed by Dubois and Prade [4, 5] and Radzikowska and Kerre [9]. They assumed the reflexivity of P and $I(1,a) = T(1,a) = a$, for all $a \in [0,1]$. Moreover, the definitions in Eq. (8) are used even when P is not reflexive and neither I nor T satisfy the boundary conditions $I(1,a) = T(1,a) = a$, for all $a \in [0,1]$ (see [**?**, 15]). In such generalized situation, we may loose the inclusiveness of $P_*(X)$ in X and that of X in $P^*(X)$ for $P_*(X)$ and $P^*(X)$ defined by Eq. (8). The definitions of $P_*(X)$ and $P^*(X)$ by Eqs. (6) and (7) obtained from the interpretations of positive and possible regions of X satisfy the inclusiveness of $P_*(X)$ in X and that of X in $P^*(X)$ even in the generalized situation.

Using the positive region $P_*(X)$ and the possible region $P^*(X)$, we can define a fuzzy rough set of X as a pair $(P_*(X), P^*(X))$. We can call such fuzzy rough sets as classification-oriented fuzzy rough sets under a positively extensive relation P of X (for short CP-fuzzy rough sets). Note that the relation P depends on the meaning of a set X. Thus, we cannot always define the CP-rough set of $U - X$ by the same relation P.

To define a CP-rough set of $U-X$, we should introduce another fuzzy relation $Q \in \mathcal{F}(U \times U)$ such that $\mu_{Q(x)}(y) = \mu_Q(y,x)$ represents the degree to what extent we presume an object y as a member of $U-X$ from the fact x is a member of $U-X$, where $\mu_Q : U \times U \to [0,1]$ is a membership function of a fuzzy relation Q. In the same way, we define positive and possible regions of $U-X$ under the fuzzy relation Q by the following membership functions:

$$\mu_{Q_*(U-X)}(x) = \min\biggl(n(\mu_X(x)), \inf_{y\in U} I(\mu_Q(y,x), n(\mu_X(y)))\biggr), \tag{9}$$

$$\mu_{Q^*(U-X)}(x) = \max\biggl(n(\mu_X(x)), \sup_{y\in U} T(\mu_P(x,y), n(\mu_X(y)))\biggr), \tag{10}$$

where $U-X$ is defined by a membership function $n(\mu_X(\cdot))$ and $n : [0,1] \to [0,1]$ is a strong negation which is a decreasing function such that $n(n(a)) = a$, for all $a \in [0,1]$ (involutive). The involution implies the continuity of n.

Using $Q_*(X)$ and $Q^*(X)$, in analogy to Eq. (5), we can define positive region $\bar{Q}_*(X)$ and possible region $\bar{Q}^*(X)$ of X by the following membership functions,

$$\mu_{\bar{Q}_*(X)}(x) = \min\biggl(\mu_X(x), \inf_{y\in U} n(T(\mu_Q(x,y), n(\mu_X(y))))\biggr), \tag{11}$$

$$\mu_{\bar{Q}^*(X)}(x) = \max\biggl(\mu_X(x), \sup_{y\in U} n(I(\mu_Q(y,x), n(\mu_X(y))))\biggr). \tag{12}$$

We can define a fuzzy rough set of X as a pair $(\bar{Q}_*(X), \bar{Q}^*(X))$ with the positive region $\bar{Q}_*(X)$ and the possible region $\bar{Q}^*(X)$. We can call this type of rough sets as classification-oriented rough sets under a negatively extensive relation Q of X (for short CN-fuzzy rough sets).

Let us discuss the properties of CP- and CN-fuzzy rough sets. By definition, we have

$$P_*(X) \subseteq X \subseteq P^*(X), \quad \bar{Q}_*(X) \subseteq X \subseteq \bar{Q}^*(X), \tag{13}$$

$$P_*(\emptyset) = P^*(\emptyset) = \bar{Q}_*(\emptyset) = \bar{Q}^*(\emptyset) = \emptyset, \tag{14}$$

$$P_*(U) = P^*(U) = \bar{Q}_*(U) = \bar{Q}^*(U) = U, \tag{15}$$

$$P_*(X \cap Y) = P_*(X) \cap P_*(Y), \quad P^*(X \cup Y) = P^*(X) \cup P^*(Y), \tag{16}$$

$$\bar{Q}_*(X \cap Y) = \bar{Q}_*(X) \cap \bar{Q}_*(Y), \quad \bar{Q}^*(X \cup Y) = \bar{Q}^*(X) \cup \bar{Q}^*(Y), \tag{17}$$

$$X \subseteq Y \text{ implies } P_*(X) \subseteq P_*(Y), \quad X \subseteq Y \text{ implies } P^*(X) \subseteq P^*(Y), \tag{18}$$

$$X \subseteq Y \text{ implies } \bar{Q}_*(X) \subseteq \bar{Q}_*(Y), \quad X \subseteq Y \text{ implies } \bar{Q}^*(X) \subseteq \bar{Q}^*(Y), \tag{19}$$

$$P_*(X \cup Y) \supseteq P_*(X) \cup P_*(Y), \quad P^*(X \cap Y) \subseteq P^*(X) \cap P^*(Y), \tag{20}$$

$$\bar{Q}_*(X \cup Y) \supseteq \bar{Q}_*(X) \cup \bar{Q}_*(Y), \quad \bar{Q}^*(X \cap Y) \subseteq \bar{Q}^*(X) \cap \bar{Q}^*(Y), \tag{21}$$

where the inclusion relation between two fuzzy sets A and B is defined by $\mu_A(x) \le \mu_B(x)$, for all $x \in U$.

The properties satisfied under some conditions are listed as follows (see Inuiguchi [26]):

⟨1⟩ When $I(a,b) = n(T(a, n(b)))$, for all $a, b \in [0,1]$ and Q is the converse of P, i.e., $\mu_Q(x,y) = \mu_P(y,x)$, for all $x, y \in U$, we have

$$P_*(X) = U - Q^*(U-X) = \bar{Q}_*(X), \tag{22}$$

$$P^*(X) = U - Q_*(U-X) = \bar{Q}^*(X). \tag{23}$$

⟨2⟩ When $T(a, I(a,b)) \le b$ holds for all $a, b \in [0,1]$, we have

$$X \supseteq P^*(P_*(X)) \supseteq P_*(X) \supseteq P_*(P_*(X)), \tag{24}$$

$$X \subseteq \bar{Q}_*(\bar{Q}^*(X)) \subseteq \bar{Q}^*(X) \subseteq \bar{Q}^*(\bar{Q}^*(X)). \tag{25}$$

⟨3⟩ When $I(a, T(a,b)) \ge b$ holds for all $a, b \in [0,1]$, we have

$$X \subseteq P_*(P^*(X)) \subseteq P^*(X) \subseteq P^*(P^*(X)), \tag{26}$$

$$X \supseteq \bar{Q}^*(\bar{Q}_*(X)) \supseteq \bar{Q}_*(X) \supseteq \bar{Q}_*(\bar{Q}_*(X)). \tag{27}$$

⟨4⟩ Let P and Q be T'-transitive. The following assertions are valid.

(a) When I is upper semi-continuous and satisfies $I(a, I(b, c)) = I(T'(b, a), c)$ for all $a, b, c \in [0, 1]$, we have

$$P_*(P_*(X)) = P_*(X), \quad \bar{Q}^*(\bar{Q}^*(X)) = \bar{Q}^*(X). \tag{28}$$

(b) When $T = T'$ is lower semi-continuous and satisfies $T(a, T(b, c)) = T(T(a,\ b), c)$ for all $a, b, c \in [0, 1]$ (associativity), we have

$$P^*(P^*(X)) = P^*(X), \quad \bar{Q}_*(\bar{Q}_*(X)) = \bar{Q}_*(X). \tag{29}$$

⟨5⟩ When P and Q are reflexive and T-transitive, the following assertions are valid:

(a) If $I(a, \cdot)$ is upper semi-continuous, $I(1, a) \leq a$, and $T = \xi[I]$ is associative, then we have

$$P^*(P_*(X)) = P_*(X), \quad \bar{Q}_*(\bar{Q}^*(X)) = \bar{Q}^*(X). \tag{30}$$

(b) If $I(a, b) = n(\xi[I](a, n(b)))$ and the conditions of (a) are satisfied, then we have

$$P_*(P^*(X)) = P^*(X), \quad \bar{Q}^*(\bar{Q}_*(X)) = \bar{Q}_*(X). \tag{31}$$

Here a fuzzy relation P is said to be T'-transitive, if and only if P satisfies $\mu_P(x, z) \geq T'(\mu_P(x, y), \mu_P(y, z))$ for all $x, y, z \in U$ and for a conjunction function T'. We can generate a function $\xi[I] : [0, 1] \times [0, 1] \to [0, 1]$ by $\xi[I](a, b) = \inf\{s \in [0, 1] \mid I(a, s) \geq b\}$ when a function $I : [0, 1] \times [0, 1] \to [0, 1]$ is given. $\xi[I]$ is a conjunction function when I satisfies $I(1, a) < 1$ for all $a \in [0, 1)$.

Concerning to the assumption of ⟨1⟩, it is known that a function I' defined by $I'(a, b) = n(T(a, n(b)))$ is an implication function and that a function T' defined by $T'(a, b) = n(I(a, n(b)))$ is a conjunction function (see, for example, Inuiguchi and Sakawa [51, 52]). The assumption of ⟨2⟩ corresponds to modus ponens, i.e., A and $(A \to B)$ implies B. Therefore, it is a natural assumption. However, this cannot hold for any implication and conjunction functions. For example, consider functions $T(a, b) = \min(a, b)$ and $I(a, b) = \max(1 - a, b)$ which are often used in possibility theory. $T(a, I(a, b)) \leq b$ does not always hold. On the other hand, the assumption holds for any T and I such that $T(a, b) \leq \min(a, b)$ for all $a, b \in [0, 1]$ and $I(a, b) \leq b$ for all $a, b \in [0, 1]$ satisfying $a > b$. Thus, a t-norm T and a residual implication I of a t-norm satisfies the assumption, i.e., I is defined by $I(a, b) = \sup\{s \in [0, 1] \mid t(a, s) \leq b\}$, for $a, b \in [0, 1]$. The assumption of ⟨3⟩ is dual with that of ⟨2⟩. Namely, for any implication function I, there exists a conjunction function T' such that $I(a, b) = n(T'(a, n(b)))$, and for any conjunction function T, there exists an implication function I' such that $T(a, b) = n(I'(a, n(b)))$. Using T' and I', the assumption $I(a, T(a, b)) \geq b$ is equivalent to $T'(a, I'(a, b)) \leq b$ which is the same as the assumption of ⟨2⟩.

The assumption of ⟨3⟩ is satisfied with I and T such that $I(a, b) \geq \max(n(a), b)$ for all $a, b \in [0, 1]$ and $T(a, b) \geq b$ for all $a, b \in [0, 1]$ satisfying $a > n(b)$. The assumption of ⟨4⟩-(a) is satisfied with residual implication functions of lower semi-continuous t-norms T' and S-implication functions with respect to lower semi-continuous t-norms T', where an S-implication function I with respect to t-norm T' is defined by $I(a, b) = n(T'(a, n(b)))$, $a, b \in [0, 1]$ with a strong negation n. The assumption of ⟨4⟩-(b) is satisfied with lower semi-continuous t-norms T. These assumptions are satisfied with a lot of famous implication and conjunction functions.

2.3 Approximation-oriented fuzzy rough sets

2.3.1 Definitions in crisp setting

In this subsection, we define fuzzy rough sets under the interpretation of rough sets as approximation of sets and describe their properties. We first describe the definitions of lower and upper approximations in the crisp setting. We assume a family of subsets in U, $\mathcal{F} = \{F_i \mid i = 1, 2, \ldots, p\}$ is given. Each elementary set F_i is a meaningful set of objects such as a set of objects satisfying some properties. F_i's can be seen as information granules with which we would like to express a set of objects. Given a set $X \subseteq U$, an understated expression of X, or in other words, an inner approximation of X by means of unions of F_i's is obtained by

$$\mathcal{F}_*^{\cup}(X) = \bigcup\{F_i \in \mathcal{F} \mid F_i \subseteq X\}. \tag{32}$$

On the other hand, an overstated expression of X, or in other words, an outer approximation of X by means of unions of F_i's is obtained by

$$\mathcal{F}_{\cup}^*(X) = \bigcap\left\{\bigcup_{i\in J} F_i \;\middle|\; \bigcup_{i\in J} F_i \supseteq X,\ J \subseteq \{1,2,\ldots,p,\circ\}\right\}, \tag{33}$$

where we define $F_\circ = U$. We add $F_\circ$ considering cases where there is no $J \subseteq \{1,2,\ldots,p\}$ such that $\bigcup_{i\in J} F_i \supseteq X$. In such cases, we obtain $\mathcal{F}_{\cup}^*(X) = U$ owing to the existence of $F_\circ = U$. $\mathcal{F}_*^{\cup}(X)$ and $\mathcal{F}_{\cup}^*(X)$ are called lower and upper approximations of X, respectively.

Applying those approximations to $U - X$, we obtain $\mathcal{F}_*^{\cup}(U-X)$ and $\mathcal{F}_{\cup}^*(U-X)$. From those, we obtain

$$\begin{aligned}\mathcal{F}_*^{\cap}(X) &= U - \mathcal{F}_{\cup}^*(U-X)\\ &= \bigcup\left\{U - \bigcup_{i\in J} F_i \;\middle|\; \bigcup_{i\in J} F_i \supseteq U - X,\ J \subseteq \{1,2,\ldots,p\}\right\}, \qquad (34)\\ \mathcal{F}_{\cap}^*(X) &= U - \mathcal{F}_*^{\cup}(U-X)\\ &= \bigcap\{U - F_i \mid F_i \subseteq U - X,\ i \in \{1,2,\ldots,p,\bullet\}\}, \qquad (35)\end{aligned}$$

where we define $F_\bullet = \emptyset$. We note that $\mathcal{F}_*^{\cap}(X)$ and $\mathcal{F}_{\cap}^*(X)$ are not always the same as $\mathcal{F}_*^{\cup}(X)$ and $\mathcal{F}_{\cup}^*(X)$, respectively. The properties of those lower and upper approximations are studied in Inuiguchi [50].

2.3.2 Definitions by certainty qualifications in fuzzy setting

We extend those lower and upper approximations to cases where $\mathcal{F}$ is a family of fuzzy sets in U and X is a fuzzy set in U. To do this, we extend the intersection, union, complement and the inclusion relation into the fuzzy setting. The intersection, union and complement are defined by the min operation, the max operation and a strong negation n, i.e., $C \cap D$, $C \cup D$ and $U - C$ for fuzzy sets C and D are defined by membership functions $\mu_{C\cap D}(x) = \min(\mu_C(x), \mu_D(x))$, $\forall x \in C$, $\mu_{C\cup D}(x) = \max(\mu_C(x), \mu_D(x))$, $\forall x \in C$, $\mu_{U-C}(x) = n(\mu_C(x))$, $\forall x \in C$, respectively. The inclusion relation $C \subseteq D$ is extended to inclusion relation with degree $Inc(C, D) = \inf_x I(\mu_C(x), \mu_D(x))$, where I is an implication function.

First let us define a lower approximation by extending Eq. (32). In Eq. (32), before applying the union, we collect F_i such that $F_i \subseteq X$. This procedure cannot be extended simply into the fuzzy setting, because the inclusion relation has a degree showing to what extent the inclusion holds in the fuzzy setting. Namely, each F_i has a degree $q_i = Inc(F_i, X)$. This means that X includes F_i to a degree q_i. Therefore, by using F_i, X is expressed as a fuzzy set including F_i to a degree q_i. In other words, X is a fuzzy set Y satisfying

$$Inc(F_i, Y) = \inf_x I(\mu_{F_i}(x), \mu_Y(x)) = q_i. \tag{36}$$

We note that there exists a solution satisfying Eq. (36) because q_i is defined by $Inc(F_i, X)$. There can be many solutions Y satisfying Eq. (36) and the intersection and union of those solutions can be seen as inner and outer approximations of X by F_i. Because we are now extending Eq. (32) and interested in the lower approximation, we consider the intersection of fuzzy sets including F_i to a degree q_i. Let us consider

$$Inc(F_i, Y) = \inf_x I(\mu_{F_i}(x), \mu_Y(x)) \geq q_i, \tag{37}$$

instead of Eq. (36). Eq. (37) is called a converse-certainty qualification [10] (or possibility-qualification). Because $I(a, \cdot)$ is increasing for any $a \in [0,1]$ for an implication function I and Eq. (36) has a solution, the intersection of solutions of Eq. (36) is the same as the intersection of solutions of Eq. (37). Moreover, because I is upper semi-continuous, we obtain the intersection of solutions of Eq. (37) as the smallest solution $\check{Y}$ of Eq. (37) defined by the following membership function:

$$\mu_{\check{Y}}(x) = \inf\{s \in [0,1] \mid I(\mu_{F_i}(x), s) \geq q_i\} = \xi[I](\mu_{F_i}(x), q_i). \tag{38}$$

We have $\check{Y} \subseteq X$. Because we have many $F_i \in \mathcal{F}$, the lower approximation $\mathcal{F}_*^\xi(X)$ of X is defined by the following membership function:

$$\mu_{\mathcal{F}_*^\xi(X)}(x) = \sup_{F \in \mathcal{F}} \xi[I] \left(\mu_F(x), \inf_{y \in U} I(\mu_F(y), \mu_X(y)) \right), \tag{39}$$

where $\mathcal{F}$ can have infinitely many elementary fuzzy sets F.

Because Eq. (32) is extended to Eq. (39), Eq. (35) is extended to the following equation in the sense that $\mathcal{F}_\xi^*(X) = U - F_*^\xi(U - X)$:

$$\mu_{\mathcal{F}_\xi^*(X)}(x) = \inf_{F \in \mathcal{F}} n \left(\xi[I] \left(\mu_F(x), \inf_{y \in U} I \left(\mu_F(y), n(\mu_X(y)) \right) \right) \right), \tag{40}$$

where $\mu_{\mathcal{F}_\xi^*(X)}$ is the membership function of the upper approximation $\mathcal{F}_\xi^*(X)$ of X.

Now let us consider the extension of Eq. (33). In this case, before applying the intersection, we collect $\bigcup_{i \in J} F_i$ such that $\bigcup_{i \in J} F_i \supseteq X$. In the fuzzy setting, each $\bigcup_{i \in J} F_i$ has a degree $r_J = Inc(X, \bigcup_{i \in J} F_i)$. This means that X is included in $\bigcup_{i \in J} F_i$ to a degree r_J. Therefore, by using F_i, $i \in J$, X is expressed as a fuzzy set included in $\bigcup_{i \in J} F_i$ to a degree r_J. In other words, X is a fuzzy set Y satisfying

$$Inc \left(Y, \bigcup_{i \in J} F_i \right) = \inf_x I \left(\mu_Y(x), \max_{i \in J} \mu_{F_i}(x) \right) = r_J. \tag{41}$$

We note that there exists a solution satisfying Eq. (41) because r_J is defined by $Inc(X, \bigcup_{i \in J} F_i)$. There can be many solutions Y satisfying Eq. (41) and the intersection and union of those solutions can be seen as inner and outer approximations of X by $\bigcup_{i \in J} F_i$. Because we are now extending Eq. (33) and interested in the upper approximation, we consider the union of fuzzy sets including $\bigcup_{i \in J} F_i$ to a degree r_J. Let us consider

$$Inc \left(Y, \bigcup_{i \in J} F_i \right) = \inf_x I \left(\mu_Y(x), \max_{i \in J} \mu_{F_i}(x) \right) \geq r_J, \tag{42}$$

instead of Eq. (41). Eq. (42) is called a certainty qualification [10, 57]. Because $I(\cdot, a)$ is decreasing for any $a \in [0,1]$ for an implication function I and Eq. (41) has a solution, the union of solutions of Eq. (41) is the same as the union of solutions of Eq. (42). Moreover, because I is upper semi-continuous, we obtain the union of solutions of Eq. (42) as the greatest solution $\hat{Y}$ of Eq. (42) defined by the following membership function:

$$\mu_{\hat{Y}}(x) = \sup \left\{ s \in [0,1] \;\middle|\; I \left(s, \max_{i \in J} \mu_{F_i}(x) \right) \geq q_i \right\} = \sigma[I] \left(r_J, \max_{i \in J} \mu_{F_i}(x) \right), \tag{43}$$

where we define $\sigma[I](a,b) = \sup\{s \in [0,1] \mid I(s,b) \geq a\}$ for $a, b \in [0,1]$. We have $X \subseteq \hat{Y}$. Because we have many $\bigcup_{i \in J} F_i$, the upper approximation $\mathcal{F}_\sigma^*(X)$ of X is defined by the following membership function:

$$\mu_{\mathcal{F}_\sigma^*(X)}(x) = \inf_{\mathcal{T} \subseteq \mathcal{F}} \sigma[I] \left(\inf_{y \in U} I \left(\mu_X(y), \sup_{F \in \mathcal{T}} \mu_F(y) \right), \sup_{F \in \mathcal{T}} \mu_F(x) \right), \tag{44}$$

where $\mathcal{F}$ can have infinitely many elementary fuzzy sets F. We note that $\sigma[I]$ becomes an implication function.

Because Eq. (33) is extended to Eq. (44), Eq. (34) is extended to the following equation in the sense that $\mathcal{F}_*^\sigma(X) = U - F_\sigma^*(U - X)$:

$$\mu_{\mathcal{F}_*^\sigma(X)}(x) = \sup_{\mathcal{T} \subseteq \mathcal{F}} n \left(\sigma[I] \left(\inf_{y \in U} I \left(n(\mu_X(y)), \sup_{F \in \mathcal{T}} \mu_F(y) \right), \sup_{F \in \mathcal{T}} \mu_F(x) \right) \right), \tag{45}$$

where $\mu_{\mathcal{F}_\sigma^*(X)}$ is the membership function of the upper approximation $\mathcal{F}_\sigma^*(X)$ of X.

These four approximations were originally proposed by Inuiguchi and Tanino [10]. They selected a pair $(\mathcal{F}_*^\xi(X), \mathcal{F}_\xi^*(X))$ to define a rough set of X. However, in connection with the crisp case, Inuiguchi [26] selected pairs $(\mathcal{F}_*^\xi(X), \mathcal{F}_\sigma^*(X))$ and $(\mathcal{F}_*^\sigma(X), \mathcal{F}_\xi^*(X))$ for the definitions of rough sets of X. In this paper, a pair $(\mathcal{F}_*^\xi(X), \mathcal{F}_\xi^*(X))$ is called a ξ-fuzzy rough set and a pair $(\mathcal{F}_*^\sigma(X), \mathcal{F}_\sigma^*(X))$ a σ-fuzzy rough set.

2.3.3 Properties

First we show properties about the representations of lower and upper approximations defined by Eqs. (39), (40), (44) and (45). We have the following equalities (see Inuiguchi [26]):

$$\mu_{\mathcal{F}_*^\xi(X)}(x) = \sup\{\xi[I](\mu_F(x), h) \mid F \in \mathcal{F}, h \in [0,1] \text{ such that } \xi[I](\mu_F(y), h) \leq \mu_X(y), \forall y \in U\}, \tag{46}$$

$$\mu_{\mathcal{F}_*^\sigma(X)}(x) = \sup\left\{n\left(\sigma[I]\left(h, \sup_{F\in\mathcal{T}} \mu_F(x)\right)\right) \,\middle|\, \mathcal{T} \subseteq \mathcal{F}, h \in [0,1] \text{ such that } \sigma[I]\left(h, \sup_{F\in\mathcal{T}} \mu_F(y)\right) \geq n(\mu_X(y)), \forall y \in U\right\}, \tag{47}$$

$$\mu_{\mathcal{F}_\xi^*(X)}(x) = \inf\{n(\xi[I](\mu_F(x), h)) \mid F \in \mathcal{F}, h \in [0,1] \text{ such that } \xi[I](\mu_F(y), h) \leq n(\mu_X(y)), \forall y \in U\}, \tag{48}$$

$$\mu_{\mathcal{F}_\sigma^*(X)}(x) = \inf\left\{\sigma[I]\left(h, \sup_{F\in\mathcal{T}} \mu_F(x)\right) \,\middle|\, \mathcal{T} \subseteq \mathcal{F}, h \in [0,1] \text{ such that } \sigma[I]\left(h, \sup_{F\in\mathcal{T}} \mu_F(y)\right) \geq \mu_X(y), \forall y \in U\right\}. \tag{49}$$

Using these equations, the following properties can be obtained easily:

$$\mathcal{F}_*^\xi(X) \subseteq X \subseteq \mathcal{F}_\xi^*(X), \quad \mathcal{F}_*^\sigma(X) \subseteq X \subseteq \mathcal{F}_\sigma^*(X), \tag{50}$$

$$\mathcal{F}_*^\xi(\emptyset) = \mathcal{F}_*^\sigma(\emptyset) = \emptyset, \quad \mathcal{F}_\xi^*(U) = \mathcal{F}_\sigma^*(U) = U, \tag{51}$$

$$X \subseteq Y \text{ implies } \mathcal{F}_*^\xi(X) \subseteq \mathcal{F}_*^\xi(Y), \quad X \subseteq Y \text{ implies } \mathcal{F}_*^\sigma(X) \subseteq \mathcal{F}_*^\sigma(Y), \tag{52}$$

$$X \subseteq Y \text{ implies } \mathcal{F}_\xi^*(X) \subseteq \mathcal{F}_\xi^*(Y), \quad X \subseteq Y \text{ implies } \mathcal{F}_\sigma^*(X) \subseteq \mathcal{F}_\sigma^*(Y), \tag{53}$$

$$\mathcal{F}_*^\xi(X \cup Y) \supseteq \mathcal{F}_*^\xi(X) \cup \mathcal{F}_*^\xi(Y), \quad \mathcal{F}_*^\sigma(X \cup Y) \supseteq \mathcal{F}_*^\sigma(X) \cup \mathcal{F}_*^\sigma(Y), \tag{54}$$

$$\mathcal{F}_\xi^*(X \cap Y) \subseteq \mathcal{F}_\xi^*(X) \cap \mathcal{F}_\xi^*(Y), \quad \mathcal{F}_\sigma^*(X \cap Y) \subseteq \mathcal{F}_\sigma^*(X) \cap \mathcal{F}_\sigma^*(Y), \tag{55}$$

$$\mathcal{F}_*^\xi(U - X) = U - \mathcal{F}_\xi^*(X), \quad \mathcal{F}_*^\sigma(U - X) = U - \mathcal{F}_\sigma^*(X). \tag{56}$$

Furthermore, we can prove the following properties (see Inuiguchi [26]):

⟨7⟩ The following assertions are valid:

(a) If $a > 0$, $b < 1$ imply $I(a,b) < 1$ and $\inf_{x\in U} \sup_{F\in\mathcal{F}} \mu_F(x) > 0$, then we have $\mathcal{F}_*^\xi(U) = U$ and $\mathcal{F}_\xi^*(\emptyset) = \emptyset$.

(b) If $b < 1$ implies $I(1,b) < 1$ and $\inf_{x\in U} \sup_{F\in\mathcal{F}} \mu_F(x) = 1$, then we have $\mathcal{F}_*^\xi(U) = U$ and $\mathcal{F}_\xi^*(\emptyset) = \emptyset$.

(c) If $a > 0$, $b < 1$ imply $I(a,b) < 1$ and $\forall x \in U$, $\exists F \in \mathcal{F}$ such that $\mu_F(x) < 1$, then we have $\mathcal{F}_*^\sigma(U) = U$ and $\mathcal{F}_\sigma^*(\emptyset) = \emptyset$.

(d) If $a > 0$ implies $I(a,0) < 1$ and $\forall x \in U$, $\exists F \in \mathcal{F}$ such that $\mu_F(x) = 0$, then we have $\mathcal{F}_*^\sigma(U) = U$ and $\mathcal{F}_\sigma^*(\emptyset) = \emptyset$.

⟨8⟩ We have

$$\mathcal{F}_*^\sigma(X \cap Y) = \mathcal{F}_*^\sigma(X) \cap \mathcal{F}_*^\sigma(Y), \quad \mathcal{F}_\sigma^*(X \cup Y) = \mathcal{F}_\sigma^*(X) \cup \mathcal{F}_\sigma^*(Y). \tag{57}$$

Moreover, if $\forall a \in [0,1]$, $I(a,a) = 1$ and $\forall F_i, F_j \in \mathcal{F}$, $F_i \neq F_j$, $F_i \cap F_j = \emptyset$, we have

$$\mathcal{F}_*^\xi(X \cap Y) = \mathcal{F}_*^\xi(X) \cap \mathcal{F}_*^\xi(Y), \quad \mathcal{F}_\xi^*(X \cup Y) = \mathcal{F}_\xi^*(X) \cup \mathcal{F}_\xi^*(Y). \tag{58}$$

⟨9⟩ We have

$$\mathcal{F}_*^\xi(\mathcal{F}_*^\xi(X)) = \mathcal{F}_*^\xi(X), \quad \mathcal{F}_*^\sigma(\mathcal{F}_*^\sigma(X)) = \mathcal{F}_*^\sigma(X), \tag{59}$$

$$\mathcal{F}_\xi^*(\mathcal{F}_\xi^*(X)) = \mathcal{F}_\xi^*(X), \quad \mathcal{F}_\sigma^*(\mathcal{F}_\sigma^*(X)) = \mathcal{F}_\sigma^*(X). \tag{60}$$

Inuiguchi and Tanino [10] proposed first this type of fuzzy rough sets. They demonstrated the advantage in approximation when P is reflexive and symmetric, I is Dienes implication and T is minimum operation. Inuiguchi and Tanino [58] showed that by selection of a necessity measure expressible various inclusion situations, the approximations become better, i.e., the differences between lower and upper approximations satisfying Eq. (58) become smaller. Moreover, Inuiguchi and Tanino [59] applied these fuzzy rough sets to function approximation.

2.4 Relations between two kinds of fuzzy rough sets

Under given fuzzy relations P and Q described in subsection 2.2, we discuss the relations between two kinds of fuzzy rough sets. Families of fuzzy sets are defined by $\mathcal{P} = \{P(x) | \ x \in U\}$ and $\mathcal{Q} = \{Q(x) | \ x \in U\}$. We have the following assertions.

⟨10⟩ When P and Q are reflexive, $I(1, a) = a$, we have

$$P_*(X) \subseteq \mathcal{P}^\xi_*(X), \quad \mathcal{Q}^*_\xi(X) \subseteq \bar{Q}^*(X). \tag{61}$$

⟨11⟩ When P and Q are reflexive, X is a crisp set, $a \le b$ if and only if $I(a, b) = 1$ and $T(a, 1) = a$ for all $a \in [0, 1]$, we have

$$\mathcal{P}^*_\sigma(X) \subseteq P^*(X), \quad \bar{Q}_*(X) \subseteq \mathcal{Q}^\sigma_*(X). \tag{62}$$

⟨12⟩ When P and Q are T-transitive, the following assertions are valid:

(a) When $T = \xi[I]$ is associative, we have

$$\mathcal{P}^\xi_*(X) \subseteq P_*(X), \quad \bar{Q}^*(X) \subseteq \mathcal{Q}^*_\xi(X). \tag{63}$$

(b) When $T = \xi[\sigma[I]]$ and $\sigma[I](a, \sigma[I](b, c)) = \sigma[I](b, \sigma[I](a, c))$ for all $a, b, c \in [0, 1]$, we have

$$P^*(X) \subseteq \mathcal{P}^*_\sigma(X), \quad \mathcal{Q}^\sigma_*(X) \subseteq \bar{Q}_*(X). \tag{64}$$

Here we define $\zeta[T](a, b) = \sup\{s \in [0, 1] \mid T(a, s) \le b\}$. This functional ζ can produce an implication function from a conjunction function T. Note that $\zeta[\xi[I]] = I$ and $\xi[\zeta[T]] = I$ for upper semi-continuous I and lower semi-continuous T (see Inuiguchi and Sakawa [52]).

2.5 The other approximation-oriented fuzzy rough sets

Greco et al. [24, 25] proposed fuzzy rough sets corresponding to a gradual rule [53], 'the more an object is in G, the more it is in X' with fuzzy sets F and X. Corresponding to this gradual rule, we may define the lower approximation $G^+_*(X)$ of X and the upper approximation $G^*_+(X)$ of X respectively by the following membership functions

$$\mu_{G^+_*(X)}(x) = \inf\{\mu_X(z) \mid z \in U, \ \mu_G(z) \ge \mu_G(x)\}, \tag{65}$$

$$\mu_{G^*_+(X)}(x) = \sup\{\mu_X(z) \mid z \in U, \ \mu_G(z) \le \mu_G(x)\}. \tag{66}$$

When we have a gradual rule, 'the less an object is in G, the more it is in X', we define the lower approximation $G^-_*(X)$ of X and the upper approximation $G^*_-(X)$ of X respectively by the following membership functions

$$\mu_{G^-_*(X)}(x) = \inf\{\mu_X(z) \mid z \in U, \ \mu_G(z) \le \mu_G(x)\}, \tag{67}$$

$$\mu_{G^*_-(X)}(x) = \sup\{\mu_X(z) \mid z \in U, \ \mu_G(z) \ge \mu_G(x)\}. \tag{68}$$

Moreover, when a complex gradual rule, 'the more an object is in G^+ and the less it is in G^-, the more it is in X' is given, the lower approximation $\mathcal{G}^\pm_*(X)$ and upper approximation $\mathcal{G}^*_\pm(X)$ are defined respectively by the following equations:

$$\mu_{\mathcal{G}^\pm_*(X)}(x) = \inf\{\mu_X(z) \mid z \in U, \ \mu^+_G(z) \ge \mu^+_G(x), \ \mu^-_G(z) \le \mu^-_G(x)\}, \tag{69}$$

$$\mu_{\mathcal{G}^*_\pm(X)}(x) = \sup\{\mu_X(z) \mid z \in U, \ \mu^+_G(z) \le \mu^+_G(x), \ \mu^-_G(z) \ge \mu^-_G(x)\}, \tag{70}$$

where we define $\mathcal{G} = \{G^+, G^-\}$.

The fuzzy rough sets are defined by pairs of those lower and upper approximations. This approach is advantageous in (i) no logical connectives such as implication function, conjunction function, etc., are used and (ii) the fuzzy rough sets correspond to gradual rules (see Greco et al. [24, 25]). However, we need a background knowledge about the monotone properties between G (or $\mathcal{G}$) and X.

This approach can be seen from a view point of modifier functions of fuzzy sets. A modifier function φ is generally a function from $[0,1]$ to $[0,1]$ (see [54]). Functions defined by $\varphi_1(x) = x^2$, $\varphi_2(x) = \sqrt{x}$ and $\varphi_3(x) = 1 - x$ are known as modifier functions corresponding to modifying words 'very', 'more or less' and 'not'. Namely, given a fuzzy set A, we may define fuzzy sets 'very A', 'more or less A' and 'not-A' by the following membership functions:

$$\mu_{\text{very } A}(x) = (\mu_A(x))^2, \mu_{\text{more or less } A}(x) = \sqrt{\mu_A(x)}, \mu_{\text{not-}A}(x) = 1 - \mu_A(x). \tag{71}$$

Such modifier functions are often used in approximate/fuzzy reasoning [55, 56], especially in the indirect method of fuzzy reasoning which is called also, truth value space method.

Namely, we may define the lower approximation $\Phi_*(X)$ of X and the upper approximation $\Phi^*(X)$ of X by means of a fuzzy set G by the following membership functions:

$$\mu_{\Phi_*(X)}(x) = \varphi_*^{G\to X}(\mu_G(x)), \quad \mu_{\Phi^*(X)}(x) = \varphi^*_{G\to X}(\mu_G(x)), \tag{72}$$

where $\Phi = \{\varphi_*^{G\to X}, \varphi^*_{G\to X}\}$ and modifier functions $\varphi_*^{G\to X}$ and $\varphi^*_{G\to X}$ are selected to satisfy

$$\varphi_*^{F\to X}(\mu_G(x)) \le \mu_X(x), \quad \varphi^*_{F\to X}(\mu_G(x)) \le \mu_X(x), \ \forall x \in U. \tag{73}$$

Indeed, $\xi[I](\cdot, \inf_{y\in U} I(\mu_G(y), \mu_X(y))$ and $\sigma[I](\inf_{y\in U} I(\mu_G(y), \mu_X(y), \cdot)$ are modifier functions satisfying Eq. (73) and these are used to define $\mathcal{F}_*^\xi(X)$ and $\mathcal{F}_\sigma^*(X)$ in Eqs. (39) and (44), respectively. We note that we consider multiple fuzzy set $G = F \in \mathcal{F}$ in Eq. (39) and apply the union because we have $\xi[I](\mu_G, \inf_{y\in U} I(\mu_G(y),\ \mu_X(x)) \le \mu_X(x)$, $\forall x \in U$ for all $G \in \mathcal{F}$. Similarly we consider multiple fuzzy sets G defined by $\mu_G(x) = \sup_{F\in\mathcal{T}\subseteq\mathcal{F}} \mu_F(x)$, $x \in U$ in Eq. (44) and we apply the intersection because we have $\sigma[I](\inf_{y\in U} I(\mu_G(y), \mu_X(y), \mu_G(x)) \ge \mu_X(x)$, $\forall x \in U$ for all those fuzzy sets G.

In the definitions of Eqs. (65) to (68), the following modifier functions are used respectively:

$$\varphi_*^+(\alpha) = \sup\{\psi_*^+(\beta) \mid \beta \in [0,\alpha]\}, \quad \varphi_+^*(\alpha) = \inf\{\psi_+^*(\beta) \mid \beta \in [\alpha, 1]\}, \tag{74}$$

$$\varphi_*^-(\alpha) = \sup\{\psi_*^-(\beta) \mid \beta \in [\alpha,1]\}, \quad \varphi_-^*(\alpha) = \inf\{\psi_-^*(\beta) \mid \beta \in [0, \alpha]\}, \tag{75}$$

where we define

$$\psi_*^+(\beta) = \inf\{\mu_X(z) \mid z \in U,\ \mu_G(z) \ge \beta\}, \tag{76}$$

$$\psi_-^*(\alpha) = \sup\{\mu_X(z) \mid z \in U,\ \mu_G(z) \le \beta\}, \tag{77}$$

$$\psi_*^-(\alpha) = \inf\{\mu_X(z) \mid z \in U,\ \mu_G(z) \le \beta\}, \tag{78}$$

$$\psi_-^*(\alpha) = \sup\{\mu_X(z) \mid z \in U,\ \mu_G(z) \ge \beta\}, \tag{79}$$

with $\inf\emptyset = 0$ and $\sup\emptyset = 1$. We note that φ_*^+ and φ_+^* are monotonically increasing which φ_*^- and φ_-^* are monotonically decreasing. These monotonicities are imposed in order to fit the supposed gradual rules. However, such monotonicities do not hold for functions ψ_*^+, ψ_+^*, ψ_*^- and ψ_-^*. In the cases of Eqs. (69) and (70), we should extend the modifier function to a generalized modifier function which is a function from $[0,1] \times [0,1]$ to $[0,1]$ because we have two fuzzy sets in the premise of the corresponding gradual rule. The associated generalized modifier functions with Eqs. (69) and (70) are obtained as

$$\varphi_*^\pm(\alpha_1, \alpha_2) = \sup\{\psi_*^\pm(\beta_1, \beta_2) \mid \beta_1 \in [0, \alpha_1],\ \beta_2 \in [\alpha_2, 1]\}, \tag{80}$$

$$\varphi_\pm^*(\alpha_1, \alpha_2) = \sup\{\psi_\pm^*(\beta_1, \beta_2) \mid \beta_1 \in [\alpha_1, 1],\ \beta_2 \in [0, \alpha_2]\}, \tag{81}$$

where we define

$$\psi_*^\pm(\beta_1, \beta_2) = \inf\{\mu_X(z) \mid z \in U,\ \mu_G^+(z) \ge \beta_1,\ \mu_G^-(z) \le \beta_2\}, \tag{82}$$

$$\psi_\pm^*(\beta_1, \beta_2) = \sup\{\mu_X(z) \mid z \in U,\ \mu_G^+(z) \le \beta_1,\ \mu_G^-(z) \ge \beta_2\}, \tag{83}$$

with $\inf\emptyset = 0$ and $\sup\emptyset = 1$. We note $\varphi_*^{\pm}$ and $\varphi_{\pm}^*$ are monotonically increasing in the first argument and monotonically decreasing in the second argument.

Moreover when we do not have any background knowledge about the relation between G and X which is expressed by a gradual rule. We may define the lower approximation $G_*(X)$ and the upper approximation $G^*(X)$ by the following membership functions:

$$\mu_{G_*(X)}(x) = \inf\{\mu_X(z) \mid z \in U,\ \mu_G(z) = \mu_G(x)\}, \tag{84}$$
$$\mu_{G^*(X)}(x) = \sup\{\mu_X(z) \mid z \in U,\ \mu_G(z) = \mu_G(x)\}. \tag{85}$$

The modifier functions associate with these approximations are obtained as

$$\varphi_*(\alpha) = \inf\{\mu_X(z) \mid z \in U,\ \mu_G(z) = \alpha\}, \tag{86}$$
$$\varphi^*(\alpha) = \sup\{\mu_X(z) \mid z \in U,\ \mu_G(z) = \alpha\}. \tag{87}$$

where we define $\inf\emptyset = 0$ and $\sup\emptyset = 1$. Eq. (87) is the same as the inverse truth qualification [55, 56] of X based on G.

We describe the properties of the approximations defined by Eqs. (65) to (68). However the other approximations defined by Eqs. (69), (70), (84) and (85) have the similar results. We have the following properties for the approximations defined by Eqs. (65) to (68) (see Greco et al. [25] for a part of these properties):

$$G_*^+(X) \subseteq X \subseteq G_+^*(X), \quad G_*^-(X) \subseteq X \subseteq G_-^*(X), \tag{88}$$
$$G_*^+(\emptyset) = G_+^*(\emptyset) = G_*^-(\emptyset) = G_-^*(\emptyset) = \emptyset, \tag{89}$$
$$G_*^+(U) = G_+^*(U) = G_*^-(U) = G_-^*(U) = U, \tag{90}$$
$$G_*^+(X \cap Y) = G_*^+(X) \cap G_*^+(Y), \quad G_*^-(X \cap Y) = G_*^-(X) \cap G_*^-(Y), \tag{91}$$
$$G_+^*(X \cup Y) = G_+^*(X) \cup G_+^*(Y), \quad G_-^*(X \cup Y) = G_-^*(X) \cup G_-^*(Y), \tag{92}$$
$$X \subseteq Y \text{ implies } G_*^+(X) \subseteq G_*^+(Y), \quad X \subseteq Y \text{ implies } G_+^*(X) \subseteq G_+^*(Y), \tag{93}$$
$$X \subseteq Y \text{ implies } G_*^-(X) \subseteq G_*^-(Y), \quad X \subseteq Y \text{ implies } G_-^*(X) \subseteq G_-^*(Y), \tag{94}$$
$$G_*^+(X \cup Y) \supseteq G_*^+(X) \cup G_*^+(Y), \quad G_+^*(X \cap Y) \subseteq G_+^*(X) \cap G_+^*(Y), \tag{95}$$
$$G_*^-(X \cup Y) \supseteq G_*^-(X) \cup G_*^-(Y), \quad G_-^*(X \cap Y) \subseteq G_-^*(X) \cap G_-^*(Y), \tag{96}$$
$$G_*^+(U \backslash X) = U \backslash G_-^*(X) = U \backslash (U \backslash G)_+^*(X) = (U \backslash G)_*^-(U \backslash X), \tag{97}$$
$$G_*^-(U \backslash X) = U \backslash G_+^*(X) = U \backslash (U \backslash G)_-^*(X) = (U \backslash G)_*^+(U \backslash X), \tag{98}$$
$$G_*^+(G_*^+(X)) = G_+^*(G_*^+(X)) = G_*^+(X), \quad G_+^*(G_+^*(X)) = G_*^+(G_+^*(X)) = G_+^*(X), \tag{99}$$
$$G_*^-(G_*^-(X)) = G_-^*(G_*^-(X)) = G_*^-(X), \quad G_-^*(G_-^*(X)) = G_*^-(G_-^*(X)) = G_-^*(X), \tag{100}$$

where $U \backslash X$ is a fuzzy set defined by its membership function $\mu_{U\backslash X}(x) = N(\mu_X(x))$, $\forall x \in U$ with a strictly decreasing function $N : [0,1] \to [0,1]$. We found that all fundamental properties [2] of the classical rough set are preserved.

2.6 Remarks

Three types of fuzzy rough set models have been described, divided into two groups: classification-oriented and approximation-oriented models. The classification-oriented fuzzy rough set models are much more investigated by many researchers. However, the approximation-oriented fuzzy rough set models would be more important because they are associated with rules. While approximation-oriented fuzzy rough set models based on modifiers are strongly related to the gradual rules, approximation-oriented fuzzy rough set models based on certainty qualification have relations to uncertain generation rule (uncertain qualification rule: certainty rule and possibility rule) [57], i.e., a rule such as 'the more an object is in A, the more certain (possible) it is in B'. While approximation-oriented fuzzy rough set models based on modifiers need a modifier function for each granule G, the approximation-oriented fuzzy rough set models based on certainty qualification need only a degree of inclusion for each granule F. Therefore, the latter may work well for data compression such as image compression, speech compression and so on.

3 Generalized Fuzzy Belief Structures with Application in Fuzzy Information Systems

In rough set theory there exists a pair of approximation operators, the lower and upper approximations, whereas in the Dempster-Shafer theory of evidence there exists a dual pair of uncertainty measures, the belief and plausibility functions. In this section, general types of belief structures and their induced dual pairs of belief and plausibility functions are first introduced. Relationships between belief and plausibility functions in the Dempster-Shafer theory of evidence and the lower and upper approximations in rough set theory are then established. It is shown that the probabilities of lower and upper approximations induced from an approximation space yield a dual pair of belief and plausibility functions. And for any belief structure there must exist a probability approximation space such that the belief and plausibility functions defined by the given belief structure are, respectively, the lower and upper probabilities induced by the approximation space. The pair of lower and upper approximations of a set capture the non-numeric aspect of uncertainty of the set which can be interpreted as the qualitative representation of the set, whereas the pair of the belief and plausibility measures of the set capture the numeric aspect of uncertainty of the set which can be treated as the quantitative characterization of the set. Finally, the potential applications of the main results to intelligent information systems are explored.

3.1 Belief structures and belief functions

In this subsection we recall some basic notions related to belief structures with their induced belief and plausibility functions.

3.1.1 Belief and plausibility functions derived from a crisp belief structure

The basic representational structure in the Dempster-Shafer theory of evidence is a belief structure.

Definition 1 *Let U be a non-empty set which may be infinite, a set function $m : \mathcal{P}(U) \to [0,1]$ is referred to as a crisp basic probability assignment if it satisfies axioms* (M1) *and* (M2)*:*

(M1) $m(\emptyset) = 0$, (M2) $\sum\limits_{X \subseteq U} m(X) = 1$.

A set $X \in \mathcal{P}(U)$ with nonzero basic probability assignment is referred to as a focal element of m. We denote by $\mathcal{M}$ the family of all focal elements of m. The pair $(\mathcal{M}, m)$ is called a crisp belief structure on U.

Lemma 1 *Let $(\mathcal{M}, m)$ be a crisp belief structure on U. Then the focal elements of m constitute a countable set.*

Proof. For any $n \in \{1, 2, \ldots\}$, let

$$H_n = \{A \in \mathcal{M} | m(A) > 1/n\}.$$

By axiom (M2) we can see that for each $n \in \{1, 2, \ldots\}$, H_n is a finite set. Since $\mathcal{M} = \bigcup\limits_{n=1}^{\infty} H_n$, we conclude that $\mathcal{M}$ is countable.

Associated with each belief structure, a pair of belief and plausibility functions can be defined.

Definition 2 *Let $(\mathcal{M}, m)$ be a crisp belief structure on U. A set function* $\mathrm{Bel} : \mathcal{P}(U) \to [0,1]$ *is called a CC-belief function on U if*

$$\mathrm{Bel}(X) = \sum_{M \subseteq X} m(M), \quad \forall X \in \mathcal{P}(U). \tag{101}$$

A set function $\mathrm{Pl} : \mathcal{P}(U) \to [0,1]$ *is called a CC-plausibility function on U if*

$$\mathrm{Pl}(X) = \sum_{M \cap X \neq \emptyset} m(M), \quad \forall X \in \mathcal{P}(U). \tag{102}$$

Remark 1 *Since $\mathcal{M}$ is a countable set, the change of convergence may not change the values of the infinite (countable) sums in Eqs.* (101) *and* (102). *Therefore, Definition* 2 *is reasonable.*

The CC-belief function and CC-plausibility function based on the same belief structure are connected by the dual property

$$\mathrm{Pl}(X) = 1 - \mathrm{Bel}(\sim X), \quad \forall X \in \mathcal{P}(U) \tag{103}$$

and moreover,

$$\mathrm{Bel}(X) \leq \mathrm{Pl}(X), \quad \forall X \in \mathcal{P}(U). \tag{104}$$

When U is finite, a CC-belief function can be equivalently defined as a monotone Choquet capacity [60] on U which satisfies the following properties [27]:
(MC1) $\mathrm{Bel}(\emptyset) = 0$,
(MC2) $\mathrm{Bel}(U) = 1$,
(MC3) for all $X_i \in \mathcal{P}(U), i = 1, 2, \ldots, k$,

$$\mathrm{Bel}(\bigcup_{i=1}^{k} X_i) \geq \sum_{\emptyset \neq J \subseteq \{1,2,\ldots,k\}} (-1)^{|J|+1} \mathrm{Bel}(\bigcap_{i \in J} X_i). \tag{105}$$

Similarly, a CC-plausibility function can be equivalently defined as an alternating Choquet capacity on U which satisfies the following properties:
(AC1) $\mathrm{Pl}(\emptyset) = 0$,
(AC2) $\mathrm{Pl}(U) = 1$,
(AC3) for all $X_i \in \mathcal{P}(U), i = 1, 2, \ldots, k$,

$$\mathrm{Pl}(\bigcap_{i=1}^{k} X_i) \leq \sum_{\emptyset \neq J \subseteq \{1,2,\ldots,k\}} (-1)^{|J|+1} \mathrm{Pl}(\bigcup_{i \in J} X_i). \tag{106}$$

A monotone Choquet capacity is a belief function in which the basic probability assignment can be calculated by using the Möbius transform:

$$m(X) = \sum_{Y \subseteq X} (-1)^{|X \setminus Y|} \mathrm{Bel}(Y), X \in \mathcal{P}(U). \tag{107}$$

A crisp belief structure can also be induced a dual pair of fuzzy belief and plausibility functions.

Definition 3 *Let $(\mathcal{M}, m)$ be a crisp belief structure on U. A fuzzy set function $\mathrm{Bel} : \mathcal{F}(U) \to [0, 1]$ is called a CF-belief function on U if*

$$\mathrm{Bel}(X) = \sum_{A \in \mathcal{M}} m(A) \mathrm{N}_A(X), \quad \forall X \in \mathcal{F}(U). \tag{108}$$

A fuzzy set function $\mathrm{Pl} : \mathcal{F}(U) \to [0, 1]$ is called a CF-plausibility function on U if

$$\mathrm{Pl}(X) = \sum_{A \in \mathcal{M}} m(A) \Pi_A(X), \quad \forall X \in \mathcal{F}(U). \tag{109}$$

Where $\mathrm{N}_A : \mathcal{F}(U) \to [0, 1]$ and $\Pi_A : \mathcal{F}(U) \to [0, 1]$ are, respectively, the necessity measure and the possibility measure determined by the crisp set A defined as follows:

$$\mathrm{N}_A(X) = \bigwedge_{u \in A} X(u), X \in \mathcal{F}(U), \tag{110}$$

$$\Pi_A(X) = \bigvee_{u \in A} X(u), X \in \mathcal{F}(U). \tag{111}$$

3.1.2 Belief and plausibility functions derived from a fuzzy belief structure

Definition 4 *Let U be a non-empty set which may be infinite. A set function $m : \mathcal{F}(U) \to [0,1]$ is referred to as a fuzzy basic probability assignment, if it satisfies axioms* (FM1) *and* (FM2)*:*

$$\text{(FM1) } m(\emptyset) = 0, \quad \text{(FM2) } \sum_{X \in \mathcal{F}(U)} m(X) = 1.$$

A fuzzy set $X \in \mathcal{F}(U)$ with $m(X) > 0$ is referred to as a focal element of m. We denote by $\mathcal{M}$ the family of all focal elements of m. The pair $(\mathcal{M}, m)$ is called a fuzzy belief structure.

Lemma 2 [61] *Let $(\mathcal{M}, m)$ be a fuzzy belief structure on W. Then the focal elements of m constitute a countable set.*

In the discussion to follow, all the focal elements are supposed to be normal, i.e., for any $A \in \mathcal{M}$, there exists an $x \in U$ such that $A(x) = 1$. Associated with the fuzzy belief structure $(\mathcal{M}, m)$, two pairs of fuzzy belief and plausibility functions can be derived.

Definition 5 *Let U be a non-empty set which may be infinite, and $(\mathcal{M}, m)$ a fuzzy belief structure on U. A crisp set function* $\mathrm{Bel} : \mathcal{P}(U) \to [0,1]$ *is referred to as a FC-belief function on U if*

$$\mathrm{Bel}(X) = \sum_{A \in \mathcal{M}} m(A) \mathrm{N}_A(X), \quad \forall X \in \mathcal{P}(U). \tag{112}$$

A crisp set function $\mathrm{Pl} : \mathcal{P}(U) \to [0,1]$ *is called a FC-plausibility function on U if*

$$\mathrm{Pl}(X) = \sum_{A \in \mathcal{M}} m(A) \Pi_A(X), \quad \forall X \in \mathcal{P}(U). \tag{113}$$

Where $\mathrm{N}_A : \mathcal{P}(U) \to [0,1]$ *and* $\Pi_A : \mathcal{P}(U) \to [0,1]$ *are, respectively, the necessity measure and the possibility measure determined by the fuzzy set A defined as follows:*

$$\mathrm{N}_A(X) = \bigwedge_{u \notin X} (1 - A(u)), X \in \mathcal{P}(U) \tag{114}$$

$$\Pi_A(X) = \bigvee_{u \in X} A(u), X \in \mathcal{P}(U). \tag{115}$$

Definition 6 *Let U be a non-empty set which may be infinite, and $(\mathcal{M}, m)$ a fuzzy belief structure on U. A fuzzy set function* $\mathrm{Bel} : \mathcal{F}(U) \to [0,1]$ *is referred to as a FF-belief function on U if*

$$\mathrm{Bel}(X) = \sum_{A \in \mathcal{M}} m(A) \mathrm{N}_A(X), \quad \forall X \in \mathcal{F}(U). \tag{116}$$

A fuzzy set function $\mathrm{Pl} : \mathcal{F}(U) \to [0,1]$ *is called a FF-plausibility function on U if*

$$\mathrm{Pl}(X) = \sum_{A \in \mathcal{M}} m(A) \Pi_A(X), \quad \forall X \in \mathcal{F}(U). \tag{117}$$

Where $\mathrm{N}_A : \mathcal{F}(U) \to [0,1]$ *and* $\Pi_A : \mathcal{F}(U) \to [0,1]$ *are, respectively, the necessity measure and the possibility measure determined by the fuzzy set A defined as follows:*

$$\mathrm{N}_A(X) = \bigwedge_{u \in U} (X(u) \vee (1 - A(u))), X \in \mathcal{F}(U), \tag{118}$$

$$\Pi_A(X) = \bigvee_{u \in U} (X(u) \wedge A(u)), X \in \mathcal{F}(U). \tag{119}$$

It can be proved that the belief and plausibility functions derived from the same fuzzy belief structure $(\mathcal{M}, m)$ are dual, that is,

$$\mathrm{Bel}(X) = 1 - \mathrm{Pl}(\sim \mathrm{X}), \quad \forall X \in \mathcal{F}(U). \tag{120}$$

And

$$\mathrm{Bel}(X) \leq \mathrm{Pl}(\mathrm{X}), \quad \forall X \in \mathcal{F}(U). \tag{121}$$

Moreover, Bel is a fuzzy monotone Choquet capacity of infinite order on U which satisfies axioms (FMC1)-(FMC3),

(FMC1) $\mathrm{Bel}(\emptyset) = 0$,
(FMC2) $\mathrm{Bel}(U) = 1$,
(FMC3) For $X_i \in \mathcal{F}(U)$, $i = 1, 2, \ldots, n$, $n \in \mathbf{N}$,

$$\mathrm{Bel}(\bigcup_{i=1}^{n} X_i) \geq \sum_{\emptyset \neq J \subseteq \{1,2,\ldots,n\}} (-1)^{|J|+1} \mathrm{Bel}(\bigcap_{j \in J} X_j). \tag{122}$$

And Pl is a fuzzy alternating Choquet capacity of infinite order on U which obeys axioms (FAC1)-(FAC3),

(FAC1) $\mathrm{Pl}(\emptyset) = 0$,
(FAC2) $\mathrm{Pl}(U) = 1$,
(FAC3) For $X_i \in \mathcal{F}(U)$, $i = 1, 2, \ldots, n$, $n \in \mathbf{N}$,

$$\mathrm{Pl}(\bigcap_{i=1}^{n} X_i) \leq \sum_{\emptyset \neq J \subseteq \{1,2,\ldots,n\}} (-1)^{|J|+1} \mathrm{Pl}(\bigcup_{i \in J} X_i). \tag{123}$$

3.2 Belief structures of rough approximations

In this subsection, we show relationships between various belief and plausibility functions in Dempster-Shafer theory of evidence and the lower and upper approximations in rough set theory with potential applications.

3.2.1 Belief functions VS crisp rough approximations

Definition 7 *Let U and W be two nonempty universes of discourse. A subset $R \in \mathcal{P}(U \times W)$ is referred to as a binary relation from U to W. The relation R is referred to as serial if for any $x \in U$ there exists $y \in W$ such that $(x, y) \in R$. If $U = W$, $R \in \mathcal{P}(U \times U)$ is called a binary relation on U, $R \in \mathcal{P}(U \times U)$ is referred to as reflexive if $(x, x) \in R$ for all $x \in U$; R is referred to as symmetric if $(x, y) \in R$ implies $(y, x) \in R$ for all $x, y \in U$; R is referred to as transitive if for any $x, y, z \in U$, $(x, y) \in R$ and $(y, z) \in R$ imply $(x, z) \in R$; R is referred to as Euclidean if for any $x, y, z \in U$, $(x, y) \in R$ and $(x, z) \in R$ imply $(y, z) \in R$; R is referred to as an equivalence relation if R is reflexive, symmetric and transitive.*

Assume that R is an arbitrary binary relation from U to W. One can define a set-valued mapping $R_s : U \to \mathcal{P}(W)$ by:

$$R_s(x) = \{y \in W | (x, y) \in R\}, \quad x \in U. \tag{124}$$

$R_s(x)$ is called the successor neighborhood of x with respect to (w.r.t.) R [62]. Obviously, any set-valued mapping F from U to W defines a binary relation from U to W by setting $R = \{(x, y) \in U \times W | y \in F(x)\}$. For $A \in \mathcal{P}(W)$, let $j(A) = R_s^{-1}(A)$ be the counter-image of A under the set-valued mapping R_s, i.e.,

$$j(A) = \begin{cases} R_s^{-1}(A) = \{u \in U | R_s(u) = A\}, & \text{if } A \in \{R_s(x) | x \in U\}, \\ \emptyset, & \text{otherwise}. \end{cases} \tag{125}$$

Then it is well-known that j satisfies the properties (J1) and (J2):

$$\text{(J1) } A \neq B \Longrightarrow j(A) \cap j(B) = \emptyset, \qquad \text{(J2) } \bigcup_{A \in \mathcal{P}(W)} j(A) = U.$$

Definition 8 *If R is an arbitrary relation from U to W, then the triple (U, W, R) is referred to as a generalized approximation space. For any set $A \subseteq W$, a pair of lower and upper approximations, $\underline{R}(A)$ and $\overline{R}(A)$, are, respectively, defined by*

$$\underline{R}(A) = \{x \in U | R_s(x) \subseteq A\}, \quad \overline{R}(A) = \{x \in U | R_s(x) \cap A \neq \emptyset\}. \tag{126}$$

The pair $(\underline{R}(A), \overline{R}(A))$ is referred to as a generalized crisp rough set, and $\underline{R}$ and $\overline{R} : \mathcal{P}(W) \to \mathcal{P}(U)$ are called the lower and upper generalized approximation operators respectively.

If U is countable set, P a normalized probability measure on U, i.e. $P(\{x\}) > 0$ for all $x \in U$, and R an arbitrary relation from U to W, then $((U, P), W, R)$ is referred to as a probability approximation space.

Theorem 1 [32] *Assume that $((U, P), W, R)$ is a serial probability approximation space, for $X \in \mathcal{P}(W)$, define*

$$m(X) = P(j(X)), \quad \mathrm{Bel}(X) = P(\underline{R}(X)), \quad \mathrm{Pl}(X) \ = P(\overline{R}(X)). \tag{127}$$

Then $m : \mathcal{P}(W) \to [0, 1]$ is a basic probability assignment on W and $\mathrm{Bel} : \mathcal{P}(W) \to [0, 1]$ *and* $\mathrm{Pl} : \mathcal{P}(W) \to [0, 1]$ *are, respectively, the CC-belief and CC-plausibility functions on W.*

Conversely, for any crisp belief structure $(\mathcal{M}, m)$ on W which may be infinite. If $\mathrm{Bel} : \mathcal{P}(W) \to [0, 1]$ *and* $\mathrm{Pl} : \mathcal{P}(W) \to [0, 1]$ *are, respectively, the CC- belief and CC-plausibility functions defined in Definition 2, then there exists a countable set U, a serial relation R from U to W, and a normalized probability measure P on U such that*

$$\mathrm{Bel}(X) = P(\underline{R}(X)), \quad \mathrm{Pl}(X) \ = P(\overline{R}(X)), \quad \forall X \in \mathcal{P}(W). \tag{128}$$

The notion of information systems (sometimes called data tables, attribute-value systems, knowledge representation systems etc.) provides a convenient tool for the representation of objects in terms of their attribute values.

An information system is a pair (U, A), where $U = \{x_1, x_2, \ldots, x_n\}$ is a non-empty, finite set of objects called the universe of discourse and $A = \{a_1, a_2, \ldots, a_m\}$ is a non-empty, finite set of attributes, such that $a : U \to V_a$ for any $a \in A$, where V_a is called the domain of a.

Each non-empty subset $B \subseteq A$ determines an indiscernibility relation as follows:

$$R_B = \{(x, y) \in U \times U | a(x) = a(y), \forall a \in B\}. \tag{129}$$

Since R_B is an equivalence relation on U, it forms a partition $U/R_B = \{[x]_B | x \in U\}$ of U, where $[x]_B$ denotes the equivalence class determined by x with respect to (w.r.t.) B, i.e., $[x]_B = \{y \in U | (x, y) \in R_B\}$.

Let (U, A) be an information system, $B \subseteq A$, for any $X \subseteq U$, denote

$$\begin{aligned} \underline{R_B}(X) &= \{x \in U | [x]_B \subseteq X\}, \\ \overline{R_B}(X) &= \{x \in U | [x]_B \cap X \neq \emptyset\}. \end{aligned} \tag{130}$$

$\underline{R_B}(X)$ and $\overline{R_B}(X)$ are respectively referred to as the lower and upper approximations of X w.r.t. (U, R_B), the knowledge generated by B. Objects in $\underline{R_B}(X)$ can be with certainty classified as elements of X on the basis of knowledge in (U, R_B), whereas objects in $\overline{R_B}(X)$ can only be classified possibly as elements of X on the basis of knowledge in (U, R_B)).

For $B \subseteq A$ and $X \subseteq U$, denote $\mathrm{Bel}_B(X) = P(\underline{R_B}(X))$ and $\mathrm{Pl}_B(X) = P(\overline{R_B}(X))$, where $P(Y) = |Y|/|U|$ and $|Y|$ is the cardinality of a set Y. Then Bel_B and Pl_B are CC-belief function and CC-plausibility function on U respectively, and the corresponding mass distribution is

$$m_{_B}(Y) = \begin{cases} P(Y), & \text{if } Y \in U/R_B, \\ 0, & \text{otherwise.} \end{cases}$$

A decision system (sometimes called decision table) is a pair $(U, C \cup \{d\})$ where (U, C) is an information system, and d is a distinguished attribute called the decision, in this case C is called the conditional attribute set, d is a map $d : U \to V_d$ of the universe U into the value set V_d, we assume, without any loss of generality, that $V_d = \{1, 2, \dots, r\}$. Define

$$R_d = \{(x, y) \in U \times U | d(x) = d(y)\}. \tag{131}$$

Then we obtain the partition $U/R_d = \{D_1, D_2, \dots, D_r\}$ of U into decision classes, where $D_j = \{x \in U | d(x) = j\}, j \leq r$. If $R_C \subseteq R_d$, then the decision system $(U, C \cup \{d\})$ is consistent, otherwise it is inconsistent. One can acquire certainty decision rules from consistent decision systems and uncertainty decision rules from inconsistent decision systems.

3.2.2 Belief functions VS rough fuzzy approximations

Definition 9 *Let (U, W, R) be a generalized approximation space, for a fuzzy set $A \in \mathcal{F}(W)$, the lower and upper approximations of A, $\underline{RF}(A)$ and $\overline{RF}(A)$, with respect to the approximation space (U, W, R) are fuzzy sets of U whose membership functions, for each $x \in U$, are defined, respectively, by*

$$\begin{aligned} \overline{RF}(A)(x) &= \bigvee_{y \in R_s(x)} A(y), \quad x \in U, \\ \underline{RF}(A)(x) &= \bigwedge_{y \in R_s(x)} A(y), \quad x \in U. \end{aligned} \tag{132}$$

The pair $(\underline{RF}(A), \overline{RF}(A))$ is referred to as a generalized rough fuzzy set, and $\underline{RF}$ and $\overline{RF} : \mathcal{F}(W) \to \mathcal{F}(U)$ are referred to as lower and upper generalized rough fuzzy approximation operators, respectively.

In the discussion to follow, we always assume that $(U, \mathcal{A}, P)$ is a probability space, i.e., U is a nonempty set, $\mathcal{A} \subseteq \mathcal{P}(U)$ a σ-algebra on U, and P a probability measure on U.

Definition 10 *A fuzzy set $A \in \mathcal{F}(U)$ is said to be measurable w.r.t. $(U, \mathcal{A})$ if $A : U \to [0, 1]$ is a measurable function w.r.t. $\mathcal{A} - \mathcal{B}([0, 1])$, where $\mathcal{B}([0, 1])$ is the family of Borel sets on $[0, 1]$. We denote by $\mathcal{F}(U, \mathcal{A})$ the family of all measurable fuzzy sets of U w.r.t. $\mathcal{A} - \mathcal{B}([0, 1])$.*

For any measurable fuzzy set $A \in \mathcal{F}(U, \mathcal{A})$, since $A_\alpha \in \mathcal{A}$ for all $\alpha \in [0, 1]$, A_α is a measurable set on the probability space $(U, \mathcal{A}, P)$ and then $P(A_\alpha) \in [0, 1]$. Notice that $f(\alpha) = P(A_\alpha)$ is monotone decreasing and left continuous, it can be seen that $f(\alpha)$ is integrable, we denote the integrand as $\int_0^1 P(A_\alpha) d\alpha$.

Definition 11 *If a fuzzy set A is measurable w.r.t. $(U, \mathcal{A})$, and P is a probability measure on $(U, \mathcal{A})$. Denote*

$$P(A) = \int_0^1 P(A_\alpha) d\alpha, \tag{133}$$

$P(A)$ is called the probability of A.

For a singleton set $\{x\}$, we will write $P(x)$ instead of $P(\{x\})$ for short.

Proposition 1 [63, 21] *The fuzzy probability measure P in Definition 11 satisfies the following properties:*
(1) $P(A) \in [0, 1]$ and $P(A) + P(\sim A) = 1$, for all $A \in \mathcal{F}(U, \mathcal{A})$.
(2) P is countably additive, i.e., for $A_i \in \mathcal{F}(U, \mathcal{A})$, $i = 1, 2, \dots$, $A_i \cap A_j = \emptyset$, $\forall i \neq j$, then

$$P(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i). \tag{134}$$

(3) $A, B \in \mathcal{F}(U, \mathcal{A})$, $A \subseteq B \Longrightarrow P(A) \leq P(B)$.
(4) If $U = \{u_i | i = 1, 2, \dots\}$ is an infinite countable set and $\mathcal{A} = \mathcal{P}(U)$, then for all $A \in \mathcal{F}(U)$,

$$P(A) = \int_0^1 P(A_\alpha) d\alpha = \sum_{x \in U} A(x) P(x). \tag{135}$$

(5) If U is a finite set with $|U| = n$, $\mathcal{A} = \mathcal{P}(U)$, and $P(u) = 1/n$, then $P(A) = \int_0^1 P(A_\alpha) d\alpha = |A|/n$ for all $A \in \mathcal{P}(U)$.

Theorem 2 *Assume that* $((U,P),W,R)$ *is a serial probability approximation space, for* $X \in \mathcal{F}(W)$*, define*

$$m(X) = P(j(X)), \quad \mathrm{Bel}(X) = P(\underline{RF}(X)), \quad \mathrm{Pl}(X) = P(\overline{RF}(X)). \tag{136}$$

Then $m : \mathcal{P}(W) \to [0,1]$ *is a basic probability assignment on* W *and* $\mathrm{Bel} : \mathcal{F}(W) \to [0,1]$ *and* $\mathrm{Pl} : \mathcal{F}(W) \to [0,1]$ *are, respectively, the CF-belief and CF-plausibility functions on* W*.*

Conversely, for any crisp belief structure $(\mathcal{M}, m)$ *on* W *which may be infinite. If* $\mathrm{Bel} : \mathcal{F}(W) \to [0,1]$ *and* $\mathrm{Pl} : \mathcal{F}(W) \to [0,1]$ *are, respectively, the CF- belief and CF-plausibility functions defined in Definition 3, then there exists a countable set* U*, a serial relation* R *from* U *to* W*, and a normalized probability measure* P *on* U *such that*

$$\mathrm{Bel}(X) = P(\underline{RF}(X)), \quad \mathrm{Pl}(X) = P(\overline{RF}(X)), \quad \forall X \in \mathcal{F}(W). \tag{137}$$

For a decision table $(U, C \cup \{d\})$, where $V_d = \{d_1, d_2, \ldots, d_r\}$, d is called a fuzzy decision if, for each $x \in U$, $d(x)$ is a fuzzy subset of V_d, i.e., $d : U \to \mathcal{F}(V_d)$, with no lose of generality, we represent d as follows:

$$d(x_i) = d_{i1}/d_1 + d_{i2}/d_2 + \cdots + d_{ir}/d_r, i = 1, 2, \ldots, n, \tag{138}$$

where $d_{ij} \in [0,1]$. In this case, $(U, C \cup \{d\})$ is called an information system with fuzzy decision. For the fuzzy decision d, we define a fuzzy indiscernibility binary relation R_d on U as follows: For $i, k = 1, 2, \ldots, n$

$$R_d(x_i, x_k) = \min\{1 - |d_{ij} - d_{kj}| \,|\, j = 1, 2, \ldots, r\}. \tag{139}$$

Then, we obtain a fuzzy similarity class $S_d(x)$ of $x \in U$ in the system $(U, C \cup \{d\})$ as follows:

$$S_d(x)(y) = R_d(x, y), \quad y \in U. \tag{140}$$

Since $S_d(x)(x) = R_d(x,x) = 1$, we see that $S_d(x) : U \to [0,1]$ is a normalized fuzzy set of U. Denote by U/R_d the fuzzy similarity classes induced by the fuzzy decision d, i.e.

$$U/R_d = \{S_d(x) | x \in U\}. \tag{141}$$

For $B \subseteq C$ and $X \in \mathcal{F}(U)$, we define the lower and upper approximations of X w.r.t. (U, R_B) as follows:

$$\begin{aligned} \underline{RF_B}(X)(x) &= \bigwedge_{y \in S_B(x)} X(y), \quad x \in U, \\ \overline{RF_B}(X)(x) &= \bigvee_{y \in S_B(x)} X(y), \quad x \in U. \end{aligned} \tag{142}$$

Theorem 3 *Let* $(U, C \cup \{d\})$ *be an information system with fuzzy decision. For* $B \subseteq C$ *and* $X \in \mathcal{F}(U)$*, if* $\underline{RF_B}(X)$ *and* $\overline{RF_B}(X)$ *are, respectively, the lower and upper approximations of* X *w.r.t.* (U, R_B) *defined by Definition 9, denote*

$$\begin{aligned} \mathrm{Bel}_B(X) &= P(\underline{\underline{RF_B}}(X)), \\ \mathrm{Pl}_B(X) &= P(\overline{\overline{RF_B}}(X)), \end{aligned} \tag{143}$$

where $P(X) = \sum_{x \in U} X(x)/|U|$ *for* $X \in \mathcal{F}(U)$*, then* $\mathrm{Bel}_B : \mathcal{F}(U) \to [0,1]$ *and* $\mathrm{Pl}_B : \mathcal{F}(U) \to [0,1]$ *are, respectively, a CF-belief function and a CF-plausibility function on* U*, and the corresponding basic probability assignment* $m_{_B}$ *is*

$$m_{_B}(Y) = \begin{cases} P(Y) = |Y|/|U|, & \text{if } Y \in U/R_B, \\ 0, & \text{otherwise.} \end{cases} \tag{144}$$

3.2.3 Belief functions VS fuzzy rough approximations

Definition 12 *Let* U *and* W *be two nonempty universes of discourse. A fuzzy subset* $R \in \mathcal{F}(U \times W)$ *is referred to as a binary relation from* U *to* W*,* $R(x,y)$ *is the degree of relation between* x *and* y*, where* $(x,y) \in U \times W$*. The fuzzy relation* R *is referred to as serial if for each* $x \in U$*,* $\bigvee_{y \in W} R(x,y) = 1$*. If* $U = W$*,* $R \in \mathcal{F}(U \times U)$ *is called a fuzzy binary relation on* U*,* R *is referred to as a reflexive fuzzy relation if* $R(x,x) = 1$ *for all* $x \in U$*;* R *is referred to as a symmetric fuzzy relation if* $R(x,y) = R(y,x)$ *for all* $x, y \in U$*;* R *is referred to as a transitive fuzzy relation if* $R(x,z) \geq \vee_{y \in U}(R(x,y) \wedge R(y,z))$ *for all* $x, z \in U$*; and* R *is referred to as an equivalence fuzzy relation if it is reflexive, symmetric, and transitive.*

Definition 13 *Let U and W be two non-empty universes of discourse and R a fuzzy relation from U to W. The triple (U, W, R) is called a generalized fuzzy approximation space. For any set $A \in \mathcal{F}(W)$, the lower and upper approximations of A, $\underline{FR}(A)$ and $\overline{FR}(A)$, with respect to the approximation space (U, W, R) are fuzzy sets of U whose membership functions, for each $x \in U$, are defined, respectively, by*

$$\begin{aligned} \overline{FR}(A)(x) &= \bigvee_{y \in W} [R(x, y) \wedge A(y)], \qquad & x \in U, \\ \underline{FR}(A)(x) &= \bigwedge_{y \in W} [(1 - R(x, y)) \vee A(y)], \qquad & x \in U. \end{aligned} \tag{145}$$

The pair $(\underline{FR}(A), \overline{FR}(A))$ is referred to as a generalized fuzzy rough set, and $\underline{FR}$ and $\overline{FR} : \mathcal{F}(W) \to \mathcal{F}(U)$ are referred to as lower and upper generalized fuzzy rough approximation operators, respectively.

Theorem 4 *Let (U, W, R) be a serial fuzzy approximation space in which U is a countable set, P a probability measure on U. If $\underline{FR}$ and $\overline{FR}$ are the fuzzy rough approximation operators defined in Definition 13, denote*

$$\mathrm{Bel}(X) = P(\underline{FR}(X)), \quad \mathrm{Pl}(X) = P(\overline{FR}(X)), \quad X \in \mathcal{F}(W). \tag{146}$$

Then $\mathrm{Bel} : \mathcal{F}(W) \to [0, 1]$ and $\mathrm{Pl} : \mathcal{F}(W) \to [0, 1]$ are, respectively, FF-fuzzy belief and FF-plausibility functions on W.

Conversely, if $(\mathcal{M}, m)$ is a fuzzy belief structure on W, $\mathrm{Bel} : \mathcal{F}(W) \to [0, 1]$ and $\mathrm{Pl} : \mathcal{F}(W) \to [0, 1]$ are the pair of FF-fuzzy belief function and FF-plausibility function defined in Definition 6, then there exists a countable set U, a serial fuzzy relation R from U to W, and a probability measure P on U such that for all $X \in \mathcal{F}(W)$,

$$\mathrm{Bel}(X) = P(\underline{FR}(X)) = \sum_{x \in U} \underline{FR}(X)(x) P(x), \tag{147}$$

$$\mathrm{Pl}(X) \ = P(\overline{FR}(X)) = \sum_{x \in U} \overline{FR}(X)(x) P(x). \tag{148}$$

A pair (U, A) is called a fuzzy information system if each $a \in A$ is a fuzzy attribute, i.e. for each $x \in U$, $a(x)$ is a fuzzy subset of V_d, that is, $a : U \to \mathcal{F}(V_a)$. Similar to Eq. (139), we can define a reflexive fuzzy binary relation R_a on U, and consequently, for any attribute subset $B \subseteq A$ one can define a reflexive fuzzy relation R_B as follows

$$R_B = \bigcap_{a \in B} R_a. \tag{149}$$

For $X \in \mathcal{F}(U)$, denote

$$\mathrm{Bel}_B(X) = P(\underline{FR}_B(X)), \quad \mathrm{Pl}_B(X) \ = P(\overline{FR}_B(X)), \tag{150}$$

where $P(X) = \sum_{x \in U} X(x)/|U|$ for $X \in \mathcal{F}(U)$. Then, according to Theorem 4, $\mathrm{Bel}_B : \mathcal{F}(U) \to [0, 1]$ and $\mathrm{Pl}_B : \mathcal{F}(U) \to [0, 1]$ are respectively, FF-fuzzy belief function and FF-plausibility function on U. More specifically, if X in Eq. (150) is crisp subset of U, then $\mathrm{Bel}_B : \mathcal{P}(U) \to [0, 1]$ and $\mathrm{Pl}_B : \mathcal{P}(U) \to [0, 1]$ defined by Eq. (150) are respectively, FC-fuzzy belief functions and FC-plausibility functions on U. Based on these observations, we believe that FF-fuzzy belief functions and FF-plausibility functions can be used to analyze uncertainty fuzzy information systems with fuzzy decision and whereas FC-fuzzy belief functions and FC-plausibility functions can be employed to deal with knowledge discovery in fuzzy information systems with crisp decision.

3.3 Conclusion of this section

The lower and upper approximations of a set capture the non-numeric aspect of uncertainty of the set which can be interpreted as the qualitative representation of the set, whereas the pair of the belief and plausibility measures of the set characterize the numeric aspect of uncertainty of the set which can be treated as the quantitative characterization of the set. In this section, we have introduced some generalized belief and plausibility and belief functions on the Dempster-Shafer theory of evidence. We have shown that the fuzzy belief and plausibility functions can be interpreted as the lower and upper approximations in rough set theory. That is, the belief and plausibility functions in the Dempster-Shafer theory of evidence can be represented as the probabilities of lower and upper approximations in rough set theory, thus rough set theory may be regarded as the basis of the Dempster-Shafer theory of evidence. And the Dempster-Shafer theory of evidence in the fuzzy environment provides a potentially useful tool for reasoning and knowledge acquisition in fuzzy systems and fuzzy decision systems.

4 Applications of Fuzzy Rough Sets

Both fuzzy set and rough set theories have fostered broad research communities and have been applied in a wide range of settings. More recently, this has extended also to the hybrid fuzzy rough set models. This section tries to give a sample of those applications, which are in particular numerous for machine learning but which also cover many other fields, like image processing, decision making and information retrieval.

Note that we do not consider applications that simply involve a joint application of fuzzy sets and rough sets, like for instance a rough classifier that induces fuzzy rules. Rather, we focus on applications that specifically involve one of the fuzzy rough set models discussed in the previous sections.

4.1 Applications in machine learning

4.1.1 Feature selection

The most prominent application of classical rough set theory is undoubtedly semantics-preserving data dimensionality reduction: the removal of attributes (features) from information systems[1] without sacrificing the ability to discern between different objects. A minimal attribute subset $B \subseteq A$ that maintains objects' discernibility is called a *reduct*. For classification tasks, it is sufficient to be able to discern between objects belonging to different classes, in which case a *decision reduct*, also called *relative reduct*, is sought.

The traditional rough set model sets forth a crisp notion of discernibility, where two objects are either discernible or not w.r.t. a set of attributes B based on their values for all attributes in B. To be able to handle numerical data, discretization is required. Fuzzy-rough feature selection avoids this external preprocessing step by incorporating graded indiscernibility between objects directly into the data reduction process. On the other hand, by the use of fuzzy partitions, such that objects can belong to different classes to varying degrees, a more flexible data representation is obtained.

Chronologically, the oldest proposal to apply fuzzy rough sets to feature selection is due to Kuncheva [39] in 1992. However, rather than using Dubois and Prade's definition, she proposed her own notion of a fuzzy rough set based on an inclusion measure. Based on this, she defined a quality measure for evaluating attribute subsets w.r.t. their ability to approximate a predetermined fuzzy partition on the data, and illustrated its usefulness on a medical data set.

Jensen and Shen [40, 42] were the first to propose a reduction method that generalizes the classical rough set positive region and dependency function. In particular, the dependency degree γ_B, with $B \subseteq A$, is used to guide a hill-climbing search in which, starting from $B = \emptyset$, in each step an attribute a is added such that $\gamma_{B\cup\{a\}}$ is maximal. The search ends when there is no further increase in the measure. This is the QuickReduct algorithm. In [64] they replaced this simple greedy search heuristic by a more complex one based on ant colony optimization.

Hu et al. [66] formally defined the notions of reduct and decision reduct in the fuzzy-rough case, referring to the invariance of the fuzzy partition induced by the data, and of the fuzzy positive region, respectively. They also showed that minimal subsets that are invariant w.r.t. (conditional) entropy are (decision) reducts.

Tsang et al. [67] proposed a method based on the discernibility matrix and function to find all decision reducts where invariance of the fuzzy positive region defined using Dubois and Prade's definition is imposed, and proved its correctness. In [68], an extension of this method is defined that finds all decision reducts where the approximations are defined using a lower semi-continuous t-norm $\mathcal{T}$ and its R-implicator. The particular

[1]An information system (U, A) consists of a non-empty set U of objects which are described by a set of attributes A.

case using Łukasiewicz connectives was studied in [69]. Later, Zhao and Tsang [?] studied relationships that exist between different kinds of decision reducts, defined using different types of fuzzy connectives.

In [65], Jensen and Shen introduced three different quality measures for evaluating attribute subsets: the first one is a revised version of their previously defined degree of dependency, the second one is based on the fuzzy boundary region and the third one on the satisfaction of the clauses of the fuzzy discernibility function. On the other hand, in [70], Cornelis et al. proposed the definition of fuzzy $\mathcal{M}$-decision reducts, where $\mathcal{M}$ is an increasing, [0,1]-valued quality measure. They studied two measures based on the fuzzy positive region and two more based on the fuzzy discernibility function, and applied them to classification and regression problems.

In [46], Chen and Zhao study the concept of local reduction: instead of looking for a global reduction, where the whole positive region is considered as an invariant, they focus on subsets of decision classes and identify the conditional attributes that provide minimal descriptions for them.

Over the past few years, there has also been considerable interest in the application of noise-tolerant fuzzy rough set models to feature selection, where the aim is to make the reduction more robust in the presence of noisy or erroneous data. For instance, Hu et al.[71] defined fuzzy rough sets as an extension of Variable Precision Rough Sets, and used a corresponding notion of positive region to guide a greedy search algorithm. In [72], Cornelis and Jensen evaluated the Vaguely Quantified Rough Set (VQRS) approach to feature selection. They found that because the model does not satisfy monotonicity w.r.t. the fuzzy relation R, adding more attributes does not always lead to an expansion of the fuzzy positive region, and the hill-climbing search sometimes runs into troubles. Furthermore, in [73] Hu et al., inspired by the idea of soft margin support vector machines, introduced soft fuzzy rough sets and applied them to feature selection.

He et al. [74] consider the problem of fuzzy-rough feature selection for decision systems with fuzzy decisions, that is, where the decision attribute is characterized by a fuzzy T-similarity relation instead of a crisp one. This is the case of regression problems. They give an algorithm for finding all decision reducts and another one for finding a single reduction.

The relatively high complexity of fuzzy-rough feature selection algorithms somewhat limits is applicability to large datasets. In view of this, Chen et al. [75] propose a fast algorithm to obtain one reduct, based on a procedure to find the minimal elements of the discernibility matrix of [67]. The algorithm is compared w.r.t. execution time with the proposals in [65] and [67], and turns out to be a lot faster. On the other hand, Qian et al. [76] implement an efficient version of feature selection using the model of Hu et al. [71].

The use of kernel functions as fuzzy similarity relations in feature selection algorithms has also sparked researchers' interest. In particular, Du et al. [77] apply fuzzy-rough feature selection with kernelized fuzzy rough sets to yawn detection, while Chen et al. [78] propose parameterized attribute reduction with Gaussian kernel based fuzzy rough sets. He and Wu [79] develop a new method to compute membership for fuzzy support vector machines (FSVMs) by using a Gaussian kernel-based fuzzy rough set, and employ a technique of attribute reduction using Gaussian kernel-based fuzzy rough sets to perform feature selection for FSVMs.

Finally, Derrac et al. [80] combine fuzzy-rough feature selection with evolutionary instance selection.

4.1.2 Instance selection

Instance selection can be seen as the orthogonal task to feature selection: here the goal is to reduce an information system (U, A) by removing objects from U. The first work on instance selection using fuzzy rough set theory was presented in [81]. The main idea is that instances for which the fuzzy rough lower approximation membership is lower than a certain threshold are removed. This idea was improved in [82], where the selection threshold is optimized. This method has been applied in combination with evolutionary feature selection in [83] and for imbalanced classification problems in [84, 85], in combination with resampling methods.

4.1.3 Classification

Fuzzy rough sets have been widely used for classification purposes, either by means of rule induction or by plugging them into existing classifiers like nearest neighbor classifiers, decision trees and support vector machines (SVM).

The earliest work on rule induction using fuzzy rough set theory can be found in [25]. In this paper, the authors propose a fuzzy rough framework to induce fuzzy decision rules that does not use any fuzzy logical connectives. Later, in [43], an approach that generates rules from data using fuzzy reducts was presented, with a fuzzy rough feature selection preprocessing step. In [86], the authors noticed that using feature selection as a preprocessing step often leads to too specific rules, and proposed an algorithm for simultaneous feature selection and rule induction. In [87, 88], a rule-based classifier is built using the so-called consistency degree as a critical value to keep the discernibility information invariant in the rule induction process. Another approach to fuzzy rough rule induction can be found in [89], where rules are found from training data with hierarchical and quantitative attribute values. The most recent work can be found in [90], where fuzzy equivalence relations are used to model different types of attributes in order to obtain small rule sets from hybrid data, and in [91] where a harmony search algorithm is proposed to generate emerging rule sets.

In [92], the K nearest neighbor method was improved using fuzzy set theory. So far, three different fuzzy rough based approaches have been used to improve this Fuzzy Nearest Neighbor (FNN) classifier. In [93], the author introduces a fuzzy rough ownership function and plugs it into the FNN algorithm. In [94, 95, 96, 97], the extent to which the nearest neighbors belong to the fuzzy lower and upper approximations of a certain class are used to predict the class of the target instance, these techniques are applied in [98] for mammographic risk analysis. Finally, in [99], the FNN algorithm is improved using the fuzzy rough positive regions as weights for the nearest neighbors.

During the last decade, several authors have worked on fuzzy rough improvements of decision trees. The common idea of these methods is that during the construction phase of the decision tree, the feature significances are measured using fuzzy rough techniques [100, 101, 102, 103]. In [104, 105, 106], the kernel functions of the SVM are redefined using fuzzy rough sets, to take into account the inconsistency between conditional attributes and the decision class. In [79], this approach is combined with fuzzy rough feature selection. In [107], SVMs are reformulated by plugging in the fuzzy rough memberships of all training samples into the constraints of the SVMs.

4.1.4 Clustering

Many authors have worked on clustering methods that use both fuzzy set theory and rough set theory, but to the best of our knowledge, only two approaches use fuzzy rough sets for clustering. In [108], fuzzy rough sets are used to measure the intra-cluster similarity, in order to estimate the optimal number of clusters. In [109], a fuzzy rough measure is used to measure the similarity between genes in microarray analysis, in order to generate clusters such that genes within a cluster are highly correlated to the sample categories, while those in different clusters are as dissimilar as possible.

4.1.5 Neural networks

There are many approaches to incorporate fuzzy rough set theory in neural networks. One option is to use fuzzy rough set theory to reduce the problem that samples in the same input clusters can have different classes. The resulting fuzzy rough neural networks are designed such that they work as fuzzy rough membership functions [110, 111, 112, 113]. A related approach is to use fuzzy rough set theory to find the importance of each subset of information sources of subnetworks [114]. Other approaches use fuzzy rough set theory to measure the importance of each feature in the input layer of the neural network [115, 116, 117].

4.2 Other applications

4.2.1 Image processing

Fuzzy rough sets have been used in several domains of image processing. They are especially suitable for these tasks because they can capture both indiscernibility and vagueness, which are two important aspects of image processing.

In [118, 119], fuzzy rough based image segmenting methods are proposed and applied in a traditional Chinese medicine tongue image segmentation experiment. Often, fuzzy rough attribute reduction methods are proposed for image processing problems, as in [120] or in [121], where the methods are applied for face recognition. In [122], a method for edge detection is proposed by building a hierarchy of rough-fuzzy sets to exploit the uncertainty and vagueness at different image resolutions. Another aspect of image processing is texture segmentation, this problem is tackled in [123] using rough-fuzzy sets. In [124], the authors solve the image classification problem using a nearest neighbor clustering algorithm based on fuzzy rough set theory, and apply their algorithm to hand gesture recognition. In [125], a combined approach of neural network classification systems with a fuzzy rough sets based feature reduction method is presented. In [126], fuzzy rough feature reduction techniques are applied to a large-scale Mars McMurdo panorama image.

4.2.2 Decision making

Fuzzy rough set theory has many applications in decision making. In [127], the authors calculate the fuzzy rough memberships of software components in previous projects and decide based on these values which ones to reuse in a new program. In [128, 129], a multi objective decision making model based on fuzzy rough set theory is used to solve the inventory problem. In [130], variable precision fuzzy rough sets are used to develop a decision making model, that is applied for IT offshore outsourcing risk evaluation. Another approach can be found in [131] where the decision corresponds to the decision corresponding with the instance with maximal sum of lower and upper soft fuzzy rough approximation. Recent work can be found in [132], where a fuzzy rough set model over two universes is defined to develop a general decision making framework in an uncertainty environment for solving a medical diagnosis problem.

4.2.3 Information retrieval, data Mining and the web

Fuzzy rough sets have been used to model imprecision and vagueness in databases. In [133], the authors develop a fuzzy rough relational database, while in [134], a fuzzy rough extension of a rough object classifier for relational database mining is studied. In [135], fuzzy rough set theory is used to mine from incomplete datasets, while in [136], fuzzy rough sets are incorporated in mining agents for predicting stock prices. More recently, fuzzy rough sets have been applied to identify imprecision in temporal database models [137, 138].

In [139, 140], fuzzy rough set theory is used to approximate document queries. In the context of the semantic web, a lot of work has been done on fuzzy rough description logics. The first paper on this topic can be found in [141], where a fuzzy rough ontology was proposed. Later, in [142], the authors propose a fuzzy rough extension of the descriptive logic $\mathcal{SHIN}$. A fuzzy rough extension of the descriptive logic $\mathcal{ALC}$ can be found in [143]. In [144, 145], an improved and more general approach is presented.

References

[1] Z. Pawlak, Rough sets, International Journal of Computer and Information Science 11 (1982) 341–356.

[2] Z. Pawlak, Rough Sets: Theoretical Aspects of Reasoning about Data, Kluwer Academic Publishers, Boston, 1991.

[3] A. Nakamura, Fuzzy rough sets, Notes on Multiple-Valued Logic in Japan 9 (1988) 1–8.

[4] D. Dubois, H. Prade, Rough fuzzy sets and fuzzy rough sets, International Journal of General Systems 17 (1990) 191–209.

[5] D. Dubois, H. Prade, Putting rough sets and fuzzy sets together, in: R. Słowiński (ed.) Intelligent Decision Support, Kluwer, Dordrecht, 1992, pp.203–232.

[6] N.N. Morsi, M.M. Yakout, Axiomatics for fuzzy rough sets, Fuzzy Sets and Systems 100 (1998) 327–342.

[7] S. Greco, B. Matarazzo, R. Słowiński, The use of rough sets and fuzzy sets in MCDM, in: T. Gal et al.(eds.) Multicriteria Decision Making, Kluwer, Boston, MA, 1999, pp.14-1–14-59.

[8] D. Boixader, J. Jacas, J. Recasens, Upper and lower approximations of fuzzy sets, International Journal of General Systems 29 (2000) 555–568.

[9] A.M. Radzikowska, E.E. Kerre, A comparative study of fuzzy rough set, Fuzzy Sets and Systems 126 (2002) 137–155.

[10] M. Inuiguchi, T. Tanino, New fuzzy rough sets based on certainty qualification, in: K. Pal et al. (eds.) Rough-Neural Computing, Springer, Berlin-Heidelberg, 2003, pp.278–296.

[11] W.Z. Wu, J.S. Mi, W.X. Zhang, Generalized fuzzy rough sets, Information Sciences 151 (2003) 263–282.

[12] M. Inuiguchi, Generalization of rough sets: From crisp to fuzzy cases, in: S. Tsumoto, et al. (eds.) Rough Sets and Current Trends in Computing: 4th International Conference, Proceedings, LNAI 3066, Springer-Verlag (2004) pp.26–37.

[13] A.M. Radzikowska, E.E. Kerre, Fuzzy rough sets based on residuated lattices, Transactions on Rough Sets II, LNCS 3135, 2004, pp.278–296.

[14] W.Z. Wu, W.X. Zhang, Constructive and axiomatic approaches of fuzzy approximation operators, Information Sciences 159 (2004) 233–254.

[15] J.S. Mi, W.X. Zhang, An axiomatic characterization of a fuzzy generalization of rough sets, Information Sciences 160 (2004) 235–249.

[16] W.Z. Wu, Y. Leung, J.S. Mi, On characterizations of (I, T)-fuzzy rough approximation operators, Fuzzy Sets and Systems 15 (2005) 76–102.

[17] D.S. Yeung, D.G. Chen, E.C.C. Tsang, J.W.T. Lee, X.Z. Wang, On the generalization of fuzzy rough sets, IEEE Transactions on Fuzzy Systems 13 (2005) 343–361.

[18] M. De Cock, C. Cornelis, E.E. Kerre, Fuzzy rough sets: The forgotten step, IEEE Transactions on Fuzzy Systems 15 (2007) 121–130.

[19] T.J. Li, W.X. Zhang, Rough fuzzy approximations on two universes of discourse, Information Sciences 178 (2008) 892–906.

[20] J.S. Mi, Y. Leung, H.Y. Zhao, T. Feng, Generalized fuzzy rough sets determined by a triangular norm, Information Sciences 178 (2008) 3203–3213.

[21] W.Z. Wu, Y. Leung, J.S. Mi, On generalized fuzzy belief functions in infinite spaces, IEEE Transactions on Fuzzy Systems 17 (2009) 385–397.

[22] X.D. Liu, W. Pedrycz, T.Y. Chai, M.L. Song, The development of fuzzy rough sets with the use of structures and algebras of axiomatic fuzzy sets, IEEE Transactions on Knowledge and Data Engineering 21 (2009) 443–462.

[23] W.Z. Wu, On some mathematical structures of T-fuzzy rough set algebras in infinite universes of discourse, Fundamenta Informaticae 108 (2011) 337–369.

[24] S. Greco, M. Inuiguchi, R. Słowiński, Rough sets and gradual decision rules, in: G. Wang et al. (eds.) Rough Sets, Fuzzy Sets, Data Mining, and Granular Computing, LNAI 2639, Springer, Berlin-Heidelberg, 2003, pp.156–164.

[25] S. Greco, M. Inuiguchi, R. Słowiński, Fuzzy rough sets and multiple-premise gradual decision rules, Int. J. Approximate Reasoning 41(2) (2006) 179–211.

[26] M. Inuiguchi, Classification- versus approximation-oriented fuzzy rough sets, in: B. Bouchon-Meunier et al. (eds.), Proceedings of Information Processing and Management of Uncertainty in Knowledge-Based Systems, 2004, CD-ROM.

[27] G. Shafer, A Mathematical Theory of Evidence, Princeton University Press, Princeton, 1976.

[28] A. Skowron, The relationship between rough set theory and evidence theory, Bulletin of the Polish Academy of Sciences: Mathematics 37 (1989) 87–90.

[29] A. Skowron, The rough sets theory and evidence theory, Fundamenta Informaticae 13 (1990) 245–262.

[30] A. Skowron, J. Grzymala-Busse, From rough set theory to evidence theory, in: R.R. Yager, M. Fedrizzi, J. Kacprzyk (Eds.), Advance in the Dempster-Shafer Theory of Evidence, Wiley, New York, 1994, pp. 193–236.

[31] W.Z. Wu, Y. Leung, W.X. Zhang, Connections between rough set theory and Dempster-Shafer theory of evidence, International Journal of General Systems 31 (2002) 405–430.

[32] W.Z. Wu, J.S. Mi, Some mathematical structures of generalized rough sets in infinite universes of discourse, Transactions on Rough Sets XIII, LNCS 6499 (2011) 175–206.

[33] Y.Y. Yao, P.J. Lingras, Interpretations of belief functions in the theory of rough sets, Information Sciences 104 (1998) 81–106.

[34] P.J. Lingras, Y.Y. Yao, Data mining using extensions of the rough set model, Journal of the American Society for Information Science 49 (1998) 415–422.

[35] W.Z. Wu, Attribute reduction based on evidence theory in incomplete decision systems, Information Sciences 178 (2008) 1355–1371.

[36] W.Z. Wu, Knowledge reduction in random incomplete decision tables via evidence theory, Fundamenta Informaticae 115 (2012) 203–218.

[37] W.Z. Wu, M. Zhang, H.Z. Li, J.S. Mi, Knowledge reduction in random information systems via Dempster-Shafer theory of evidence, Information Sciences 174 (2005) 143–164.

[38] M. Zhang, L.D. Xu, W.X. Zhang, H.Z. Li, A rough set approach to knowledge reduction based on inclusion degree and evidence reasoning theory, Expert Systems 20 (2003) 298–304.

[39] L.I. Kuncheva, Fuzzy rough sets: Application to feature selection, Fuzzy Sets and Systems 51 (1992) 147–153.

[40] R. Jensen, Q. Shen, Fuzzy-rough attributes reduction with application to web categorization, Fuzzy Sets and Systems 141 (2004) 469–485.

[41] R. Jensen, Q. Shen, Semantics-preserving dimensionality reduction: Rough and fuzzy-rough based approaches, IEEE Transactions on Knowledge and Data Engineering 16 (2004) 1457–1471.

[42] R. Jensen, Q. Shen, Fuzzy-rough sets assisted attribute selection, IEEE Transactions on Fuzzy Systems 15 (2007) 73–89.

[43] X.Z. Wang, E.C.C. Tsang, S.Y. Zhao, D.G. Chen, D.S. Yeung, Learning fuzzy rules from fuzzy samples based on rough set technique, Fuzzy Sets and Systems 177 (2007) 4493–4514.

[44] S.Y. Zhao, E.C.C. Tsang, On fuzzy approximation operators in attribute reduction with fuzzy rough sets, Information Sciences 178 (2008) 3163–3176.

[45] S.Y. Zhao, E.C.C. Tsang, D.G. Chen, The model of fuzzy variable precision rough sets, IEEE Transactions on Fuzzy Systems 17 (2009) 451–467.

[46] D.G. Chen, S.Y. Zhao, Local reduction of decision system with fuzzy rough sets, Fuzzy Sets and Systems 161 (2010) 1871–1883.

[47] Q.H. Hu, L. Zhang, D.G. Chen, W. Pedrycz, D.R. Yu, Gaussian kernel based fuzzy rough sets: Model, uncertainty measures and applications, International Journal of Approximate Reasoning 51 (2010) 453–471.

[48] Q.H. Hu, D.R. Yu,W. Pedrycz, D.G. Chen, Kernelized fuzzy rough sets and their applications, IEEE Transactions on Knowledge and Data Engineering 23 (2011) 1649–1667.

[49] Q.H. Hu, L. Zhang, S. An, D. Zhang, D.R. Yu, On robust fuzzy rough set models, IEEE Transactions on Fuzzy Systems 20 (2012) 636–651.

[50] M. Inuiguchi, Generalization of rough sets and rule extraction, Transaction on Rough Sets, I, LNCS 3100 (2004) 96–119.

[51] E.P. Klement, R. Mesiar, E. Pap, E., Triangular Norms, Kluwer, Dordrecht (2000).

[52] M. Inuiguchi, M. Sakawa, On the closure of generation processes of implication functions from a conjunction function, in: T. Yamakawa, G. Matsumoto (eds.) Methodologies for the Conception, Design, and Application of Intelligent Systems, Vol.1, World Scientific, Singapore (1996) pp.327–330.

[53] D. Dubois, H. Prade, Gradual inference rules in approximate reasoning, Information Sciences 61 (1992) 103–122.

[54] L.A. Zadeh, A fuzzy set-theoretic interpretation of linguistic hedge, J. Cybernetics 2 (1974) 4–34.

[55] J.F. Baldwin, A new approach to approximate reasoning using a fuzzy logic, Fuzzy Sets and Systems 2(4) (1979) 309–325.

[56] Y. Tsukamoto, An approach to fuzzy reasoning method, in: M.M. Gupta, et al. (eds.) Advances in Fuzzy Set Theory and Applications, North-Holland, New-York, 1979, pp.137–149.

[57] D. Dubois, H. Prade, Fuzzy sets in approximate reasoning, part 1: Inference with possibility distributions, Fuzzy Sets and Systems 40 (1991) 143–202.

[58] M. Inuiguchi, T. Tanino, A new class of necessity measures and fuzzy rough sets based on certainty qualifications, in: W. Ziarko, Y. Yao (eds.) Rough Sets and Current Trends in Computing (RSCTC 2000), LNCS 2005 (2001) pp.261–268.

[59] M. Inuiguchi, T. Tanino, Function approximation by fuzzy rough sets, B. Bouchon-Meunier et al. (eds.) Intelligent Systems for Information Processing: From Representation to Applications, Elsevier, Amsterdam, The Netherlands (2003) pp.93–104.

[60] G. Choquet, Theory of capacities, Annales de l'institut Fourier 5 (1954) 131–295.

[61] L. Biacino, Fuzzy subsethood and belief functions of fuzzy events, Fuzzy Sets and Systems 158 (2007) 38–49.

[62] Y.Y. Yao, Generalized rough set model, in: L. Polkowski, A. Skowron(Eds.), Rough Sets in Knowledge Discovery 1. Methodology and Applications, Physica-Verlag, Heidelberg, 1998, pp.286–318.

[63] D.G. Chen, W.X. Yang, F.C. Li, Measures of general fuzzy rough sets on a probabilistic space, Information Sciences 178 (2006) 3177–3187.

[64] R. Jensen, Q. Shen, Fuzzy-rough data reduction with ant colony optimization, Fuzzy Sets and Systems 149(1) (2005) 5–20.

[65] R. Jensen, Q. Shen, New approaches to fuzzy-rough feature selectio, IEEE Transactions on Fuzzy Systems 17(4) (2009) 824–838.

[66] Q.H. Hu, D.R. Yu, Z.X. Xie, Information-preserving hybrid data reduction based on fuzzy-rough techniques, Pattern Recognition Letters, 27(5) (2006) 414–423.

[67] E.C.C. Tsang, D.G. Chen, D.S. Yeungm, X.Z. Wang, J.W.T. Lee, Attributes reduction using fuzzy rough sets, IEEE Transactions on Fuzzy Systems, 16(5) (2008) 1130–1141.

[68] D.G. Chen, E.C.C. Tsang, S.Y. Zhao, Attribute reduction based on fuzzy rough sets, in: M. Kryszkiewicz, et al. (eds.) Proc. Int. Conf. on Rough Sets and Intelligent Systems Paradigms, LNCS 4585, Springer, Heidelberg, pp.73–89, 2007.

[69] D.G. Chen, E.C.C. Tsang, S.Y. Zhao, An approach of attributes reduction based on fuzzy tl-rough sets, in: Proceedings of IEEE International Conference on Systems, Man, and Cybernetics, pp.486–491, 2007.

[70] C. Cornelis, G.H. Martin, R. Jensen, D. Slezak, Feature selection with fuzzy decision reducts, Information Sciences 180(2) (2010) 209–224.

[71] Q.H. Hu, Z.X. Xie, D.R. Yu, Hybrid attribute reduction based on a novel fuzzy-rough model and information granulation, Pattern Recognition 40(12) (2007) 3509–3521.

[72] C. Cornelis, R. Jensen, A noise-tolerant approach to fuzzy-rough feature selection. in: Proceedings of IEEE International Conference on Fuzzy Systems (FUZZ- IEEE 2008), pp.1598–1605, 2008.

[73] Q.H. Hu, A. An, D.R. Yu, Soft fuzzy rough sets for robust feature evaluation and selection, Information Sciences 180(22) (2010) 4384–4440.

[74] Q. He, C.X. Wu, D.G. Chen, S.Y. Zhao, Fuzzy rough set based attribute reduction for information systems with fuzzy decisions, Knowledge-Based Systems 24(5) (2011) 689–696.

[75] D.G. Chen, L. Zhang, S.Y. Zhao, Q.H. Hu, P.F. Zhu, A novel algorithm for finding reducts with fuzzy rough sets, IEEE Transactions on Fuzzy Systems 20(2) (2012) 385–389.

[76] Y.H. Qian, C. Li, J.Y. Liang, An efficient fuzzy-rough attribute reduction approach, in: J.T. Yao et al. (eds.), Rough Sets and Knowledge Technology, LNAI 6954, pp.63–70, 2011.

[77] Y. Du, Q.H. Hu, D.G. Chen, P.J. Ma, Kernelized fuzzy rough sets based yawn detection for driver fatigue monitoring, Fundamenta Informaticae 111(1) (2011) 65–79.

[78] D.G. Chen, Q.H. Hu, Y.P. Yang, Parameterized attribute reduction with gaussian kernel based fuzzy rough sets, Information Sciences 181(23) (2011) 5169–5179.

[79] Q. He, C.X. Wu, Membership evaluation and feature selection for fuzzy support vector machine based on fuzzy rough sets, Soft Computing 15(6) (2011) 1105–114.

[80] J. Derrac, C. Cornelis, S. Garcia, F. Herrera, Enhancing evolutionary instance selection algorithms by means of fuzzy rough set based feature selection, Information Sciences 186(1) (2012) 73–92.

[81] R. Jensen, C. Cornelis, Fuzzy-rough instance selection, in: Proceedings of IEEE International Conference on Fuzzy Systems, pp.1–7, 2010.

[82] N. Verbiest, C. Cornelis, F. Herrera, Granularity-based instance selection, in: Proceedings of 20th Annual Belgian-Dutch Conference on Machine Learning, pp.101–103, 2011.

[83] J. Derrac, N. Verbiest, S., Garcia, C. Cornelis, F. Herrera, On the use of evolutionary feature selection for improving fuzzy rough set based prototype selection, Soft Computing (2012) 1–16.

[84] E. Ramentol, N. Verbiest, R. Bello, Y. Caballero, C. Cornelis, F. Herrera, Smote-frst: A new resampling method using fuzzy rough set theory, in: Proceedings of The 10th International FLINS Conference on uncertainty Modeling in Knowledge Engineering and Decision Making, 800–805, 2012.

[85] N. Verbiest, E. Ramentol, C. Cornelis, F. Herrera, Improving smote with fuzzy rough prototype selection to detect noise in imbalanced classification data, in: Proceedings of the 13th Ibero-American Conference on Artificial Intelligence (IBERAMIA), pp.169–178, 2012.

[86] R. Jensen, C. Cornelis, Q. Shen, Hybrid fuzzy-rough rule induction and feature selection, in: Proceedings of IEEE International Conference on Fuzzy Systems (FUZZ-IEEE), pp.1151–1156, 2009.

[87] E.C.C. Tsang, S.Y. Zhao, J. Lee, Rule induction based on fuzzy rough sets, in: Proceedings of International Conference on Machine Learning and Cybernetics, Vol. 5, pp.3028–3033, 2007.

[88] S.Y. Zhao, E.C.C. Tsang, D.G. Chen, X.Z. Wang, Building a rule-based classifier — a fuzzy-rough set approach, IEEE Transactions on Knowledge and Data Engineering 22 (2010) 624–638.

[89] T.P. Hong, Y.L. Liou, S.L. Wang, Fuzzy rough sets with hierarchical quantitative attributes, Expert Systems and Applications, 36(3) (2009) 6790–6799.

[90] Y. Liu, Q. Zhou, E. Rakus-Andersson, G. Bai, A fuzzy-rough sets based compact rule induction method for classifying hybrid data, In: T. Li, et al. (eds.), Rough Sets and Knowledge Technology (RSKT), LNCS 7414, pp.63–70, 2012.

[91] R. Diao, Q. Shen, A harmony search based approach to hybrid fuzzy-rough rule induction, in: Proceedings of the 21st International Conference on Fuzzy Systems, 1549–1556, 2012.

[92] J.M. Keller, M.R. Gray, J.R. Givens, A fuzzy k-nearest neighbor algorithm, IEEE Transactions on Systems, Man, and Cybernetics 15 (1985) 580–585.

[93] M. Sarkar, Fuzzy-rough nearest neighbor algorithms in classification, Fuzzy Sets and Systems 158 (2007) 2134–2152.

[94] R. Jensen, C. Cornelis: A new approach to fuzzy-rough nearest neighbour classification, in: C.-C. Chan et al. (eds.) Rough Sets and Current Trends in Computing, LNCS 5306, pp.310–319, 2008.

[95] R. Jensen, C. Cornelis, Fuzzy-rough nearest neighbour classification and prediction, Theoretical Computer Science 412 (2011) 5871–5884.

[96] Y. Qu, C. Shang, Q. Shen, N.M. Parthalain, W. Wu, Kernel-based fuzzy-rough nearest neighbour classification, in: IEEE International Conference onFuzzy Systems (FUZZ-IEEE), pp.1523 –1529, 2011.

[97] H. Bian, L. Mazlack, Fuzzy-rough nearest-neighbor classification approach, in: 22nd International Conference of the North American Fuzzy Information Processing Society (NAFIPS), pp.500–505, 2003.

[98] M.N. Parthalain, R. Jensen, Q. Shen, R. Zwiggelaar, Fuzzy-rough approaches for mammographic risk analysis, Intelligent Data Analysis 13 (2010) 225–244.

[99] N. Verbiest, C. Cornelis, R. Jensen, Fuzzy rough positive region-based nearest neighbour classification, in: Proceedings of the 20th International Conference on Fuzzy Systems (FUZZ-IEEE2012), pp.1961–1967, 2012.

[100] R. Jensen, Q. Shen, Fuzzy-rough feature significance for decision trees, in: Proceedings of the 2005 UK Workshop on Computational Intelligence, 89–96, 2005.

[101] R. Bhatt, M. Gopal, Frct: fuzzy-rough classification trees, Pattern Analysis and Applications 11 (2008) 73–88.

[102] M. Elashiri, H. Hefny, A.A. Elwahab, Induction of fuzzy decision trees based on fuzzy rough set techniques, in: Proceedings of the International Conference on Computer Engineering Systems (ICCES), pp.134–139, 2011.

[103] J. Zhai, Fuzzy decision tree based on fuzzy-rough technique, Soft Computing 15 (2011) 1087–1096.

[104] D.G. Chen, Q. He, X.Z. Wang, Frsvms: Fuzzy rough set based support vector machines, Fuzzy Sets and Systems 161 (2010) 596–607.

[105] Z. Zhang, D.G. Chen, Q. He, H. Wang, Least squares support vector machines based on fuzzy rough set, in: IEEE International Conference on Systems Man and Cybernetics (SMC), pp.3834–3838, 2010.

[106] Z. Xue, W. Liu, A fuzzy rough support vector regression machine, in: 9th International Conference on Fuzzy Systems and Knowledge Discovery (FSKD), pp.840–844, 2012.

[107] D.G. Chen, S. Kwong, Q. He, H. Wang, Geometrical interpretation and applications of membership functions with fuzzy rough sets, Fuzzy Sets and Systems 193 (2012) 122–135.

[108] F. Li, F. Min, Q. Liu, Intra-cluster similarity index based on fuzzy rough sets for fuzzy c-means algorithm, in: G. Wang et al. (eds.) Rough Sets and Knowledge Technology, LNCS 5009, pp.316–323, 2008.

[109] P. Maji, Fuzzy rough supervised attribute clustering algorithm and classification of microarray data, IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics 41 (2011) 222 –233.

[110] M. Sarkar, B. Yegnanarayana, Fuzzy-rough neural networks for vowel classification, in: IEEE International Conference on Systems, Man, and Cybernetics. Vol. 5, 4160–4165, 1998.

[111] J.Y. Zhao, Z. Zhang, Fuzzy rough neural network and its application to feature selection, in: Fourth International Workshop on Advanced Computational Intelligence (IWACI), pp.684–687, 2011.

[112] D. Zhang, Y. Wang, Fuzzy-rough neural network and its application to vowel recognition, in: 45th IEEE conference on Control and Decision, Vol. 21, pp.221–224, 2006.

[113] M. JianXu, L. Caiping, W. Yaonan, Remote sensing images classification using fuzzy-rough neural network, in: IEEE Fifth International Conference on Bio-Inspired Computing: Theories and Applications (BIC-TA), pp.761–765, 2010.

[114] M. Sarkar, B. Yegnanarayana, Application of fuzzy-rough sets in modular neural networks, in: IEEE Joint World Congress on Computational Intelligence and Neural Networks, Vol. 1, pp.741 –746, 1998.

[115] A. Ganivada, P. Sankar, A novel fuzzy rough granular neural network for classification, International Journal of Computational Intelligence Systems 4 (2011) 1042–1051.

[116] M. Sarkar, B. Yegnanarayana, Rough-fuzzy set theoretic approach to evaluate the importance of input features in classification, in: International Conference on Neural Networks, Vol. 3, pp.1590 –1595, 1997.

[117] A. Ganivada, S.S. Ray, S.K. Pal, Fuzzy rough granular self-organizing map and fuzzy rough entropy, Theoretical Computer Science 466 (2012) 37 –63.

[118] L. Jiangping, P. Baochang, W. Yuke, Tongue image segmentation based on fuzzy rough sets, in: Proceedings of the International Conference on Environmental Science and Information Application Technology (ESIAT), 367–369, 2009.

[119] L. Jiangping, W. Yuke, A shortest path algorithm of image segmentation based on fuzzy-rough grid, in: Proceedings of the International Conference on Computational Intelligence and Software Engineering (CiSE), 1–4, 2009.

[120] A. Petrosino, A. Ferone, Rough fuzzy set-based image compression, Fuzzy Sets and Systems 160 (2009) 1485–1506.

[121] L. Zhou, W. Li, Y. Wu, Face recognition based on fuzzy rough set reduction, in: Proceedings of the International Conference on Hybrid Information Technology (ICHIT), pp.642–646, 2006.

[122] A. Petrosino, G. Salvi, Rough fuzzy set based scale space transforms and their use in image analysis, International Journal of Approximate Reasoning 41 (2006) 212–228.

[123] A. Petrosino, M. Ceccarelli, Unsupervised texture discrimination based on rough fuzzy sets and parallel hierarchical clustering, in: Proceedings of the IEEE International Conference on Pattern Recognition, pp.1100–1103, 2000.

[124] X. Wang, J. Yang, X. Teng, N. Peng, Fuzzy-rough set based nearest neighbor clustering classification algorithm, in: Proceedings of the Second international conference on Fuzzy Systems and Knowledge Discovery, pp.370–373, 2005.

[125] C. Shang, Q. Shen, Aiding neural network based image classification with fuzzy-rough feature selection, in: Proceedings of the IEEE International Conference on Fuzzy Systems (FUZZ-IEEE), pp.976–982, 2008.

[126] S. Changjing, D. Barnes, S. Qiang, Effective feature selection for mars mcmurdo terrain image classification, in: Proceedings of the International Conference on Intelligent Systems, Design and Applications (ISDA), pp.1419–1424, 2009.

[127] D.V. Rao, V.V.S. Sarma, A rough-fuzzy approach for retrieval of candidate components for software reuse, Pattern Recognition Letters 24 (2003) 875–886.

[128] G. Cong, J. Zhang, T. Huazhong, K. Lai, A variable precision fuzzy rough group decision-making model for it offshore outsourcing risk evaluation, Journal of Global Information Management 16 (2008) 18–34.

[129] J. Xu, L. Zhao, A multi-objective decision-making model with fuzzy rough coefficients and its application to the inventory problem, Information Sciences 180 (2010) 679–696.

[130] J. Xu, L. Zhao, A class of fuzzy rough expected value multi-objective decision making model and its application to inventory problems, Computers and Mathematics with Applications 56(8) (2008) 2107-2119.

[131] B.Z. Sun, W.M. Ma, Soft fuzzy rough sets and its application in decision making, Artificial Intelligence Review (2011) 1–14.

[132] B.Z. Sun, W.M. Ma, Q. Liu, An approach to decision making based on intuitionistic fuzzy rough sets over two universes, Journal of the Operational Research Society 64(7) (2012) 1079–1089.

[133] T. Beaubouef, F. Petry, Fuzzy rough set techniques for uncertainty processing in a relational database, International Journal of Intelligent Systems 15(5) (2000) 389–424.

[134] R.R. Hashemi, F.F. Choobineh, A fuzzy rough sets classifier for database mining, International Journal of Smart Engineering System Design 4 (2002) 107–114.

[135] T.P. Hong, L.H. Tseng, B.C. Chien, Mining from incomplete quantitative data by fuzzy rough sets, Expert Systems with Applications 37 (2010) 2644 –2653.

[136] Y.F. Wang, Mining stock price using fuzzy rough set system, Expert Systems with Applications 24 (2003) 13 –23.

[137] A. Burney, N. Mahmood, Z. Abbas, Advances in fuzzy rough set theory for temporal databases, in: Proceedings of the 11th WSEAS international conference on Artificial Intelligence, Knowledge Engineering and Data Bases, pp.237–242, 2012.

[138] A. Burney, Z. Abbas, N. Mahmood, Q. ul Arifeen, Application of fuzzy rough temporal approach in patient data management (frt-pdm), International Journal of Computers 6 (2012) 149–157.

[139] P. Srinivasan, M. Ruiz, D.H. Kraft, J. Chen, Vocabulary mining for information retrieval: Rough sets and fuzzy sets, Information Processing and Management 37 (2001) 15–38.

[140] M. DeCock, C. Cornelis, Fuzzy rough set based web query expansion, in: Proceedings of Rough Sets and Soft Computing in Intelligent Agent and Web Technology, International Workshop at WIIAT2005, pp.9–16, 2005.

[141] L. Dey, M. Abulaish, R. Goyal, K. Shubham, A rough-fuzzy ontology generation framework and its application to bio-medical text processing, in: Proceedings of the 5th Atlantic Web Intelligence Conference (AWIC), pp.74–79, 2007.

[142] Y. Jiang, J. Wang, P. Deng, S. Tang, Reasoning within expressive fuzzy rough description logics, Fuzzy Sets and Systems 160 (2009) 3403 –3424.

[143] F. Bobillo, U. Straccia, Generalized fuzzy rough description logics, Information Sciences 189 (2012) 43–62.

[144] Y. Jiang, Y. Tang, J. Wang, S. Tang, Reasoning within intuitionistic fuzzy rough description logics, Information Sciences 179 (2009) 2362– 2378.

[145] F. Bobillo, U. Straccia, Supporting fuzzy rough sets in fuzzy description logics, in C. Sossai, G. Chemello (eds.) Symbolic and Quantitative Approaches to Reasoning with Uncertainty, LNCS 5590, 676–687, 2009.