International Journal on Approximate Reasoning 52 (2011) 1283–1297

Contents lists available at SciVerse ScienceDirect

International Journal on Approximate Reasoning

journal homepage: www.elsevier.com/locate/ijar

Dominance-based fuzzy rough set analysis of uncertain and possibilistic data tables

Tuan-Fang Fan [a], Churn-Jung Liau [b,*], Duen-Ren Liu [c]

[a] *Department of Computer Science and Information Engineering, National Penghu University of Science and Technology, Penghu 880, Taiwan*
[b] *Institute of Information Science, Academia Sinica, Taipei 115, Taiwan*
[c] *Institute of Information Management, National Chiao-Tung University, Hsinchu 300, Taiwan*

ARTICLE INFO

Article history:
Available online 1 February 2011

Keywords:
Dominance-based rough set approach
Multi-criteria decision analysis
Preference-ordered data tables
Rough set theory
Uncertain data tables
Possibilistic data table

ABSTRACT

In this paper, we propose a dominance-based fuzzy rough set approach for the decision analysis of a preference-ordered uncertain or possibilistic data table, which is comprised of a finite set of objects described by a finite set of criteria. The domains of the criteria may have ordinal properties that express preference scales. In the proposed approach, we first compute the degree of dominance between any two objects based on their imprecise evaluations with respect to each criterion. This results in a valued dominance relation on the universe. Then, we define the degree of adherence to the dominance principle by every pair of objects and the degree of consistency of each object. The consistency degrees of all objects are aggregated to derive the quality of the classification, which we use to define the reducts of a data table. In addition, the upward and downward unions of decision classes are fuzzy subsets of the universe. Thus, the lower and upper approximations of the decision classes based on the valued dominance relation are fuzzy rough sets. By using the lower approximations of the decision classes, we can derive two types of decision rules that can be applied to new decision cases.

1. Introduction

The theory of knowledge has long been an important topic in many academic disciplines, such as philosophy, psychology, economics, and artificial intelligence, whereas the storage and retrieval of data is the main concern of information science. In modern experimental science, knowledge is usually acquired from observed data, which is a valuable resource for researchers and decision-makers. In this respect, reasoning about data is a primary task in science. With the aid of computers, the data can be transformed into symbolic knowledge automatically. Thus, intelligent data analysis has received a great deal of attention in recent years. Rough set theory proposed in [27] provides a theoretical foundation for reasoning about data.

When rough set theory is applied to *multi-criteria decision analysis* (MCDA), it is crucial to deal with preference-ordered attribute domains and decision classes [8,9,12–15,18]. The original rough set theory cannot handle inconsistencies arising from violations of the dominance principle due to its use of the indiscernibility relation. Therefore, in the above-mentioned works, the indiscernibility relation is replaced by a dominance relation to solve the multi-criteria sorting problem; and the data table is replaced by a pairwise comparison table to solve multi-criteria choice and ranking problems. The approach is called the *dominance-based rough set approach* (DRSA). For MCDA problems, DRSA can induce a set of decision rules from sample decisions provided by decision-makers. The induced rules form a comprehensive preference model and can provide recommendations about a new decision-making environment.

A strong assumption about data tables is that each object takes exactly one value with respect to an attribute. However, in practice, we may only have incomplete information about the values of an object's attributes. Thus, more general data

* Corresponding author.
E-mail addresses: dffan@npu.edu.tw (T.-F. Fan), liaucj@iis.sinica.edu.tw (C.-J. Liau), dliu@iim.nctu.edu.tw (D.-R. Liu).

0888-613X/$ - see front matter
doi:10.1016/j.ijar.2011.01.009

tables are needed to represent incomplete information. For example, set-valued and interval-valued data tables have been introduced to represent incomplete information [20–22,24,34]; and DRSA has been extended to deal with missing values in MCDA problems [13]. Since a data table with missing values is a special case of uncertain or possibilistic data tables, we propose further extending DRSA to the decision analysis of these two kinds of data tables. This paper contains a theoretical investigation of such an extension based on the valued dominance principle.

While efficient rule induction algorithms are important for KDD applications, we believe that a clear interpretation of uncertain data and unambiguous specification of the induced model are prerequisites for further computational implementations. Therefore, we focus on the declarative aspects instead of the procedural aspects of the method. In other words, rather than provide an efficient implementation of the proposed method, we show the kinds of rules that can be induced given our interpretation of the uncertain data, and how the rules can be applied to new decision environments.

The proposed method first computes the degree of dominance between any two objects based on their imprecise evaluations with respect to each criterion. This results in a valued dominance relation on the universe. Then, we define the degree of adherence to the dominance principle by every pair of objects and the degree of consistency of each object. The consistency degrees of all objects are aggregated to derive the quality of the classification, which we use to define the reducts of decision tables. In addition, the upward and downward unions of decision classes are fuzzy subsets of the universe. Therefore, lower and upper approximations of the decision classes based on the valued dominance relation are fuzzy rough sets. By using the lower approximations of the decision classes, we can derive two types of decision rules that can be applied in new decision-making environments.

The remainder of this paper is organized as follows: In Section 2, we review the dominance-based rough set approach. In Section 3, we present a general framework of the dominance-based fuzzy rough set approach(DFRSA). Then, we reify the framework to deal with uncertain data tables and possibilistic data tables in Sections 4 and 5, respectively. In Section 6, we discuss our approaches and provide an extensive comparison with related works. Section 7 contains some concluding remarks.

2. Review of rough set theory and DRSA

The basic construct of rough set theory is an *approximation space*, which is defined as a pair (U, R), where U is a finite universe and $R \subseteq U \times U$ is an equivalence relation on U. We write an equivalence class of R as $[x]_R$ if it contains the element x. For any subset X of the universe, the lower approximation and upper approximation of X are denoted by $\underline{R}X$ and $\overline{R}X$, respectively, and defined as follows:

$$\underline{R}X = \{x \in U \mid [x]_R \subseteq X\}, \tag{1}$$

$$\overline{R}X = \{x \in U \mid [x]_R \cap X \neq \emptyset\}. \tag{2}$$

Although an approximation space is an abstract framework used to represent classification knowledge, it can easily be derived from a concrete data table (DT). In [28], a data table[1] is defined as a tuple $T = (U, A, \{V_i \mid i \in A\}, \{f_i \mid i \in A\})$, where U is a nonempty finite set, called the universe; A is a nonempty finite set of primitive attributes; for each $i \in A$, V_i is the domain of values of i; and for each $i \in A$, $f_i : U \to V_i$ is a total function. An attribute in A is usually denoted by the lower-case letters i or a. In decision analysis (and throughout this paper), we assume the set of attributes is partitioned into $\{d\} \cup (A - \{d\})$, where d is called the *decision attribute*, and the remaining attributes in $C = A - \{d\}$ are called *condition attributes*. Given a subset of attributes B, we can define an equivalence relation, called the *indiscernibility relation*, as follows:

$$ind(B) = \{(x, y) \mid x, y \in U, f_i(x) = f_i(y) \forall i \in B\}. \tag{3}$$

Consequently, for each $B \subseteq A$, $(U, ind(B))$ is an approximation space.

For MCDA problems, each object in a data table can be seen as a sample decision, and each condition attribute is a criterion for that decision. Since a criterion's domain of values is usually ordered according to the decision-maker's preferences, we define a preference-ordered data table (PODT) as a tuple $T = (U, A, \{(V_i, \succeq_i) \mid i \in A\}, \{f_i \mid i \in A\})$, where $(U, A, \{V_i \mid i \in A\}, \{f_i \mid i \in A\})$ is a classical data table; and for each $i \in A$, $\succeq_i \subseteq V_i \times V_i$ is a binary relation over V_i. The relation $\succeq_i$ is called a *weak preference relation* or *outranking* on V_i, and represents a preference over the set of objects with respect to the criterion i [31].

To deal with inconsistencies arising from violations of the dominance principle, the indiscernibility relation is replaced by a dominance relation in DRSA. Let P be a subset of criteria. Then, we can define the *P-dominance relation* $D_P \subseteq U \times U$ as follows:

$$(x, y) \in D_P \Leftrightarrow f_i(x) \succeq_i f_i(y) \forall i \in P. \tag{4}$$

When $(x, y) \in D_P$, we say that x P-dominates y, and that y is P-dominated by x. We usually use the infix notation xD_Py to denote $(x, y) \in D_P$. Given the dominance relation D_P, the P-dominating set and P-dominated set of x are defined as

[1] Also called knowledge representation systems, information systems, or attribute-value systems.

$D_P^+(x) = \{y \in U \mid yD_Px\}$ and $D_P^-(x) = \{y \in U \mid xD_Py\}$, respectively. In addition, for each $t \in V_d$, we define the decision class Cl_t as $\{x \in U \mid f_d(x) = t\}$. Then, the *upward and downward unions of classes* are defined as $Cl_t^{\geq} = \bigcup_{s \geq t} Cl_s$ and $Cl_t^{\leq} = \bigcup_{s \leq t} Cl_s$, respectively. We can then define the P-lower and P-upper approximations of $Cl_t^{\geq}$ and $Cl_t^{\leq}$ by using the P-dominating sets and P-dominated sets instead of the equivalence classes.

3. Dominance-based fuzzy rough set approach

In [10,16], fuzzy set extensions of the DRSA are proposed to refine the concept of dominance. In this section, we present a framework called the *dominance-based fuzzy rough set approach*(DFRSA), which is a slightly modified version of the extension in [16]. In the framework, the universe is endowed with a *valued preference relation* $D_i : U \times U \to [0, 1]$ for each criterion i and each decision class is a fuzzy subset of U. Thus, a *DFRSA frame* is a tuple $\mathfrak{F} = (U, A, \{D_i \mid i \in A\}, \{Cl_t^{\geq}, Cl_t^{\leq} \mid 1 \leq t \leq n\})$, where U is the universe, A is the set of criteria, each D_i $(i \in A)$ is a valued preference relation on U, and $Cl_t^{\geq}$ and $Cl_t^{\leq}$ $(1 \leq t \leq n$ for some $n > 1)$ are fuzzy subsets of U.

For any subset of criteria P, we can aggregate the valued preference relations $D_i(i \in P)$ into P-dominance relations. Let $\otimes$, $\oplus$ and $\to$ denote, respectively, a t-norm operation, an s-norm operation and an implication operation on [0, 1]. Then, the valued P dominance relation $D_P : U \times U \to [0, 1]$ is defined as

$$D_P(x, y) = \bigotimes_{i \in P} D_i(x, y). \tag{5}$$

Since the dominance relation is a valued relation, the satisfaction of the dominance principle is a matter of degree. Thus, the *degree of adherence* of (x, y) to the dominance principle with respect to a subset of condition criteria P is defined as

$$\delta_P(x, y) = D_P(x, y) \to D_d(x, y), \tag{6}$$

and the degree of P-consistency of x is defined as

$$\delta_P(x) = \bigotimes_{y \in U} (\delta_P(x, y) \otimes \delta_P(y, x)). \tag{7}$$

Let $\mathfrak{F}$ be a DFRSA frame. Then, the *quality of the classification* of $\mathfrak{F}$ based on the set of criteria P is defined as

$$\gamma_P(\mathfrak{F}) = \frac{\sum_{x \in U} \delta_P(x)}{|U|}. \tag{8}$$

Note that $\gamma_P(\mathfrak{F})$ is monotonic with respect to P, i.e., $\gamma_Q(\mathfrak{F}) \leq \gamma_P(\mathfrak{F})$ if $Q \subseteq P$. Thus, we can define every minimal subset $P \subseteq C$ such that $\gamma_P(\mathfrak{F}) = \gamma_C(\mathfrak{F})$ as a *reduct* of C, where $C = A - \{d\}$ is the set of all condition criteria. In addition, the degree of P-consistency is monotonic with respect to P, so a reduct is also a minimal subset $P \subseteq C$ such that $\delta_P(x) = \delta_C(x)$ for all $x \in U$. However, because $\delta_P(x)$ is less sensitive to individual changes in $\delta_P(x, y)$, we cannot guarantee that a reduct will preserve the degree of adherence to the dominance principle for each pair of objects. To resolve this difficulty, we adopt the following alternative definition of the quality of the classification:

$$\eta_P(\mathfrak{F}) = \frac{\sum_{x, y \in U} \delta_P(x, y)}{|U|^2}. \tag{9}$$

The reducts can also be defined in terms of the above equation.

Since our dominance relation is a valued relation and the decision classes are fuzzy sets, the lower and upper approximations of the classes are defined in the same way as those for fuzzy rough sets[2] [4,29,26,23,25,33]. More specifically, the P-lower and P-upper approximations of $Cl_t^{\geq}$ and $Cl_t^{\leq}$ for each $1 \leq t \leq n$ are defined as fuzzy subsets of U with the following membership functions:

$$\underline{P}(Cl_t^{\geq})(x) = \bigotimes_{y \in U} (D_P(y, x) \to Cl_t^{\geq}(y)), \tag{10}$$

$$\overline{P}(Cl_t^{\geq})(x) = \bigoplus_{y \in U} (D_P(x, y) \otimes Cl_t^{\geq}(y)), \tag{11}$$

$$\underline{P}(Cl_t^{\leq})(x) = \bigotimes_{y \in U} (D_P(x, y) \to Cl_t^{\leq}(y)), \tag{12}$$

$$\overline{P}(Cl_t^{\leq})(x) = \bigoplus_{y \in U} (D_P(y, x) \otimes Cl_t^{\leq}(y)). \tag{13}$$

[2] Although the [0, 1]-valued dominance relation is not restricted to being a fuzzy relation, the formal definition of the lower and upper approximations is the same as that for fuzzy rough sets.

Let P denote a reduct of a DFRSA frame and let $1 \leq t \leq n$. Then, for each object x, where $\underline{P}(Cl_t^{\geq})(x) > 0$ (or above some pre-determined threshold), we can derive the first type of fuzzy rule:

$$\underline{P}(Cl_t^{\geq})(x) : \bigwedge_{i \in P} (\succeq_i, f_i(x)) \longrightarrow (\succeq_d, t); \quad (14)$$

and for each object x, where $\underline{P}(Cl_t^{\leq})(x) > 0$ (or above some pre-determined threshold), we can derive the second type of fuzzy rule:

$$\underline{P}(Cl_t^{\leq})(x) : \bigwedge_{i \in P} (\preceq_i, f_i(x)) \longrightarrow (\preceq_d, t), \quad (15)$$

where $\underline{P}(Cl_t^{\geq})(x)$ and $\underline{P}(Cl_t^{\leq})(x)$ are the respective degrees of truth of the rules.

The DFRSA frame is simply an abstract framework. The valued preference relation is a primitive notion in the framework and the upward and downward unions of decision classes are not connected with the valued dominance relation $D_{\{d\}}$ in an explicit way. By contrast, in the application of DRSA to MCDA, the dominance relations and the upward and downward unions of decision classes are both derived from the evaluations and class assignments of sample decisions in a PODT. Thus, we are interested in reifying the DFRSA frame for specific MCDA problems. In particular, a DFRSA frame can be derived from sample decisions with imprecise evaluations and assignments. In the next two sections, we consider two concrete instances of the DFRSA frame derived through the reification process, where the valued preference relation is derived from the imprecise evaluations in a probabilistic way and a possibilistic way.

4. Preference-ordered uncertain data tables

Although a PODT can represent multi-criteria decision cases effectively, it inherits the restriction of the classical DT in that uncertain information cannot be represented. An uncertain data table is a generalization of a DT, so the values of some or all of its attributes are imprecise [20–22]. An analogous generalization can be applied to a PODT to define preference-ordered uncertain data tables (POUDT). Formally, a POUDT is a tuple $T = (U, A, \{(V_i, \succeq_i) \mid i \in A\}, \{f_i \mid i \in A\})$, where $U, A, \{(V_i, \succeq_i) \mid i \in A\}$ are defined as above; and for each $i \in A$, $f_i : U \rightarrow 2^{V_i} - \{\emptyset\}$. The intuition about a POUDT is that the evaluation of criterion i for an object x belongs to $f_i(x)$, although the evaluation is not known exactly. When $f_i(x)$ is a singleton, we say that the evaluation is precise. If all evaluations of T are precise, then T is said to be precise.

Since we do not have any further information about the evaluation, we employ the *principle of indifference* to assess the degrees of belief in its possible values. (For a justification of the principle, see [19].) This motivates us to assign the simplest non-informative prior, i.e., the uniform epistemic probability to the possible values of the evaluation. Therefore, we assume that, for each criterion i, the Cartesian plane $V_i \times V_i$ is endowed with a uniform measure μ_i. Thus, for each subset $S \subseteq V_i \times V_i$, $\mu_i(S)$ is a non-negative real number. When V_i is a finite set, we take $\mu_i(S)$ as the cardinality of S; and when V_i is a real interval, we take $\mu_i(S)$ as the area of S. The assumption may bias our treatment of incomplete information, since it does not necessarily reflect the proper epistemic status (i.e., partial ignorance) of the analyst who has the uncertain data [5,7]. However, the assumption is sometimes realistic when the data is sanitized via well-known privacy-preserving techniques, as shown in Section 6.2.

4.1. Valued preference relation

In a POUDT, the objects may have imprecise evaluations with respect to the condition criteria and imprecise assignments to decision classes. Thus, the preference relations between the objects cannot be determined with certainty. Instead, we can derive a degree of preference between two objects with respect to each criterion i based on the associated measures μ_i. Formally, the valued preference relation with respect to the criterion i is $D_i : U \times U \rightarrow [0, 1]$ such that for all $x \neq y$,

$$D_i(x, y) = \frac{\mu_i(\{(v_1, v_2) \mid v_1 \succeq_i v_2, v_1 \in f_i(x), v_2 \in f_i(y)\})}{\mu_i(f_i(x) \times f_i(y))}, \quad (16)$$

and $D_i(x, x) = 1$ for any $x \in U$.

Example 1. Fig. 1 shows an example of computing the degree of preference, where the evaluation of x with respect to criterion i, denoted by $s(x)$, is in a continuous interval $f_i(x) = [l_x, u_x]$. In this example, $D_i(x, y)$ is the ratio of the area of ABC over the area of $ABDE$, i.e., $\frac{u_x - l_x}{2(u_y - l_y)}$.

From the example, it is clear that $D_i(x, y) + D_i(y, x) = 1$ holds for the continuous domain V_i if $x \neq y$. This is a direct result of probability calculus, since $D_i(x, y)$ is actually the probability of $f_i(x) \succeq_i f_i(y)$ under the uniformity assumption. However, we note that, although $D_P(x, y)$ is equivalent to the conjunctive statement $\bigwedge_{i \in P}(f_i(x) \succeq_i f_i(y))$ in standard DRSA, the valued dominance $D_P(x, y)$ in our framework is not necessarily equal to the probability of $\bigwedge_{i \in P}(f_i(x) \succeq_i f_i(y))$. In other words,

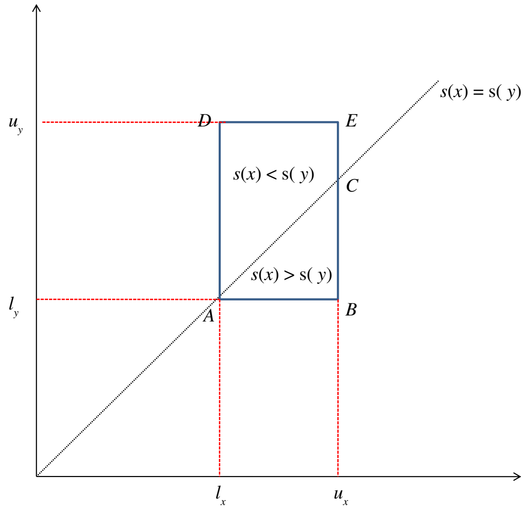

Fig. 1. The degree of dominance of x over y.

we consider $D_P(x, y)$ as simply the aggregation of individual $D_i(x, y)$ for $i \in P$ instead of the probability of the conjunctive statement, even though each $D_i(x, y)$ is regarded as the probability of the individual statement.

4.2. Upward and downward unions of decision classes

In a POUDT, the assignment of a decision label to an object may be imprecise, so the decision classes may be fuzzy subsets of the universe. First, for each decision label $t \in V_d$, the decision class $Cl_t : U \longrightarrow [0, 1]$ is defined by

$$Cl_t(x) = \begin{cases} \frac{1}{|f_d(x)|}, & \text{if } t \in f_d(x), \\ 0, & \text{otherwise.} \end{cases} \tag{17}$$

Second, the upward and downward unions of classes are defined by

$$Cl_t^{\geq}(x) = \frac{|f_d(x) \cap \{v \in V_d : v \geq t\}|}{|f_d(x)|} \tag{18}$$

and

$$Cl_t^{\leq}(x) = \frac{|f_d(x) \cap \{v \in V_d : v \leq t\}|}{|f_d(x)|} \tag{19}$$

respectively. Note that $Cl_t^{\geq} = \bigcup_{s \geq t} Cl_s$ and $Cl_t^{\leq} = \bigcup_{s \leq t} Cl_s$ only hold when we take the Łukasiewicz s-norm as the union operation, i.e., only when $(F \cup G)(x) = F(x) \oplus G(x)$, where $a \oplus b = \min(1, a + b)$.

4.3. Decision-making process

For a new decision case with (possibly imprecise) evaluations based on the condition criteria P, we can apply the decision rules (14) and (15) to derive its decision label assignment. Specifically, let x be a new object such that, for each criterion $i \in P, f_i(x) \subseteq V_i$ is given; and let α be a rule $c : \bigwedge_{i \in P}(\succeq_i, s_i) \longrightarrow (\geq_d, t)$ in the form of (14). Then, according to the rule α, we can derive that the degree of satisfaction of $f_d(x) \succeq_d t$, denoted by $\varepsilon(\alpha, f_d(x) \succeq_d t)$, is

$$c \otimes \bigotimes_{i \in P} \frac{\mu_i(\{(v_1, v_2) \mid v_1 \succeq_i v_2, v_1 \in f_i(x), v_2 \in s_i\})}{\mu_i(f_i(x) \times s_i)}. \tag{20}$$

Let $\mathcal{R}_t^{\geq}$ denote the set of all rules with the consequent $(\geq_d, t')$ such that $t' \geq t$. Then, the final degree of $f_d(x) \succeq_d t$ is

$$\bigoplus_{\alpha \in \mathcal{R}_t^{\geq}} \varepsilon(\alpha, f_d(x) \succeq_d t'). \tag{21}$$

We can derive the degree of $f_d(x) \preceq_d t$ from the second type of rule in a similar manner. Then, the degree of $f_d(x) = t$ is the t-norm conjunction of the degree of $f_d(x) \preceq_d t$ and the degree of $f_d(x) \succeq_d t$. In this way, we can apply the induced rules to rank the possible decision assignments for the new case.

4.4. Properties

In [1], it is shown that the DRSA approach to multiple criteria classification with imprecise evaluations and assignments preserves some well-known properties of rough approximations, such as rough inclusion, complementarity, the identity of boundaries, and precisiation. In the following, we examine some of those properties in our approach.

Proposition 1 (Rough inclusion). *Assume that the implication operation used in* (10)–(13) *satisfies the requirement that* $1 \to a = a$ *for all* $a \in [0, 1]$. *Then, for any* $t \in V_d$ *and* $P \subseteq C$, *the following properties hold:*

1. $\underline{P}(Cl_t^{\geq}) \subseteq Cl_t^{\geq} \subseteq \overline{P}(Cl_t^{\geq})$,
2. $\underline{P}(Cl_t^{\leq}) \subseteq Cl_t^{\leq} \subseteq \overline{P}(Cl_t^{\leq})$.

Proof. By the definitions of the lower and upper approximations, the left inclusion follows from the facts that $D_p(x, x) = 1$, $1 \to a = a$, and $a \otimes b \leq a$; and the right inclusion follows from the facts that $D_p(x, x) = 1$, $1 \otimes a = a$, and $a \leq a \oplus b$. □

To present the complementarity property, we need a negation function $\neg : [0, 1] \to [0, 1]$. In this paper, we assume the negation function is defined by $\neg a = 1 - a$ for any $a \in [0, 1]$. Thus, the complement of a fuzzy subset $S \subseteq U$ is defined by the membership function $(U - S)(x) = \neg S(x) = 1 - S(x)$ for all $x \in U$.

Proposition 2 (Complementarity). *Assume that the t-norm and implication operations used in* (10)–(13) *satisfy the requirement that* $\neg(a \to b) = a \otimes \neg b$ *for all* $a, b \in [0, 1]$. *Then, for any* $P \subseteq C$, *the following properties hold for* $1 \leq t \leq n - 1$:

1. $\overline{P}(Cl_t^{\leq}) = U - \underline{P}(Cl_{t+1}^{\geq})$,
2. $\underline{P}(Cl_t^{\leq}) = U - \overline{P}(Cl_{t+1}^{\geq})$.

Proof. The properties depend on the following crucial fact, which can be derived easily from (18) and (19).

$$Cl_t^{\leq}(x) = \neg Cl_{t+1}^{\geq}(x) \tag{22}$$

for $x \in U$ and $1 \leq t \leq n - 1$. □

The *precisiation of data* means any new information about objects in U that is added to the data tables. It can be a new criterion or a piece of information that further refines the subsets of evaluations or assignments of an object. Rough set approaches usually require that more precise information about objects does not reduce the lower approximations of the decision classes. Thus, two types of precisiation properties are investigated in [1]. One is for the situation when new attributes or criteria are added to the decision table; and the other is for when more specific information about the evaluations and assignments of objects is added. The first type depends on the monotonicity of the t-norm, s-norm, and implication operations. Recall that the t-norm and s-norm operations are non-decreasing in their respective arguments, while the implication operation should be non-increasing in its left argument and non-decreasing in its right argument.

Proposition 3 (Precisiation). *For any* $R \subseteq P \subseteq C$ *and* $t \in V_d$, *we have*

1. $\underline{R}(Cl_t^{\geq}) \subseteq \underline{P}(Cl_t^{\geq}) \subseteq \overline{P}(Cl_t^{\geq}) \subseteq \overline{R}(Cl_t^{\geq})$,
2. $\underline{R}(Cl_t^{\leq}) \subseteq \underline{P}(Cl_t^{\leq}) \subseteq \overline{P}(Cl_t^{\leq}) \subseteq \overline{R}(Cl_t^{\leq})$.

Proof. According to the t-norm aggregation, $D_P(x, y) \leq D_R(x, y)$ for all $x, y \in U$. Then, the results follow from the monotonicity of the t-norm operation with respect to both of its arguments and the anti-monotonicity of the implication function with respect to its left argument. □

However, the second type of precisiation property does not hold for our fuzzy rough approximations. This is because, according to (16), the degree of dominance between two objects does not change monotonically with the precision of their evaluations or assignments. Furthermore, according to (18) and (19), the upward and downward unions of decision classes may also change non-monotonically with the precision of the assignments to objects. We consider this reasonable because the qualitative precisiation property may not capture the quantitative uncertainty encoded in our valued dominance relation.

5. Preference-ordered possibilistic data tables

In a POUDT, the information function f_i of a criterion i assigns a subset of possible values $f_i(x)$ to each object $x \in U$. As mentioned earlier, the subset $f_i(x)$ can be interpreted as follows: the precise evaluation (or assignment) of x on criterion i is not known precisely, although a range for the value is known and restricted to the subset $f_i(x)$. Moreover, since no further information is specified, it is assumed that each value in the subset $f_i(x)$ is equally possible. In terms of possibility theory [36], $f_i(x)$ is simply a special kind of possibility distribution, where the degrees of possibility are restricted to 0 or 1. Thus, it is natural to extend the result in the previous section to more general preference-ordered uncertain data tables, in which the uncertainty of information is specified by general possibility distributions. To present our approach in this general setting, we first define *preference-ordered possibilistic data tables* (POPDT). Recall that a possibility distribution on a domain V is simply a function $\pi : V \to [0, 1]$. Intuitively, π specifies the degree of possibility of each element in the domain V. Here, $\pi(v) = 1$ and $\pi(v) = 0$ mean that the element v is fully possible and totally impossible, respectively; while the intermediate values in $(0, 1)$ mean partial possibilities of v. We usually assume that a possibility distribution is *normalized*, i.e., $\sup_{v \in V} \pi(v) = 1$. Let π_1 and π_2 be two possibility distributions on V. Then, we say that π_1 is *at least as specific as* π_2, denoted by $\pi_1 \leq \pi_2$, if $\pi_1(v) \leq \pi_2(v)$ for each $v \in V$. Let us denote the set of all normalized possibility distributions on V by $(V \to [0, 1])^+$. Then, a POPDT is a tuple $T = (U, A, \{(V_i, \succeq_i) \mid i \in A\}, \{f_i \mid i \in A\})$, where $U, A, \{(V_i, \succeq_i) \mid i \in A\}$ are defined as above, and for each $i \in A, f_i : U \to (V_i \to [0, 1])^+$.

5.1. Fuzzy weak preference relations

In a POUDT, each value in a subset $f_i(x)$ is considered equally possible; therefore, we can use a uniform measure μ_i to determine the probability of x being preferred to y based on the criterion i. However, since possibility information for each possible value is available in a POPDT, we can use the extension principle in fuzzy set theory to compute the degree of preference [35]. The extension principle extends an operation or a relation over a base domain to the class of all fuzzy sets or possibility distributions over the domain. In our context, we use the extension principle to extend the preference relation $\succeq_i$ on V_i to a valued preference relation between two possibility distributions on V_i. Consequently, the valued preference relation between two objects with respect to the criterion i is determined by their respective possibility distributions over the domain of the criterion. Thus, the valued preference relation is also called the fuzzy weak preference relation. Formally, the fuzzy weak preference relation with respect to the criterion i is a fuzzy relation $D_i : U \times U \to [0, 1]$ such that

$$D_i(x, y) = \sup\{f_i(x)(v_1) \otimes f_i(y)(v_2) \mid v_1 \succeq_i v_2, v_1, v_2 \in V_i\}. \tag{23}$$

5.2. Upward and downward unions of classes

For a given POPDT, the decision classes are still fuzzy subsets of U, but their membership functions are derived from the possibility distributions associated with the assignments of the objects. For each $t \in V_d$, the decision class $Cl_t : U \to [0, 1]$ is defined by

$$Cl_t(x) = f_d(x)(t); \tag{24}$$

thus, the the upward and downward unions of classes are defined by

$$Cl_t^{\geq}(x) = \sup_{v \geq t} f_d(x)(v) = \Pi_x(\{v \geq t\}) \tag{25}$$

and

$$Cl_t^{\leq}(x) = \sup_{v \leq t} f_d(x)(v) = \Pi_x(\{v \leq t\}) \tag{26}$$

respectively, where Π_x is the possibility measure corresponding to the possibility distribution $f_d(x)$.

5.3. Decision-making process

For a new decision case with evaluations based on the condition criteria P, we can apply the decision rules in (14) and (15) to derive the case's decision label assignment. Specifically, let x be a new object such that, for each criterion $i \in P$, $f_i(x) \in (V_i \to [0, 1])^+$ is given; and let α be a rule $c : \bigwedge_{i \in P}(\succeq_i, \pi_i) \longrightarrow (\succeq_d, t)$. Then, according to the rule α, we can derive that the degree of satisfaction of $f_d(x) \succeq_d t$, denoted by $\varepsilon(\alpha, f_d(x) \succeq_d t)$, is

$$c \otimes \bigotimes_{i \in P} \bigoplus_{v_1 \succeq_i v_2} (f_i(x)(v_1) \otimes \pi_i(v_2)). \tag{27}$$

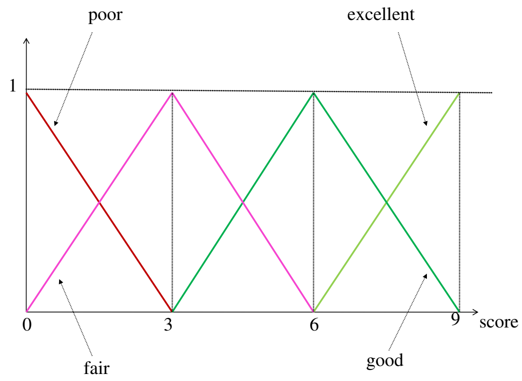

Fig. 2. The possibility distributions on the domain of "score".

Let $\mathcal{R}_t^{\geq}$ denote the set of all rules with the consequent $(\succeq_d, t')$ such that $t' \geq t$. Then, the final degree of $f_d(x) \succeq_d t$ is

$$\bigoplus_{\alpha \in \mathcal{R}_t^{\geq}} \varepsilon(\alpha, f_d(x) \succeq_d t'). \tag{28}$$

We can derive the degree of $f_d(x) \preceq_d t$ from the second type of rule in a similar manner. Therefore, as in the case of a POUDT, we can apply the induced rules to rank the possible decision assignments for the new case.

Mathematically, the evaluations and assignments in a POPDT are possibility distributions, so the antecedents of the rules may also include any possibility distributions on the domain. However, in general, the set of all (normalized) possibility distributions is infinite, even though the domain is finite. This may result in a very large set of rules. Moreover, most of the possibility distributions may lack semantically meaningful interpretations for human users; hence, the induced rules may be hard to use. To resolve the difficulty, the standard practice in fuzzy logic is to use a set of meaningful linguistic labels whose interpretations are simply possibility distributions on the domain. Thus, the evaluations and assignments given in a POPDT are restricted to the (usually finite) set of linguistic labels, so the set of atomic formulas in our rules only contains $(\succeq_i, \pi_i)$ or $(\preceq_i, \pi_i)$, where π_i is a linguistic label. For example, if the evaluated criterion is "score" and its domain is $[0, 9]$, then the set of linguistic labels may be {poor, fair, good, excellent}, and their corresponding interpretations are possibility distributions, as shown in Fig. 2.

5.4. Properties

In this section, we investigate the properties mentioned in the preceding section for fuzzy rough approximations in a POPDT. The rough inclusion property is exactly the same as above. However, since (22) no longer holds for the upward and downward unions of decision classes defined by (25) and (26), the complementarity property is not the same as that described above. Instead, we utilize the following fact to formulate an alternative complementarity property

$$\max(Cl_t^{\leq}(x), Cl_{t+1}^{\geq}(x)) = 1 \tag{29}$$

for $x \in U$ and $1 \leq t \leq n-1$.

Proposition 4 (Complementarity)**.** *Assume that the t-norm, s-norm, and implication operations applied in* (10)–(13) *satisfy the following condition for any finite index set I and $\{a_i, b_i \mid i \in I\} \subseteq [0, 1]$:*

$$\bigoplus_{i \in I} a_i \oplus \bigotimes_{i \in I} (a_i \rightarrow b_i) = 1. \tag{30}$$

Then, for any $P \subseteq C$, the following properties hold for $1 \leq t \leq n-1$:

1. $\overline{P}(Cl_t^{\leq}) \cup \underline{P}(Cl_{t+1}^{\geq}) = U,$
2. $\underline{P}(Cl_t^{\leq}) \cup \overline{P}(Cl_{t+1}^{\geq}) = U.$

Proof. For a given $x \in U$ and $1 \le t \le n-1$, let $a_y = D_P(y, x)$, $b_y = Cl^{\ge}_{t+1}(y)$, and $c_y = Cl^{\le}_t(y)$. Then, $\max(b_y, c_y) = 1$ for any $y \in U$. Let $I = \{y \mid b_y < 1\}$. Then, $c_y = 1$ for $y \in I$. Thus, by (10) and (13), we have

$$\underline{P}(Cl^{\ge}_{t+1})(x) = \bigotimes_{y \in I}(a_y \to b_y),$$

$$\overline{P}(Cl^{\le}_t)(x) = \bigoplus_{y \in I} a_y \oplus \bigoplus_{y \in U-I}(a_y \otimes c_y).$$

Therefore,

$$\underline{P}(Cl^{\ge}_{t+1})(x) \oplus \overline{P}(Cl^{\le}_t)(x) = 1 \oplus \bigoplus_{y \in U-I}(a_y \otimes c_y) = 1$$

by assumption (30). Since this holds for any $x \in U$, we have $\overline{P}(Cl^{\le}_t) \cup \underline{P}(Cl^{\ge}_{t+1}) = U$. The second equation can be proved in an analogous manner. □

The first type of precisiation property still holds for the fuzzy rough approximations in a POPDT. However, in contrast to the complete absence of the second type of precisiation property in a POUDT, we have the following result for a POPDT because of the monotonic change in the degree of dominance with respect to the specificity of the possibility distributions.

Proposition 5 (The second type of precisiation). *Let $T = (U, A, \{(V_i, \succeq_i) \mid i \in A\}, \{f_i \mid i \in A\})$ and $T' = (U, A, \{(V_i, \succeq_i) \mid i \in A\}, \{f'_i \mid i \in A\})$ be two POPDTs, where $A = C \cup \{d\}$ such that $f_d = f'_d$ and $f_i(x) \le f'_i(x)$ for all $i \in C$ and $x \in U$. In other words, the evaluations of objects in T are at least as specific as those in T'. Then, for any $P \subseteq C$ and $t \in V_d$, we have*

1. $\underline{P'}(Cl^{\ge}_t) \subseteq \underline{P}(Cl^{\ge}_t) \subseteq \overline{P}(Cl^{\ge}_t) \subseteq \overline{P'}(Cl^{\ge}_t)$,
2. $\underline{P'}(Cl^{\le}_t) \subseteq \underline{P}(Cl^{\le}_t) \subseteq \overline{P}(Cl^{\le}_t) \subseteq \overline{P'}(Cl^{\le}_t)$,

where the P'-approximations indicate the approximations based on the fuzzy dominance relations derived from T' with the same set of criteria P.

Proof. Note that $f_d = f'_d$ implies that $Cl^{\ge}_t$ (resp. $Cl^{\le}_t$) remains the same in both tables. Thus, we can consider the P-approximations and P'-approximations of $Cl^{\ge}_t$ (resp. $Cl^{\le}_t$) at the same time. By Eq. (23), $D_i(x, y) \le D'_i(x, y)$ for any $x, y \in U$ and $i \in A$ due to the monotonicity of the t-norm and sup operations, so $D_P(x, y) \le D'_P(x, y)$ for any $x, y \in U$. Then, by Eqs. (10)–(13), the left inclusion follows from the anti-monotonicity of the implication operator with respect to its first argument; the right inclusion follows from the monotonicity of the t-norm; and the middle inclusion is simply the property of rough inclusion (Proposition 1). □

6. Related works and discussion

A large number of works use rough set-based approaches to deal with different types of data for MCDA problems. The works can be divided into six categories based on the techniques employed (DRSA or DFRSA) and the types of data dealt with (precise, uncertain, or fuzzy data). Our studies of POUDTs and POPDTs belong to the categories of DFRSA for uncertain data and DFRSA for fuzzy data, respectively. Thus, we compare our work with the representative works in the remaining four categories.

6.1. DRSA for precise data

The first category comprises works on standard applications of DRSA to PODTs. Several representative works in this category were cited in Section 1 and the basic concept of the approach was reviewed in Section 2; therefore, we only include it here as a baseline for comparison. Obviously, a POUDT and a POPDT are both generalizations of a PODT, and our approach is reduced to the standard DRSA when the given data table is precise.

6.2. DRSA for uncertain data

As mentioned in Section 1, the concept of incomplete or multi-valued information systems has been considered in classical rough set approaches. In the context of DRSA, the problem of missing values was examined in [11]. Since an information system with missing values is a special case of a POUDT, it is also desirable to extend DRSA to deal with more general information systems that have "partially missing values". This issue was addressed in Dembczynski et al.'s recent work [1], which is representative of the second category. In [1], the imprecision of data is represented by an interval of possible values. Thus, for each decision criterion i and sample decision x, $f_i(x)$ is represented as an interval $[l_i(x), u_i(x)]$, where

$l_i(x), u_i(x) \in V_i$ and $l_i(x) \le u_i(x)$. Based on the interval-valued evaluations, three kinds of weak preference relations are defined with respect to a criterion i:

- A possible weak preference (PWP) relation: $x \overline{\succeq}_i y$ iff $u_i(x) \ge l_i(y)$.
- A lower-end weak preference (LWP) relation: $x \succeq_i^l y$ iff $l_i(x) \ge l_i(y)$.
- An upper-end weak preference (UWP) relation: $x \succeq_i^u y$ iff $u_i(x) \ge u_i(y)$.

For any $x, y \in U$ and $P \subseteq C$, the *generalized dominance principle* in [1] is formulated as follows:

$$[x\overline{D}_P y \Rightarrow xD_d^u y] \wedge [y\overline{D}_P x \Rightarrow yD_d^l x], \tag{31}$$

where $\overline{D}_P$, D_d^u, and D_d^l are the aggregated dominance relations of the respective weak preference relations.

Unlike interval-valued data, an uncertain value in a POUDT cannot be characterized simply by its lower and upper bounds because it may be an arbitrary subset of the domain of values. This results in a semantic difference between the weak preference relations over intervals and those over subsets of values. While PWP, LWP, and UWP are appropriate for the comparison of two intervals, they cannot completely characterize the preferences between two subsets of values. In a POUDT, two subsets of values may have the same lower and upper bounds. Thus, instead of using three qualitative preference relations that may be unable to distinguish between the two subsets, we use a quantitative relation to compare the degree of preference between them.

An implication of such a semantic difference is that these two approaches may yield different sets of consistent objects, even when they are only applied to interval-valued data. The following example illustrates the difference.

Example 2. Let us consider a single criterion i and the decision attribute d. Assume that x and y are two objects such that $f_i(x) = [1, 3], f_i(y) = [3, 5], f_d(x) = [2, 2]$, and $f_d(y) = [3, 3]$. Then, since $x\overline{D}_i y$ and $y\overline{D}_i x$ hold simultaneously, but $f_d(x) = f_d(y)$ does not hold, x and y violate the generalized dominance principle in [1]. Thus, x and y are regarded as $\{i\}$-inconsistent objects in [1]. However, according to DFRSA in this paper, $D_i(x, y) = 0$, $D_d(x, y) = 0$, $D_i(y, x) = 1$, and $D_d(y, x) = 1$, so the degree of adherence of (x, y) to the dominance principle with respect to $\{i\}$ is 1. This seems quite reasonable since we are (almost) certain that y is preferred over x in criterion i, and that y is assigned to a better decision class than x, although we do not know the i-evaluations of x and y exactly.

The semantic difference also explains why the *reclassification property* holds for the model in [1], but it does not apply to our induced rules. The property requires that the set of rules applied to the training samples should restore the original assignments of consistent objects given by the decision-makers. In fact, the decision rules induced in [1] treat the lower bound and the upper bound of the decision assignment as individual values in the same way as the standard DRSA. Thus, the original bounds of the interval decision can be restored if the condition parts of the new object are the same as those of the training sample. However, in our interpretation of uncertainty, the same interval evaluation does not mean the same decision situation, so requiring that the same decision interval must be assigned is unreasonable.

Example 3. For ease of explanation, let us consider a single criterion decision problem. Let i be the criterion and let x be a sample decision such that $f_i(x) = [1, 4]$ and $f_d(x) = \{2\}$. Then, based on our interpretation, the real evaluation of the case with respect to the criterion i lies somewhere between 1 and 4. Based on the principle of indifference, we can assume that all values in the range $[1, 4]$ are equally possible for the real evaluation. A cautious interpretation of the decision is that an evaluation in that range will assign the sample to class 2. Thus, if we have a new object whose evaluation is 3, the probability that it will be preferred over the sample object is $\frac{3}{4}$, so the certainty of it being assigned to at least class 2 should also be $\frac{3}{4}$. Taking the same cautious approach, if we have a new object with the same imprecise evaluation $[1, 4]$, it cannot be assumed that the new object will have the same evaluation as the sample object. Hence, requiring that the new object should be assigned to the same class as the sample object is unreasonable. Instead, the new object is only given a preferred evaluation with the probability $\frac{1}{2}$, so we can only be half certain that it will be assigned to a class at least class 2.

The next example describes further application scenarios of the two approaches. The scenarios show that our theoretical framework is potentially applicable to realistic privacy-preserving decision analysis.

Example 4. Let us consider a sample decision case x with a single criterion i such that $f_i(x) = [1, 3]$ and $f_d(x) = \{2\}$. For ease of presentation, we assume that the domain of the criterion i is discrete, so actually $f_i(x) = \{1, 2, 3\}$. The sample decision may arise from different scenarios. In one scenario, the decision maker was presented with an incomplete specification of the criterion i for the case, i.e., $\{1, 2, 3\}$, and was asked to judge the appropriate decision class of the case. He finally assigned it to class 2. Thus, we have a sample decision with uncertain information. In another scenario, the decision maker was presented with a precise specification of the criterion i of the case, say 3, and was asked to judge the appropriate decision class of the case. He finally assigned it to class 2. However, due to the privacy concern (e.g., the criterion may represent the income of the case), the sample decision was sanitized before it was released to the public. One of the most popular sanitization approaches in privacy protection is called *generalization*, i.e., the i-evaluation is replaced with a subset of V_i [32]. Assume that the precise sample decision is then sanitized by replacing its i-evaluation with $\{1, 2, 3\}$. Then, we have the same uncertain sample decision as that obtained in the first scenario.

Although both of the scenarios generate the same sample decision, different decision rules should be induced from the decision data. In the first scenario, the decision maker's intention is that anyone with the i-evaluation falling in $[1, 3]$ could be assigned to class 2. The following two qualitative rules can reasonably be derived from such an intention:

$$f_i(x) \geq 1 \longrightarrow f_d(x) \geq 2,$$

$$f_i(x) \leq 3 \longrightarrow f_d(x) \leq 2.$$

Thus, the qualitative interpretation of the uncertain sample decision fits this kind of scenario well. Moreover, when the system is presented with a new decision case y with $f_i(y) = [1, 3]$, it should assign y to at least class 2 since this is exactly the same as the situation under which the sample decision was made. Therefore, the reclassification property is appropriate for the scenario.[3]

However, these two rules are not well-suited to the second scenario, since we do not know the original decision maker's situation exactly. It may be $f_i(x) = 1 \wedge f_d(x) = 2$, $f_i(x) = 2 \wedge f_d(x) = 2$, or $f_i(x) = 3 \wedge f_d(x) = 2$. For each of the above situations, one of the following three (upward) rules can reasonably be derived:

(i) $f_i(x) \geq 1 \longrightarrow f_d(x) \geq 2$,
(ii) $f_i(x) \geq 2 \longrightarrow f_d(x) \geq 2$,
(iii) $f_i(x) \geq 3 \longrightarrow f_d(x) \geq 2$.

In the first situation, the induced rule is the same as that induced in the qualitative approach. However, in the other two situations, we can only induce one of the weaker rules (ii) or (iii). Since do not know x's real situation, we apply the principle of indifference to the three situations and assume that the probability of each situation is $\frac{1}{3}$. Since rule (i) implies rule (ii), which in turn implies rule (iii), rule (iii) can be derived in all three situations, so we can attach a certainty degree 1 to rule (iii). Similarly, rule (ii) (resp. (i)) can be induced in two (resp. one) of the three situations, so its certainty degree is $\frac{2}{3}$ (resp. $\frac{1}{3}$). The three rules are summarized as a rule of type (14) as follows:

$$1 : (\geq_i, [1, 3]) \longrightarrow (\geq_d 2).$$

Thus, our DFRSA could be applied to such scenario. Furthermore, when a sanitized case y with $f_i(y) = [1, 3]$ is input to the system, the system must also consider the three possible situations $f_i(y) = 1, 2$, or 3. Combining these situations with the possible situations for the sample decision x, there are in total 9 situations; however, the system can only conclude that $f_d(y) \geq 2$ in 6 of them. Thus, by using the above quantitative rule, the degree of certainty that y will be assigned to at least class 2 is $\frac{2}{3}$, as calculated with (20). This explains why the reclassification property is not appropriate for applications based on this scenario.

The above example highlights the fact that the same form of uncertain data could be interpreted and processed in different ways. Qualitative rules can be induced from past decision cases made under *uncertain* environments. When the induced decision-making model faces the same uncertain input as a previous consistent case, it should restore the original assignment of the previous case. However, this kind of approach is not appropriate for sanitized previous decision cases in *precise* environments. Since we are not sure about the original decision environments of these previous cases, we can only be partially certain about the decision assignments. This is the reason that a quantitative framework, such as DFRSA, is used. With a mild assumption of uniformity, DFRSA can quantify different possibilities of the decision assignments. When such a quantitative rule is applied to a new case of sanitized input, we cannot guarantee that the decision environment is exactly the same as that of a previous case, even though the cases have the same sanitized evaluations. Thus, the original assignment of the previous case cannot be fully restored. This explains why the reclassification property fails in such applications. However, the failure of the property does not render the induced model useless. In fact, as shown in Section 5.3, our quantitative rules can be used to rank the possible decision assignments for the new case. This is also verified in the example. Although we cannot restore the original assignment due to the uncertainty about the real situations, we can quantitatively assess the possibility of the case being assigned to at least class 2 in a reasonable way.

The semantic difference between these two scenarios can be expressed more precisely in logical terms. Let $S_i \subseteq V_i$ and $S_d \subseteq V_d$ denote, respectively, a finite imprecise evaluation and a finite imprecise assignment of an uncertain decision sample.

[3] The example is over-simplified so that, for the first scenario, we can replace the sample decision with three precise decisions, namely, $(f_i(x) = 1, f_d(x) = 2)$, $(f_i(x) = 2, f_d(x) = 2)$, $(f_i(x) = 3, f_d(x) = 2)$, and apply the standard DRSA to their analysis. However, for continuous domains or sample decisions with imprecise assignments, such a straightforward replacement may be not available.

Then, the qualitative interpretation considers the uncertain decision case (S_i, S_d) as a decision logic [28] rule:

$$\left(\bigvee_{v\in S_i}(f_i, v)\right) \longrightarrow \left(\bigvee_{w\in S_d}(f_d, w)\right);$$

whereas our DFRSA interpretation considers it as the disjunction of several rules:

$$\bigvee_{v\in S_i, w\in S_d}((f_i, v) \longrightarrow (f_d, w)).$$

Thus, when a premise $\bigvee_{v\in S_i}(f_i, v)$ is given, the former rule can be used to derive the conclusion $\bigvee_{w\in S_d}(f_d, w)$ by modus ponens, so the reclassification property holds. However, the conclusion is not derivable by the latter formula. This is why the reclassification property fails in the latter interpretation of uncertain samples.

Finally, while we focus on comparing the model in [1] with our DFRSA for a POUDT, we note that a POUDT (and an interval-valued data table) is a special kind of POPDT. Moreover, when the definition of a fuzzy weak preference in Eq. (23) is applied to interval data tables, D_i is reduced to the PWP relation in [1]. We have demonstrated the difference between the PWP relation and the valued preference defined in Eq. (16). This also implies that our DFRSA for a POPDT is not a proper generalization of our DFRSA for a POUDT, even though a POPDT is more general than a POUDT.

6.3. DRSA for fuzzy data

A less invasive use of fuzzy techniques when processing fuzzy data is provided by the model in [16], which belongs to the category of DRSA for fuzzy data. Instead of using fuzzy connectives, the model basically applies DRSA to the analysis of fuzzy data. The rationale is that the membership functions of fuzzy features are regarded as attributes with the domain [0, 1]. Formally, a *fuzzy information base* is a triplet $(U, A, \{f_i \mid i \in A\})$, where U is the universe, A is a finite set of features, and $f_i : U \to [0, 1]$ is the membership function of feature i. A fuzzy information base can easily be regarded as a PODT if $V_i = [0, 1]$ and $\succeq_i$ is the ordinary numerical ordering $\geq$ on [0, 1] for each feature i. Then, for any fuzzy subset X of the universe, the standard DRSA can be applied to find the lower and upper approximations of the (strict) c-cuts of X. From those approximations, the induced decision rules take the following form:

> if x has feature i_1 in degree at least h_1, feature i_2 in degree at least $h_2, \cdots$, and feature i_m in degree at least h_m, then x belongs to the subset X in degree at least c.

Since a PODT is a special case of both a POUDT and a POPDT, our approaches trivially reduce to the model in [16] for fuzzy information bases. However, in addition to the trivial translation of a fuzzy information base into a PODT, an alternative (but quite natural) interpretation is to consider the membership function of a feature as a possibility distribution on {0, 1}. This results in a somewhat less trivial translation of a fuzzy information base into a POPDT. In this translation, the POPDT corresponding to the fuzzy information base $(U, A, \{f_i \mid i \in A\})$ is $(U, A, \{(V_i, \succeq_i) \mid i \in A\}, \{p_i \mid i \in A\}$ such that $V_i = \{0, 1\}$, $1 \succ_i 0$ for $i \in A$, and $p_i(x)(1) = f_i(x)$ and $p_i(x)(0) = 1 - f_i(x)$ with necessary normalization for $x \in U$ and $i \in A$.

Example 5. Let us consider a case where the t-norm and s-norm operations are *min* and *max*, respectively. Assume that $x, y \in U$ are two objects in a fuzzy information base such that $f_i(x) = r \geq f_i(y) = s$ for some feature i. Then, according to the translation above and Eq. (23), we can obtain the following results on the fuzzy dominance relation D_i:

(1) if $r \geq s \geq \frac{1}{2}$ or $s \leq r \leq \frac{1}{2}$, then $D_i(x, y) = D_i(y, x) = 1$;
(2) if $r > \frac{1}{2} > s$, then $D_i(x, y) = 1$ and $D_i(y, x) = \max\left(\frac{1-r}{r}, \frac{s}{1-s}\right)$.

On the other hand, if we view the fuzzy information base as a PODT and take it as a special case of POPDT, then $D_i(x, y) = 1$ and $D_i(y, x) = 0$ if $r > s$ and $D_i(x, y) = D_i(y, x) = 1$ if $r = s$.

The above example demonstrates the difference between the two translations of a fuzzy information base into a POPDT. The phenomenon arises because of the well-known distinction between fuzziness and uncertainty [6]. In the translation via a PODT, $f_i(x) = r$ is interpreted as "the possibility that x has feature i with degree r is 1, and the possibility that x has feature i with degree other than r is 0", which is equivalent to "x has feature i with degree r". In other words, i is a fuzzy feature and an object can have the feature partially. This is the fuzziness interpretation of a fuzzy information base. On the other hand, the alternative translation is based on the uncertain interpretation of a fuzzy information base, whereby each feature is considered crisp. Thus, an object may or may not have the feature, although we are uncertain about which one is true. Consequently, $f_i(x) = r$ is interpreted as "the possibility that x has feature i is r, and the possibility that x does not have feature i is $1 - r$". The two translations show that our treatment of a POPDT can accommodate both interpretations of fuzzy information bases.

6.4. DFRSA for PODTs

In contrast to the preceding category in which the classical technique is applied in the processing of fuzzy data, it is also possible to use the fuzzy technique to deal with classical MCDA problems. The concept was elucidated in a recent work on the fuzzy preference-based rough set approach [17]. The work presents a method for extracting fuzzy preference relations from samples characterized by numerical criteria. Although the dominance relation in the standard DRSA can represent the qualitative preference between objects, it does not utilize the quantitative difference between their numerical evaluations. Since the quantitative difference can measure how much one object is better than another object, it is natural to derive a fuzzy preference relation from such a measure. In [17], it is suggested that the Logsig sigmoid transformation of $f_i(x) - f_i(y)$ should be used as the degree of preference of x over y with respect to the numerical criterion i.

While the Logsig transformation may be crucial for the approach in [17], other transformations also seem feasible. In fact, the standard DRSA is simply the result of choosing a two-valued transformation $\delta : \Re \to \{0, 1\}$ such that $\delta(v) = 1$ if $v \geq 0$ and $\delta(v) = 0$ if $v < 0$. Since our treatment of a POUDT and a POPDT is a generalization of the standard DRSA, our approach only considers the qualitative preference between the objects. Nevertheless, our approach can easily be generalized to accommodate the quantitative difference between their numerical evaluations. Let $\tau : \Re \times \Re \to [0, 1]$ denote an arbitrary transformation function. Then, for the POUDT, the degree of preference in Eq. (16) can be generalized as follows:

$$D_i(x, y) = \frac{1}{\mu_i(f_i(x) \times f_i(y))} \cdot \int_{(v_1, v_2) \in f_i(x) \times f_i(y)} \tau(v_1, v_2) d\mu_i. \tag{32}$$

On the other hand, when the evaluations of the objects are possibility distributions on numerical domains, we can regard them as fuzzy numbers. Thus, for a POPDT, the degree of preference in Eq. (23) becomes a fuzzy number $D_i(x, y) = \hat{\tau}(f_i(x), f_i(y))$, where $\hat{\tau}$ is the operation on fuzzy numbers derived from τ according to the extension principle. We defer the investigation of the properties of these more generalized treatments to a future work.

6.5. DFRSA for POUDTs and POPDTs

It seems paradoxical that the second precisiation property (Proposition 5) holds in DFRSA for a POPDT, but not for a POUDT, since the former is more general than the latter. However, this is completely reasonable given our treatment of POUDTs and POPDTs. We indicated in Section 6.2 that the definition of the fuzzy preference for a POPDT is not a generalization of the valued preference for a POUDT. Let us take a closer look at the difference for the case of finite domains. We consider a single criterion i with a finite preference-ordered domain V_i. The notations are the same as those in Proposition 5, but we assume the data tables are POUDTs. Then, we can associate two possibility distributions on V_i with any $f_i(x) \subseteq V_i$.

The first distribution, denoted by $\pi_\alpha(x)$, corresponds to the characteristic (or membership) function of $f_i(x)$, i.e., $\pi_\alpha(x)(v) = 1$ if $v \in f_i(x)$, and 0 otherwise. Thus, it is a $\{0, 1\}$-valued possibility distribution. Furthermore, $\pi_\alpha(x)$ is normalized if $f_i(x)$ is not empty. If we replace f_i with π_α in Eq. (23) and take the t-norm operation as the min, then $D_i(x, y)$ is reduced to the PWP relation. Thus, Proposition 5 holds if we regard the approximations in terms of PWP relations. However, this is not the same as our treatment of a POUDT based on the valued preference relation defined in (16).

The definition of the valued preference relation in (16) can be seen as a variant of that in (23) if we consider the second possibility distribution associated with $f_i(x)$, denoted by $\pi_\beta(x)$, which corresponds to the uniform probability function on $f_i(x)$, i.e., $\pi_\beta(x)(v) = \frac{1}{|f_i(x)|}$ if $v \in f_i(x)$, and 0 otherwise. Generally, it is not normalized in the sense of possibility distributions. If we take the t-norm operation in Eq. (23) as $\cdot$ and replace sup and f_i with $+$ and π_β, respectively, then $D_i(x, y)$ is reduced to the relation defined in (16). Thus, Proposition 5 only holds for our treatment of a POUDT if $\pi_\beta(x)(v) \leq \pi'_\beta(x)(v)$ for all $x \in U$ and $v \in V_i$. However, since $\pi_\beta(x)$ and $\pi'_\beta(x)$ are both probability distributions on V_i, the condition is only satisfied when $\pi_\beta(x) = \pi'_\beta(x)$. Therefore, Proposition 5 does not hold for our treatment of POUDTs, except for the trivial case of $T = T'$.

6.6. Complexity analysis

As shown by the above comparison, the main advantage of our work is that it can deal with very general types of uncertain data. However, the generality incurs an additional computational overhead. Without considering possible optimizations, we briefly compare the computational complexity of a naive implementation of our approaches with that of the standard DRSA.

The main difference between DFRSA and DRSA is that the preference relations in DRSA are primitive notions, whereas we have to compute the degrees of preference $D_i(x, y)$ from the primitive relations based on (16) and (23). The computation can be performed in $O(|V_i|^2)$ time for the criterion i if V_i is finite. For infinite V_i, we can sometimes compute $D_i(x, y)$ in constant time, as in Example 1; however, in other cases, we may also need to employ symbolic computation techniques to compute $D_i(x, y)$. In every case, it seems that the computation of $D_i(x, y)$ is quite time-consuming compared to the one-step evaluation of $x \succeq_i y$ in DRSA. Since $D_i(x, y)$ is an essential component of the valued dominance relation D_P and it is used frequently, it seems that the computation of $D_i(x, y)$ would slow down our approaches significantly. However, we can pre-compute $D_i(x, y)$ for all $x, y \in U$ and $i \in A$ and store them in $O(|A| \cdot |U|^2)$ space. In the same way, we can pre-compute the membership functions of $Cl_t^{\geq}$ and $Cl_t^{\leq}$ for each $t \in V_d$; and their storage requires $O(|U| \cdot |V_d|)$ space. It is reasonable to assume that $|V_d| \ll |U|$; thus, the total space requirement should be $O(|A| \cdot |U|^2)$. Note that, in the original DRSA, we also need $O(|A| \cdot |U|^2)$ bits of space to store the preference relations if we use a bit-array to store their incidence matrix. Since $D_i(x, y) \in [0, 1]$, we can use floating point numbers to represent the degrees of preference, which require a constant number of bits (e.g., 32 or 64). Consequently, the space requirement of our approaches is in the same order as that of DRSA.

After the degrees of preference have been pre-computed, the time complexity of the remaining steps in our approach is in the same order as that of the standard DRSA. The steps include the aggregation of valued preferences into valued dominance relations and the computation of reducts and lower and upper approximations. The main difference is that the Boolean logic operations in DRSA are replaced by more complicated t-norm, s-norm, or implication operations in DFRSA.

Finally, we compare the sizes of the rule sets and the time required for the decision-making process based on our approaches and DRSA. In DRSA, each object in each lower approximation can induce a decision rule, so the size of the exhaustive rule set is $\sum_{t \in V_d}(|\underline{P}(Cl_t^{\geq})| + |\underline{P}(Cl_t^{\leq})|)$. Consequently, in *the worst case*, the decision-making process may require the same number of steps to determine which rule is applicable. On the other hand, the size of the rule set induced by our approaches is $\sum_{t \in V_d}(|\underline{P}(Cl_t^{\geq})_c| + |\underline{P}(Cl_t^{\leq})_c|)$ if $c > 0$ is our threshold for rule generation, where S_c denotes the c-cut of a fuzzy subset S. The size seems to be in the same order as that of DRSA. However, our rule application process must apply the whole set of rules *in all cases*. Of course, we can also implement a bound-and-cut technique to optimize the process. Furthermore, the computation of the degree of satisfaction of a rule with respect to a given case would take the same time as the computation of $D_i(x, y)$; however, this process is more time-consuming than simply testing if a rule is applicable in standard DRSA.

7. Conclusion

This paper extends DRSA to a dominance-based fuzzy rough set approach (DFRSA), which can be applied to the reduction of criteria and the induction of rules for decision analysis in a POUDT or POPDT. In contrast to other approaches that deal with imprecise evaluations and assignments, DFRSA induces quantitative rules instead of qualitative rules. The proposed approach may also be useful for sparse data sets, so we will explore possible applications of DFRSA to such data sets in a future work [30].[4]

Since DFRSA is a general framework, we do not specify the t-norm operations used in the aggregation of consistency degrees or the implication operations used in the definition of adherence to the dominance principle. Hence, we do not present detailed algorithms for the computation of reducts. The computational aspects of DFRSA for specialized t-norm and implication operations will also be addressed in a future work.

Accommodating uncertainty in data incurs an additional computational overhead. In particular, the exhaustive application of rules seems quite time-consuming. A number of simplifications of the decision-making process have been proposed for DRSA [2,3]. In a future work, we will investigate if these techniques can also be adapted to our framework.

Finally, the computational complexity of the naive implementation of DFRSA prevents us applying it to any real data at this stage. Consequently, no empirical study or statistical analysis is included in the work. While the work would definitely benefit from empirical validation, we reiterate that the focus of the work is to clarify the interpretations of uncertainty in data and develop a framework based on the given interpretation. Our purpose is to determine *what* rules can be reasonably induced from the data, instead of *how* they can be induced efficiently. Thus, the framework is primarily declarative instead of procedural. However, although large-scale empirical validation of the proposed framework is still lacking, we have shown how the induced models can be used to rank the possible assignments of new decision cases. Furthermore, the complexity of the naive implementation does not exclude the possibility of more efficient implementations of the approach. The performance improvement and empirical analysis of the framework will be addressed in a future work.

Acknowledgments

This work was partially supported by the National Science Council of Taiwan under Grants NSC 98-2221-E-001-013-MY3 and NSC 99-2410-H-346-001. We wish to thank the guest editor and the anonymous referees for their constructive suggestions.

References

[1] K. Dembczynski, S. Greco, R. Slowinski, Rough set approach to multiple criteria classification with imprecise evaluations and assignments, European Journal of Operational Research 198 (2) (2009) 626–636.
[2] K. Dembczynski, R. Pindur, R. Susmaga, Dominance-based rough set classifier without induction of decision rules, Electronic Notes in Theoretical Computer Science 82 (4) (2003) 84–95.
[3] K. Dembczynski, R. Pindur, R. Susmaga, Generation of exhaustive set of rules within dominance-based rough set approach, Electronic Notes in Theoretical Computer Science 82 (4) (2003) 96–107.
[4] D. Dubois, H. Prade, Rough fuzzy sets and fuzzy rough sets, International Journal of General Systems 17 (1990) 191–209.

[4] We would like to thank an anonymous referee at IFSA/EUSFLAT 2009 for pointing out this possibility.

[5] D. Dubois, H. Prade, Formal representations of uncertainty, in: D. Bouyssou, D. Dubois, M. Pirlot, H. Prade (Eds.), Decision-Making Process: Concepts and Methods, ISTE London & Wiley, 2009, pp. 85–156.
[6] D. Dubois, H. Prade, P. Smets, Responses to Elkan (Didier Dubois, Henri Prade, Philippe Smets), IEEE Intelligent Systems 9 (4) (1994) 15–19.
[7] D. Dubois, H. Prade, P. Smets, Representing partial ignorance, IEEE Transactions on Systems, Man and Cybernetics—Part A: Systems and Humans 26 (3) (1996) 361–377.
[8] T. Fan, D. Liu, G. Tzeng, Rough set-based logics for multicriteria decision analysis, European Journal of Operational Research 182 (1) (2007) 340–355.
[9] S. Greco, B. Matarazzo, R. Slowinski, Rough approximation of a preference relation by dominance relations, European Journal of Operational Research 117 (1) (1999) 63–83.
[10] S. Greco, B. Matarazzo, R. Slowinski, The use of rough sets and fuzzy sets in mcdm, in: T. Gal, T. Stewart, T. Hanne (Eds.), Advances in Multiple Criteria Decision Making, Kluwer Academic Publishers, 1999b, pp. 14.1–14.9.
[11] S. Greco, B. Matarazzo, R. Slowinski, Dealing with missing data in rough set analysis of multi-attribute and multi-criteria decision problems, in: S. Zanakis, G.D.C. Zopounidis (Eds.), Decision Making: Recent Developments and Worldwide Applications, Kluwer Academic Publishers, 2000a, pp. 295–316.
[12] S. Greco, B. Matarazzo, R. Slowinski, Extension of the rough set approach to multicriteria decision support, INFOR Journal: Information Systems and Operational Research 38 (3) (2000) 161–195.
[13] S. Greco, B. Matarazzo, R. Slowinski, Rough set theory for multicriteria decision analysis, European Journal of Operational Research 129 (1) (2001) 1–47.
[14] S. Greco, B. Matarazzo, R. Slowinski, Rough sets methodology for sorting problems in presence of multiple attributes and criteria, European Journal of Operational Research 138 (2) (2002) 247–259.
[15] S. Greco, B. Matarazzo, R. Slowinski, Axiomatic characterization of a general utility function and its particular cases in terms of conjoint measurement and rough-set decision rules, European Journal of Operational Research 158 (2) (2004) 271–292.
[16] S. Greco, B. Matarazzo, R. Slowinski, Fuzzy set extensions of the dominance-based rough set approach, in: H. Bustince, F. Herrera, J. Montero (Eds.), Fuzzy Sets and Their Extensions: Representation, Aggregation and Models, Springer-Verlag, 2008, pp. 239–261.
[17] Q. Hu, D. Yu, M. Guo, Fuzzy preference based rough sets, Information Sciences 180 (10) (2010) 2003–2022.
[18] M. Inuiguchi, Y. Yoshioka, Y. Kusunoki, Variable-precision dominance-based rough set approach and attribute reduction, International Journal of Approximate Reasoning 50 (8) (2009) 1199–1214.
[19] E. Jaynes, Probability Theory: The Logic of Science, Cambridge University Press, 2003.
[20] M. Kryszkiewicz, Properties of incomplete information systems in the framework of rough sets, in: L. Polkowski, A. Skowron (Eds.), Rough Sets in Knowledge Discovery, Physica-Verlag, 1998, pp. 422–450.
[21] M. Kryszkiewicz, H. Rybiński, Reducing information systems with uncertain attributes, in: Z.W. Raś, M. Michalewicz (Eds.), Proceedings of the 9th ISMIS, LNAI, vol. 1079, Springer-Verlag, 1996a, pp. 285–294.
[22] M. Kryszkiewicz, H. Rybiński, Reducing information systems with uncertain real value attributes, in: Proceedings of the 6th International Conference on Information Processing and Management of Uncertainty in Knowledge-based Systems, 1996, pp. 1165–1169.
[23] T. Li, Y. Leung, W. Zhang, Generalized fuzzy rough approximation operators based on fuzzy coverings, International Journal of Approximate Reasoning 48 (3) (2008) 836–856.
[24] W. Lipski, On databases with incomplete information, Journal of the ACM 28 (1) (1981) 41–70.
[25] G. Liu, Axiomatic systems for rough sets and fuzzy rough sets, International Journal of Approximate Reasoning 48 (3) (2008) 857–867.
[26] A. Mieszkowicz-Rolka, L. Rolka, Fuzzy rough approximations of process data, International Journal of Approximate Reasoning 49 (2) (2008) 301–315.
[27] Z. Pawlak, Rough sets, International Journal of Computer and Information Sciences 11 (15) (1982) 341–356.
[28] Z. Pawlak, Rough Sets–Theoretical Aspects of Reasoning about Data, Kluwer Academic Publishers, 1991.
[29] A. Radzikowska, E. Kerre, A comparative study of fuzzy rough sets, Fuzzy Sets and Systems 126 (2) (2002) 137–155.
[30] D. Slezak, Rough sets and few-objects-many-attributes problem: the case study of analysis of gene expression data sets, in: Proceedings of the International Conference on Frontiers in the Convergence of Bioscience and Information Technologies, IEEE Press (2007) 437–442.
[31] R. Slowinski, S. Greco, B. Matarazzo, Rough set analysis of preference-ordered data, in: J. Alpigini, J. Peters, A. Skowron, N. Zhong (Eds.), Proceedings of the 3rd International Conference on Rough Sets and Current Trends in Computing, LNAI, vol. 2475, Springer-Verlag, 2002, pp. 44–59.
[32] L. Sweeney, Achieving *k*-anonymity privacy protection using generalization and suppression, International Journal on Uncertainty, Fuzziness and Knowledge-based Systems 10 (5) (2002) 571–588.
[33] T. Yang, Q. Li, Reduction about approximation spaces of covering generalized rough sets, International Journal of Approximate Reasoning 51 (3) (2010) 335–345.
[34] Y. Yao, Q. Liu, A generalized decision logic in interval-set-valued information tables, in: N. Zhong, A. Skowron, S. Ohsuga (Eds.), New Directions in Rough Sets, Data Mining, and Granular-Soft Computing, LNAI, vol. 1711, Springer-Verlag, 1999, pp. 285–293.
[35] L. Zadeh, The concept of a linguistic variable and its applications in approximate reasoning, Information Sciences 8 (1975) 199–251.
[36] L. Zadeh, Fuzzy sets as a basis for a theory of possibility, Fuzzy Sets and Systems 1 (1) (1978) 3–28.