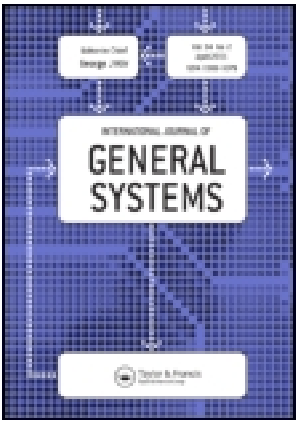

International Journal of General Systems
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/ggen20

Variable consistency and variable precision models for dominance-based fuzzy rough set analysis of possibilistic information systems

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 , Taiwan
[b] Institute of Information Science , Academia Sinica , Taipei , Taiwan
[c] Institute of Information Management , National Chiao-Tung University , Hsinchu , Taiwan
Published online: 20 May 2013.

To cite this article: Tuan-Fang Fan , Churn-Jung Liau & Duen-Ren Liu (2013) Variable consistency and variable precision models for dominance-based fuzzy rough set analysis of possibilistic information systems, International Journal of General Systems, 42:6, 659-686, DOI: 10.1080/03081079.2013.798910

To link to this article: http://dx.doi.org/10.1080/03081079.2013.798910

International Journal of General Systems, 2013
Vol. 42, No. 6, 659–686, http://dx.doi.org/10.1080/03081079.2013.798910

Variable consistency and variable precision models for dominance-based fuzzy rough set analysis of possibilistic information systems

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

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

(Received 30 September 2011; final version received 18 February 2013)

The dominance-based fuzzy rough set approach (DFRSA) is a theoretical framework that can deal with multi-criteria decision analysis of possibilistic information systems. While a set of comprehensive decision rules can be induced from a possibilistic information system by using DFRSA, generation of several intuitively justified rules is sometimes blocked by objects that only partially satisfy the antecedents of the rules. In this paper, we use the variable consistency models and variable precision models of DFRSA to cope with the problem. The models admit rules that are not satisfied by all objects. It is only required that the proportion of objects satisfying the rules must be above a threshold called a consistency level or a precision level. In the presented models, the proportion of objects is represented as a relative cardinality of a fuzzy set with respect to another fuzzy set. We investigate three types of models based on different definitions of fuzzy cardinalities including Σ-counts, possibilistic cardinalities, and probabilistic cardinalities; and the consistency levels or precision levels corresponding to the three types of models are, respectively, scalars, fuzzy numbers, and random variables.

Keywords: dominance-based fuzzy rough set approach; preference-ordered possibilistic information system; multi-criteria decision analysis; variable consistency DFRSA; variable precision DFRSA; fuzzy cardinality

1. Introduction

The rough set theory proposed by Pawlak (1982) provides an effective tool for extracting knowledge from information systems. 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 (Greco et al. 2001, 2002, 2004; Słowiński et al. 2002). The original rough set theory cannot handle inconsistencies arising from violations of the dominance principle due to its use of the indiscernibility relation. Therefore, the indiscernibility relation is replaced by a dominance relation to solve the multi-criteria sorting problem; and the information system 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.

*Corresponding author. Email: liaucj@iis.sinica.edu.tw
This is an extended version of (Fan et al. 2011b).

A strong assumption about information systems 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 information systems have been introduced to represent incomplete information (Kryszkiewicz 1998; Lipski 1981; Myszkorowski 2011; Sakai et al. 2011; Yao and Liu 1999). DRSA has also been extended to deal with missing or uncertain values in MCDA problems (Greco et al. 2001; Fan et al. 2009; Inuiguchi 2009; Słowiński et al. 2002). Since an information system with missing or uncertain values is a special case of a possibilistic information system, further extension of DRSA to the decision analysis of possibilistic information systems is desirable. In Fan (2011a), we propose such an extension based on the fuzzy dominance principle. In the proposed approach, we first compute the degree of dominance between any two objects based on their possibilistic evaluations with respect to each criterion. This leads to a fuzzy 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 possibilistic information systems. In addition, the upward and downward unions of decision classes are fuzzy subsets of the universe. Therefore, the lower and upper approximations of the decision classes based on the fuzzy 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.

As in the case of DRSA, the strict requirement of dominance principle may prevent some useful decision rules from being discovered. The situation may deteriorate in the DFRSA (dominance-based fuzzy rough set approach) framework because objects partially satisfying the antecedent of a rule may block the derivation of the rule if they do not satisfy the rule's consequent at the same time. Thus, an appropriate relaxation of the dominance principle is necessary to alleviate the problem. This is achieved by allowing variable consistency or variable precision rules in DRSA (Greco et al. 2000; Inuiguchi et al. 2009). These rules can be induced even though not satisfied by all objects. It is only required that the proportion of objects satisfying the rules must be above a threshold called a consistency level or a precision level. In this paper, we use fuzzy cardinalities to develop the variable consistency models and variable precision models in the DFRSA framework. The proportion of objects satisfying a rule is modeled as a relative fuzzy cardinality in our approach. Because a fuzzy cardinality may be a scalar, a fuzzy number, or a random variable, we can induce three types of rules for each model depending on what kinds of fuzzy cardinalities are taken as the consistency level or precision level. Furthermore, we also use examples to illustrate why an intuitively justified rule cannot be discovered in DFRSA and how the proposed approach can solve the problem.

The remainder of the paper is organized as follows. In Section 2, we review the DRSA. In Section 3, we present the extension of DRSA for decision analysis of possibilistic information systems. The extended approach is called the DFRSA. In Section 4, we introduce the variable consistency and variable precision models of DFRSA. Section 5 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 defined as $\underline{R}X = \{x \in U \mid [x]_R \subseteq X\}$ and $\overline{R}X = \{x \in U \mid [x]_R \cap X \neq \emptyset\}$, respectively.

Although an approximation space is an abstract framework used to represent classification knowledge, it can easily be derived from a concrete information system. Pawlak (1991) defined an information system[1] 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, the *indiscernibility relation* with respect to B is defined as $ind(B) = \{(x, y) \mid x, y \in U, f_i(x) = f_i(y) \forall i \in B\}$. Obviously, for each $B \subseteq A$, $(U, ind(B))$ is an approximation space.

For MCDA problems, each object in an information system 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 information system (POIS) 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 information system; 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 values with respect to the criterion i (Słowiński et al. 2002). The weak preference relation $\succeq_i$ is supposed to be a complete preorder, i.e. a complete, reflexive, and transitive relation. In addition, we assume that the domain of the decision attribute is a finite set $V_d = \{1, 2, \ldots, n\}$ such that r is strictly preferred to s if $r > s$ for any $r, s \in V_d$.

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:

$$(x, y) \in D_P \Leftrightarrow f_i(x) \succeq_i f_i(y) \quad \forall i \in P. \tag{1}$$

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 $D_P^+(x) = \{y \in U \mid yD_Px\}$ 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. DFRSA for possibilistic information systems

3.1. *Preference-ordered possibilistic information systems*

A general approach used to specify the uncertainty of information exploits possibility distributions. 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. The notion of possibility distributions should not be confused with that of probability distributions. A major difference is in the axiom about the union of events. While probability calculus satisfies the additive axiom, the possibility measure of a set is equal to the supremum of the possibility degrees of its elements (Dubois and Prade 2001). 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 preference-ordered possibilistic information system (POPIS) 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 $i \in A$, $f_i : U \to (V_i \to [0, 1])^+$.

3.2. Fuzzy dominance relation

In a POPIS, the objects may have imprecise evaluations with respect to the condition criteria and imprecise assignments to decision classes. Thus, the dominance relation between objects cannot be determined with certainty. Instead, since possibility information for each value is available in a POPIS, we can use the extension principle in fuzzy set theory to compute the degree of dominance (Zadeh 1975). The extension principle extends an operation or a relation on a base domain to the class of all fuzzy sets or possibility distributions on the domain. In our context, we use the extension principle to extend the preference relation $\succeq_i$ on V_i to a fuzzy preference relation between two possibility distributions on V_i. Consequently, the dominance relation between two objects with respect to the criterion i is determined by their respective possibility distributions on the domain of the criterion. Let $\otimes$, $\oplus$, and $\to$ denote, respectively, a t-norm operation, an s-norm operation, and an implication operation[2] on [0, 1]. Then, the dominance relation with respect to the criterion i is a fuzzy relation $D_i : U \times U \to [0, 1]$ such that

$$D_i(x, y) = \sup_{v_1, v_2 \in V_i} \{f_i(x)(v_1) \otimes f_i(y)(v_2) \mid v_1 \succeq_i v_2\}. \quad (2)$$

After deriving the fuzzy dominance relation for each criterion, we can aggregate all the relations into P-dominance relations for any subset of criteria P. Thus, the fuzzy P dominance relation $D_P : U \times U \to [0, 1]$ is defined as

$$D_P(x, y) = \bigotimes_{i \in P} D_i(x, y). \quad (3)$$

Since the dominance relation is a fuzzy 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

$$\delta_P(x, y) = D_P(x, y) \to D_d(x, y), \quad (4)$$

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

$$\delta_P(x) = \bigotimes_{y \in U} (\delta_P(x, y) \otimes \delta_P(y, x)). \quad (5)$$

Intuitively, (4) is a fuzzy logic counterpart of the dominance principle, which states that if x P-dominates y (i.e. x is at least as good as y with respect to all criteria in P), then x should be assigned to a decision class at least as good as the class assigned to y. In DRSA, an object x is consistent if for all other objects y, (x, y) and (y, x) both satisfy the dominance principle. Thus, (5) defines the degree of consistency of x in the fuzzy logic sense.

Let T be a POPIS. Then, the *quality of the classification* of T based on the set of criteria P is defined as

$$\gamma_P(T) = \frac{\sum_{x \in U} \delta_P(x)}{|U|}. \quad (6)$$

Note that $\gamma_P(T)$ is monotonic with respect to P, i.e. $\gamma_Q(T) \leq \gamma_P(T)$ if $Q \subseteq P$. Thus, we can define every minimal subset $P \subseteq C$ such that $\gamma_P(T) = \gamma_C(T)$ 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 overcome this difficulty, we can adopt the following alternative definition of the quality of the classification:

$$\eta_P(T) = \frac{\sum_{x,y \in U} \delta_P(x, y)}{|U|^2}. \tag{7}$$

The reducts can also be defined in terms of this kind of definition.

In addition, the rough set-based reduct has been extended to the notion of approximate reduct in Slezak (2002) and Nguyen and Slezak (1999) by using information entropy measure. While we use the standard notion in rough set theory to define reducts, it is also interesting to investigate the effect of approximate reduct in the DFRSA framework. Although we cannot go into much detail here, this will be further studied in the future work.

3.3. Dominance-based fuzzy rough approximations

In a POPIS, the assignment of a decision label to an object may be imprecise, so the decision classes may be fuzzy subsets of the universe. Their membership functions are derived from the possibility distributions associated with the assignments of the objects. Specifically, for each $t \in V_d$, the decision class $Cl_t : U \to [0, 1]$ is defined by

$$Cl_t(x) = f_d(x)(t). \tag{8}$$

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

$$Cl_t^{\geq}(x) = \sup_{v \geq t} f_d(x)(v) = \Pi_x(\{v \geq t\}) \tag{9}$$

and

$$Cl_t^{\leq}(x) = \sup_{v \leq t} f_d(x)(v) = \Pi_x(\{v \leq t\}), \tag{10}$$

respectively, where Π_x is the possibility measure corresponding to the possibility distribution $f_d(x)$. Finally, since our dominance relation is a fuzzy 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 (Dubois and Prade 1990; Radzikowska and Kerre 2002). More specifically, the P-lower and P-upper approximations of $Cl_t^{\geq}$ and $Cl_t^{\leq}$ for each $t \in V_d$ are defined as fuzzy subsets of U with the following membership functions:

$$\underline{P}(Cl_t^{\geq})(x) = \bigotimes_{y \in U}(D_P(y, x) \to Cl_t^{\geq}(y)), \tag{11}$$

$$\overline{P}(Cl_t^{\geq})(x) = \bigoplus_{y \in U}(D_P(x, y) \otimes Cl_t^{\geq}(y)), \tag{12}$$

$$\underline{P}(Cl_t^{\leq})(x) = \bigotimes_{y \in U}(D_P(x, y) \to Cl_t^{\leq}(y)), \tag{13}$$

$$\overline{P}(Cl_t^{\leq})(x) = \bigoplus_{y \in U}(D_P(y, x) \otimes Cl_t^{\leq}(y)). \tag{14}$$

These equations are the fuzzy version of the corresponding definitions in DRSA. For example, (11) is a fuzzification of the condition that x belongs to the P-lower approximation of $Cl_t^{\geq}$ if all objects P-dominating x belong to $Cl_t^{\geq}$.

3.4. Decision rules

To represent knowledge discovered from a POPIS, we consider a preference-ordered possibilistic decision logic (POPDL). The well-formed formulas (wff) of POPDL are Boolean combinations of atomic formulas of the form $(\geq_i, \pi_i)$ or $(\leq_i, \pi_i)$, where $i \in A$ and $\pi_i \in (V_i \to [0,1])^+$. When π_i is a singleton possibility distribution such that $\pi(x) = 1$ if $x = v$ and $\pi(x) = 0$ if $x \neq v$, we abbreviate $(\geq_i, \pi_i)$ (resp. $(\leq_i, \pi_i)$) as $(\geq_i, v)$ (resp. $(\leq_i, v)$).

Let P denote a reduct of a POPIS and let $t \in V_d$. Then, for each object x, where $\underline{P}(Cl_t^{\geq})(x) > 0$ (or above some pre-determined threshold), we can derive the $D_{\geq}$-decision rule:

$$\bigwedge_{i \in P} (\geq_i, f_i(x)) \longrightarrow_{\underline{P}(Cl_t^{\geq})(x)} (\geq_d, t); \tag{15}$$

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

$$\bigwedge_{i \in P} (\leq_i, f_i(x)) \longrightarrow_{\underline{P}(Cl_t^{\leq})(x)} (\leq_d, t), \tag{16}$$

where $\underline{P}(Cl_t^{\geq})(x)$ and $\underline{P}(Cl_t^{\leq})(x)$ are the respective confidences of the rules. Note that the symbol "$\longrightarrow$" is purely syntactic and is simply used to connect the antecedent and the consequent of a decision rule. It should be not confused with the fuzzy implication $\to$ introduced in Section 3.2 because the latter is a binary operation on [0, 1].

Now, for a new decision case with evaluations based on the condition criteria P, we can apply these two types of rules 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 $\bigwedge_{i \in P}(\geq_i, \pi_i) \longrightarrow_c (\geq_d, t)$ discovered by the proposed approach. Then, according to the rule α, we can derive that the degree of satisfaction of $f_d(x) \succeq_d t$ is $\varepsilon(\alpha, f_d(x) \succeq_d t) = c \otimes \bigotimes_{i \in P} \sup_{v_1 \succeq_i v_2}(f_i(x)(v_1) \otimes \pi_i(v_2))$. Let $\mathcal{R}_t^{\geq}$ denote the set of all rules with a consequent $(\geq_d, t^{prime})$ such that $t^{prime} \geq t$. Then, the final degree of $f_d(x) \succeq_d t$ is $\bigoplus_{\alpha \in \mathcal{R}_t^{\geq}} \varepsilon(\alpha, f_d(x) \succeq_d t^{prime})$. We can derive the degree of $f_d(x) \preceq_d t$ from the second type of rule in a similar manner.

Mathematically, the evaluations and assignments in a POPIS are possibility distributions, so the atomic formulas of POPDL 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 interpretation 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 POPIS are restricted to the (usually finite) set of linguistic labels, so the set of atomic formulas in our POPDL only contains $(\geq_i, \pi_i)$ or $(\leq_i, \pi_i)$, where π_i corresponds to a linguistic label. For example, if the evaluated criterion is "score" and its domain is [0,100], then the set of linguistic labels may be {poor, fair, good, and excellent}, and their corresponding interpretations are possibility distributions on the domain.

3.5. An illustrative example

A variety of theoretical properties about the DFRSA framework have been proved in Fan (2011a). Instead of repeating the theoretical presentation, we will use an example to illustrate the procedures. The information system in this example contains guests' evaluations of hotels. Each evaluation consists of three criteria: "service"(f_1), "rooms"(f_2), and "location"(f_3) of a hotel; and the decision attribute is a guest's overall recommendation(f_d) on the hotel. The evaluations are given in terms of linguistic labels: "terrible"(L_1), "poor"(L_2), "average"(L_3), "good"(L_4), and "excellent"(L_5). Since the guests' evaluations are always subjective and inevitably imprecise, it is quite natural to interpret the linguistic labels as possibility distributions on the underlying domain [0, 10]. The possibility distributions corresponding to the five linguistic labels are shown in Figure 1. Formally, the POPIS in this example is $T = (U, A, \{(V_i, \succeq_i) \mid i \in A\}, \{f_i \mid i \in A\})$, where $U = \{x_1, \ldots, x_8\}$, $A = \{1, 2, 3, d\}$, $V_i = [0, 10]$, and $\succeq_i$ is defined for each $i \in A$ such that $v \succeq_i v'$ iff $v \geq v^{prime}$ for any $v, v^{prime} \in V_i$, and $f_i (i \in A)$ is specified in Table 1. For simplification of presentation, we will use f_is and L_is in parentheses to denote corresponding attributes and linguistic labels, respectively.

Let us assume the t-norm $\otimes = \min$ and the corresponding implication $\rightarrow$ is the Gödel implication.[3] Then, the fuzzy dominance relations can be derived from (2) and are represented by

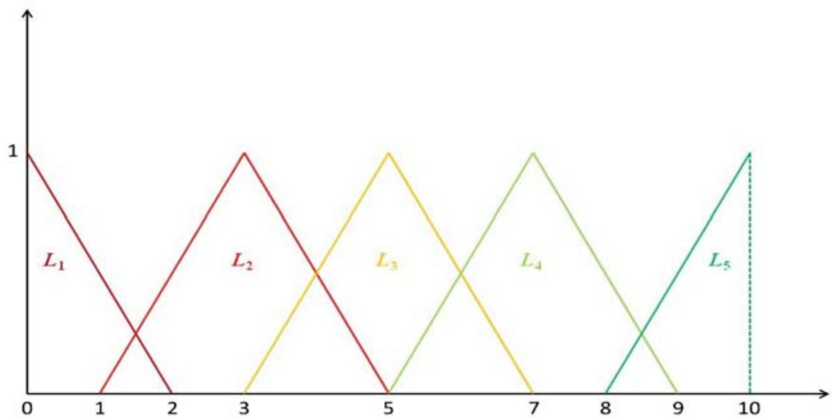

Figure 1. Possibility distributions corresponding to linguistic labels "terrible"(L_1), "poor"(L_2), "average"(L_3), "good"(L_4), and "excellent"(L_5).

Table 1. A POPIS of hotel evaluations.

U\A	Service (f_1)	Rooms (f_2)	Location (f_3)	Overall recommendation (f_d)
x_1	Good (L_4)	Excellent (L_5)	Excellent (L_5)	Excellent (L_5)
x_2	Average (L_3)	Good (L_4)	Excellent (L_5)	Good (L_4)
x_3	Average (L_3)	Excellent (L_5)	Good (L_4)	Good (L_4)
x_4	Average (L_3)	Good (L_4)	Average (L_3)	Average (L_3)
x_5	Good (L_4)	Excellent (L_5)	Good (L_4)	Good (L_4)
x_6	Average (L_3)	Poor (L_2)	Average (L_3)	Poor (L_2)
x_7	Poor (L_2)	Average (L_3)	Good (L_4)	Average (L_3)
x_8	Poor (L_2)	Poor (L_2)	Terrible (L_1)	Terrible (L_1)

Table 2. Fuzzy dominance relations for the POPIS.

	1	2	3	4	5	6	7	8
1	(1,1,1,1)	(1,1,1,1)	(1,1,1,1)	(1,1,1,1)	(1,1,1,1)	(1,1,1,1)	(1,1,1,1)	(1,1,1,1)
2	$\left(\frac{1}{2}, \frac{1}{4}, 1, \frac{1}{4}\right)$	(1,1,1,1)	$\left(1, \frac{1}{4}, 1, 1\right)$	(1,1,1,1)	$\left(\frac{1}{2}, \frac{1}{4}, 1, 1\right)$	(1,1,1,1)	(1,1,1,1)	(1,1,1,1)
3	$\left(\frac{1}{2}, 1, \frac{1}{4}, \frac{1}{4}\right)$	$\left(1, 1, \frac{1}{4}, 1\right)$	(1,1,1,1)	(1,1,1,1)	$\left(\frac{1}{2}, 1, 1, 1\right)$	(1,1,1,1)	(1,1,1,1)	(1,1,1,1)
4	$\left(\frac{1}{2}, \frac{1}{4}, 0, 0\right)$	$\left(1, 1, 0, \frac{1}{2}\right)$	$\left(1, \frac{1}{4}, \frac{1}{2}, \frac{1}{2}\right)$	(1,1,1,1)	$\left(\frac{1}{2}, \frac{1}{4}, \frac{1}{2}, \frac{1}{2}\right)$	(1,1,1,1)	$\left(1, 1, \frac{1}{2}, 1\right)$	(1,1,1,1)
5	$\left(1, 1, \frac{1}{4}, \frac{1}{4}\right)$	$\left(1, 1, \frac{1}{4}, 1\right)$	(1,1,1,1)	(1,1,1,1)	(1,1,1,1)	(1,1,1,1)	(1,1,1,1)	(1,1,1,1)
6	$\left(\frac{1}{2}, 0, 0, 0\right)$	(1,0,0,0)	$\left(1, 0, \frac{1}{2}, 0\right)$	$\left(1, 0, 1, \frac{1}{2}\right)$	$\left(\frac{1}{2}, 0, \frac{1}{2}, 0\right)$	(1,1,1,1)	$\left(1, \frac{1}{2}, \frac{1}{2}, \frac{1}{2}\right)$	(1,1,1,1)
7	$\left(0, 0, \frac{1}{4}, 0\right)$	$\left(\frac{1}{2}, \frac{1}{2}, \frac{1}{4}, \frac{1}{2}\right)$	$\left(\frac{1}{2}, 0, 1, \frac{1}{2}\right)$	$\left(\frac{1}{2}, \frac{1}{2}, 1, 1\right)$	$\left(0, 0, 1, \frac{1}{2}\right)$	$\left(\frac{1}{2}, 1, 1, 1\right)$	(1,1,1,1)	(1,1,1,1)
8	(0,0,0,0)	$\left(\frac{1}{2}, 0, 0, 0\right)$	$\left(\frac{1}{2}, 0, 0, 0\right)$	$\left(\frac{1}{2}, 0, 0, 0\right)$	(0,0,0,0)	$\left(\frac{1}{2}, 1, 0, \frac{1}{4}\right)$	$\left(1, \frac{1}{2}, 0, 0\right)$	(1,1,1,1)

a matrix in Table 2, where the entry (i, j) of the matrix denotes the tuple $(D_1(x_i, x_j), D_2(x_i, x_j), D_3(x_i, x_j), D_d(x_i, x_j))$ for $1 \leq i, j \leq 8$.

It is easy to check that $\min(c_1, c_2, c_3) \leq c_4$ for any tuple (c_1, c_2, c_3, c_4) in the matrix. This means that $\delta_P(x, y) = 1$ and $\delta_P(x) = 1$ for any $x, y \in U$ and $P = \{1, 2, 3\}$. Thus, the quality of classification of the POPIS is equal to 1, i.e. the POPIS is totally consistent. In the same way, we can check that $\delta_P(x, y) = 1$ and $\delta_P(x) = 1$ for any $x, y \in U$ and $P = \{2, 3\}$. Furthermore, when $P = \{1, 2\}$, we can find that $\delta_P(u_3, u_1) = \frac{1}{4}$, and when $P = \{1, 3\}$, we can find that $\delta_P(u_6, u_5) = 0$. Consequently, $\gamma_{\{1,2\}}(T) < \gamma_{\{1,2,3\}}(T)$ and $\gamma_{\{1,3\}}(T) < \gamma_{\{1,2,3\}}(T)$, but $\gamma_{\{2,3\}}(T) = \gamma_{\{1,2,3\}}(T)$. Thus, $\{2, 3\}$ is the only reduct of the POPIS.

To highlight the difference between POIS and POPIS, we can define a POIS $T^{prime} = (U, A, \{(V_i, \succeq_i) \mid i \in A\}, \{f_i^{prime} \mid i \in A\})$ such that $V_i = \{1, 2, 3, 4, 5\}$, $v \succeq_i v^{prime}$ iff $v \geq v^{prime}$, and $f_i^{prime}(x) = j$ iff $f_i(x) = L_j$ for any $i \in A$. In other words, T^{prime} is simply a variant of T in which the linguistic labels are treated as categorical values instead of possibility distribution. Thus, the dominance relations for T' are crisp and can be obtained by replacing all fractional numbers $\frac{1}{4}$ and $\frac{1}{2}$ in Table 2 with 0. Then, by using standard DRSA, the reducts of T' are $\{1, 3\}$ and $\{2, 3\}$. Hence, while criteria 1 and 2 seem equally important in T', the more fine-grained representation of POPIS T allows us to distinguish the relative importance between them.

Next, the upward and downward union of decision classes can be computed by using (9) and (10). For example, the membership function of $Cl_t^{\geq}$ for different ranges of t is shown in Table 3.

Furthermore, by using the reduct $P = \{2, 3\}$, we can derive $D_P(x, y) = \min(D_2(x, y), D_3(x, y))$ from Table 2. The results are recorded in Table 4.

Applying (11) and (12) to Tables 3 and 4, we can derive the lower and upper approximations of the upward union of decision classes $Cl_t^{\geq}$ for any $t \in [0, 10]$. In the same way, we can also derive the rough approximations of the downward unions. Then, we can induce rules based on the lower approximations. For example, for $t = 8$, we have $\underline{P}(Cl_t^{\geq})(x_1) = 1$, $\underline{P}(Cl_t^{\geq})(x_2) = \frac{1}{2}$, and $\underline{P}(Cl_t^{\geq})(x_i) = 0$ for $3 \leq i \leq 8$. From this, we can derive two rules as follows:

$$(\geq_2, L_5) \wedge (\geq_3, L_5) \longrightarrow_1 (\geq_d, 8),$$

$$(\geq_2, L_4) \wedge (\geq_3, L_5) \longrightarrow_{\frac{1}{2}} (\geq_d, 8),$$

Table 3. The upward union of decision classes.

$Cl_t^{\geq}$	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8
$0 \leq t \leq 3$	1	1	1	1	1	1	1	$\max\left(1-\frac{t}{2},0\right)$
$3 < t \leq 5$	1	1	1	1	1	$\frac{5-t}{2}$	1	0
$5 < t \leq 7$	1	1	1	$\frac{7-t}{2}$	1	0	$\frac{7-t}{2}$	0
$7 < t \leq 10$	1	$\max\left(\frac{9-t}{2},0\right)$	$\max\left(\frac{9-t}{2},0\right)$	0	$\max\left(\frac{9-t}{2},0\right)$	0	0	0

Table 4. The fuzzy dominance relation $D_P(x, y)$.

$D_P(x, y)$	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8
x_1	1	1	1	1	1	1	1	1
x_2	$\frac{1}{4}$	1	$\frac{1}{4}$	1	$\frac{1}{4}$	1	1	1
x_3	$\frac{1}{4}$	$\frac{1}{4}$	1	1	1	1	1	1
x_4	0	0	$\frac{1}{4}$	1	$\frac{1}{4}$	1	$\frac{1}{2}$	1
x_5	$\frac{1}{4}$	$\frac{1}{4}$	1	1	1	1	1	1
x_6	0	0	0	0	0	1	$\frac{1}{2}$	1
x_7	0	$\frac{1}{4}$	0	$\frac{1}{2}$	0	1	1	1
x_8	0	0	0	0	0	0	0	1

which mean "if the rooms and the location of a hotel are both excellent, then its overall score is at least 8" and "it is half certain that if the rooms of a hotel are at least good and its location is excellent, then its overall score is at least 8", respectively. Note that both rules belong to $\mathcal{R}_8^{\geq}$ according to the notation defined in Section 3.4. Thus, the procedure stated there can be used to derive the degree of $f_d(x) \succeq_d 8$ for a new object x. During the application of the procedure, both rules (possibly among other rules) can be applied to the new object at the same time.

4. Variable consistency and variable precision models of DRSA for possibilistic information systems

As the rough approximations defined in DRSA are based on consistency in terms of the dominance principle, it has been observed that some inconsistencies may reduce the cardinality of lower approximations to such an extent that it is impossible to discover strong patterns in the data, particularly when the data-set is large. Thus, Greco et al. (2000) proposed a relaxation of the strict dominance principle. The relaxation, which is controlled by a parameter called the consistency level, allows some inconsistent objects to be included in the lower approximations. The resulting model is called the *variable-consistency model* (VC-DRSA).

Inuiguchi et al. (2009) noted that inconsistencies in a decision table may occur for different reasons, such as: (1) hesitation in evaluating the decision attribute values; (2) errors in recording, measurement, and observation; (3) missing condition attributes related to the evaluation of decision attribute values; and (4) the unstable nature of the system represented by the decision table. The authors also showed that, although VC-DRSA can easily handle the inconsistency caused by hesitation in evaluating the attribute values, it cannot deal with other types of inconsistency. To alleviate the problem, they proposed a *variable precision model* (VP-DRSA).

Although VC-DRSA and VP-DRSA improves the robustness of DRSA, the users have to specify a consistency or precision level as the parameter for the induction of rules. Because the parameter is a scalar, the specification is usually arbitrary and a small perturbation on the specified parameter can cause abrupt change on the induced rules. Therefore, the flexibility for the imprecision-tolerance of the parameter is useful. The extension of the DFRSA framework can provide such a flexibility by allowing the consistency or precision level to be a fuzzy number or a random variable. Thus, in addition of the usual numerical level, the user can require that the rule is highly possible or probable without an explicit specification of the cutting point. In this section, we utilize the notion of fuzzy cardinality to extend the DFRSA framework along this direction.

4.1. Fuzzy cardinality

The cardinality of a fuzzy set is normally used to evaluate fuzzy quantified sentences. For example, to evaluate the truth degree of the sentence "Most students are young," we have to determine if the cardinality of the set of young students satisfies the interpretation of the fuzzy quantifier "most." In applications, two kinds of cardinality are considered: *absolute cardinality*, which measures the number of elements in a set, and *relative cardinality*, which measures the percentage of elements of one set (called the referential set) that are also present in another set (Delgado et al. 2002).[4] Since both VC-DRSA and VP-DRSA are concerned with the percentage of elements in an object's dominating set or dominated set that are also in the approximated concept, relative cardinality plays an important role in our analysis.

Several approaches for measuring the cardinality of a fuzzy set have been proposed in the literature. The approaches, which extend the classic approach in different ways, can be classified into two categories: *scalar cardinality* approaches and *fuzzy cardinality* approaches. The former measure the cardinality of a fuzzy set by means of a scalar value, either an integer or a real value; whereas the latter assume that the cardinality of a fuzzy set is just another fuzzy set over the nonnegative numbers (Delgado et al. 2002). The most simple scalar cardinality of a fuzzy set is its *power* (also called the Σ-count), which is defined as the summation of the membership degrees of all elements (de Luca and Termini, 1972). Formally, for a given fuzzy subset F on the universe U, the Σ-count of F is defined as

$$\Sigma\sharp(F) = \sum_{x \in U} \mu_F(x). \tag{17}$$

The relative cardinality of a fuzzy set G with respect to another fuzzy set F is then defined as (Zadeh 1975):

$$\Sigma\sharp(G/F) = \frac{\Sigma\sharp(F \cap G)}{\Sigma\sharp(F)}. \tag{18}$$

Subsequently, Zadeh (1979) proposed a fuzzy subset $Z(F)$ of $\mathbb{N}$ as the measure of the absolute cardinality of a fuzzy set F such that the membership degree of a natural number $k \in \mathbb{N}$ in $Z(F)$ is defined as

$$Z(F, k) = \sup\{\alpha \mid |F_\alpha| = k\}, \tag{19}$$

where F_α is the α-cut of F. In addition, a fuzzy multiset $Z(G/F)$ over [0, 1] is introduced in Zadeh (1983) to measure the fuzzy relative cardinality of G with respect to F. The membership function of $Z(G/F)$ is defined as

$$Z(G/F) = \sum_{\alpha \in \Lambda(F) \cup \Lambda(G)} \alpha / \frac{|F_\alpha \cap G_\alpha|}{|F_\alpha|}, \tag{20}$$

where $\Lambda(F)$ and $\Lambda(G)$ are the level sets of F and G, respectively, i.e. $\Lambda(F) = \{\mu_F(x) \mid x \in U\}$. Delgado et al. (2002) proposed a more compact representation of $Z(G/F)$ by transforming the fuzzy multiset into a fuzzy subset of rational numbers in [0, 1]. The representation is formulated as follows:

$$ES(G/F, q) = \sup\left\{\alpha \in \Lambda(G/F) \mid \frac{|(F \cap G)_\alpha|}{|F_\alpha|} = q\right\} \tag{21}$$

for any $q \in \mathbb{Q} \cap [0, 1]$, where $\Lambda(G/F) = \Lambda(F \cap G) \cup \Lambda(G)$.

In the context of a finite universe U, Delgado et al. (2002) proposed a family of fuzzy measures $\mathcal{E}$ for absolute cardinalities based on the evaluation of fuzzy logic sentences. To define the measures, the possibility of a fuzzy set F containing at least k elements is identified with the truth degree of the fuzzy statement "$\exists X \subseteq U$ such that $(|X| = k \wedge X \subseteq F)$," which can be formally defined as

$$L(F, k) = \begin{cases} 1, & \text{if } k = 0, \\ 0, & \text{if } k > |U|, \\ \bigoplus_{X \subseteq_k U} \bigotimes_{x \in X} \mu_{F(x)}, & \text{if } 1 \leq k \leq |U|, \end{cases} \tag{22}$$

where $X \subseteq_k U$ denotes that X is any k-element subset of U. Then, the possibility that F contains exactly k elements is formulated as follows:

$$E(F, k) = L(F, k) \otimes \neg L(F, k + 1), \tag{23}$$

where $\otimes$ is any t-norm (not necessarily the same as that used in the definition of $L(F, k)$), and $\neg$ stands for a fuzzy negation. Each member of the family $\mathcal{E}$ is determined by the choice of s-norm, t-norms, and negation in (22) and (23). Using max and min in (22) and standard negation as well as Lukasiewicz's t-norm $\max(0, a+b-1)$ in (23), a probabilistic measure of absolute cardinality ED defined as

$$ED(F, k) = \alpha_k - \alpha_{k+1} \tag{24}$$

is shown to be a member of the family $\mathcal{E}$, where α_k is the kth largest value of the multiset $\{\mu_F(x) \mid x \in U\}$ for $1 \leq k \leq |U|$, $\alpha_0 = 1$, and $\alpha_k = 0$ when $k > |U|$. The relative version of ED is also defined as

$$ER(G/F, q) = \sum_{i : \frac{|(F \cap G)_{\alpha_i}|}{|F_{\alpha_i}|} = q} (\alpha_i - \alpha_{i+1}) \tag{25}$$

for any $q \in \mathbb{Q} \cap [0, 1]$, where α_i is the ith largest value of $\Lambda(G/F)$.

The following proposition shows that all these definitions collapse to the standard one in the case of crisp sets.

Lemma 4.1 *Let F and G be crisp sets, $|F| = n$ and $\frac{|F \cap G|}{|F|} = q$. Then,*

1. *$\Sigma\sharp(F) = n$ and $\Sigma\sharp(G/F) = q$;*
2. *$Z(F, n) = 1$ and $Z(F, k) = 0$ if $k \neq n$;*
3. *$\mu_{Z(G/F)}(q) = 1$ and $\mu_{Z(G/F)}(v) = 0$ if $v \neq q$;*
4. *$ES(G/F, q) = 1$ and $ES(G/F, v) = 0$ if $v \neq q$;*
5. *$ED(F, n) = 1$ and $ED(F, k) = 0$ if $k \neq n$; and*
6. *$ER(G/F, q) = 1$ and $ER(G/F, v) = 0$ if $v \neq q$.*

Proof We note that, for crisp sets, the level sets $\Lambda(F)$, $\Lambda(F \cap G)$, and $\Lambda(G/F)$ are equal to {0, 1}. Furthermore, each crisp set is equal to its 1-cut. Then, the results follows from the definitions immediately. For example, for the proof of item (5), if $|F| = n$, then according to the definition of (24), $\alpha_k = 1$ if $0 \leq k \leq n$ and $\alpha_k = 0$ if $k > n$. Thus, we have $\alpha_k - \alpha_{k+1} = 0$ for $k \neq n$ and $\alpha_n - \alpha_{n+1} = 1$. □

4.2. Variable consistency models of DFRSA

The main difference between VC-DFRSA and DFRSA is the definition of the lower approximation, so we still use the fuzzy dominance relation and the notion of reduct defined in Section 3 to develop VC-DFRSA. The main feature of VC-DRSA is that it includes objects that partially violate the dominance principle. For example, if an object belongs to the lower approximation of an upward union of classes, it is not necessary for all objects that dominate the object to belong to the same upward union of classes. Instead, the only requirement for an object to be included in the lower approximation of a target set is that a sufficiently large portion of the object's dominating set belongs to the target set. In DFRSA, an object's dominating set and dominated set as well as the target sets are all fuzzy sets; therefore, we can determine if the portion is sufficiently large by comparing the sets' relative cardinalities with the consistency level. To formulate the VC-DFRSA model, we define the *P-dominating set* and *P-dominated set* of an object x as fuzzy subsets of U with the following membership functions:

(1) P-dominating set: $D_P^+(x)(y) = D_P(y, x)$ for every $y \in U$,
(2) P-dominated set: $D_P^-(x)(y) = D_P(x, y)$ for every $y \in U$.

Then, we can consider three kinds of generalizations of VC-DRSA to VC-DFRSA.

First, if a scalar consistency level $l \in (0.5, 1]$ is given, then we use the relative Σ-count to measure if an object satisfies the partial consistency requirement. Thus, the P-lower and P-upper approximations of $Cl_t^{\geq}$ and $Cl_t^{\leq}$ for each $t \in V_d$ are defined as fuzzy subsets of U with the following membership functions:

$$\underline{P}^l(Cl_t^{\geq})(x) = \begin{cases} Cl_t^{\geq}(x), & \text{if } \Sigma\sharp(Cl_t^{\geq}/D_P^+(x)) \geq l, \\ 0, & \text{otherwise,} \end{cases} \tag{26}$$

$$\underline{P}^l(Cl_t^{\leq})(x) = \begin{cases} Cl_t^{\leq}(x), & \text{if } \Sigma\sharp(Cl_t^{\leq}/D_P^-(x)) \geq l, \\ 0, & \text{otherwise,} \end{cases} \tag{27}$$

$$\overline{P}^l(Cl_t^{\geq})(x) = \begin{cases} Cl_t^{\geq}(x), & \text{if } \Sigma\sharp(Cl_t^{\geq}/D_P^-(x)) \leq 1-l, \\ Cl_t^{\geq}(x) \oplus Cl_t^{\leq}(x), & \text{otherwise,} \end{cases} \tag{28}$$

$$\overline{P}^l(Cl_t^{\leq})(x) = \begin{cases} Cl_t^{\leq}(x), & \text{if } \Sigma\sharp(Cl_t^{\leq}/D_P^+(x)) \leq 1-l, \\ Cl_t^{\geq}(x) \oplus Cl_t^{\leq}(x), & \text{otherwise.} \end{cases} \tag{29}$$

Second, if the consistency level is a fuzzy number over $(0.5, 1] \cap \mathbb{Q}$, then we use the relative cardinality ES to measure an object's degree of consistency. Hence, for a fuzzy number $\tilde{l} : (0.5, 1] \cap \mathbb{Q} \to [0, 1]$, the membership functions for the P-lower and P-upper approximations of $Cl_t^{\geq}$ and $Cl_t^{\leq}$ are defined as follows:

$$\underline{P}^{\tilde{l}}(Cl_t^{\geq})(x) = Cl_t^{\geq}(x) \otimes \pi(ES(Cl_t^{\geq}/D_P^+(x)) \geq \tilde{l}), \tag{30}$$

$$\underline{P}^{\tilde{l}}(Cl_t^{\leq})(x) = Cl_t^{\leq}(x) \otimes \pi(ES(Cl_t^{\leq}/D_P^-(x)) \geq \tilde{l}), \tag{31}$$

$$\overline{P}^{\tilde{l}}(Cl_t^{\geq})(x) = Cl_t^{\geq}(x) \oplus (Cl_t^{\leq}(x) \otimes \pi(ES(Cl_t^{\geq}/D_P^-(x)) > 1-\tilde{l})), \tag{32}$$

$$\overline{P}^{\tilde{l}}(Cl_t^{\leq})(x) = Cl_t^{\leq}(x) \oplus (Cl_t^{\geq}(x) \otimes \pi(ES(Cl_t^{\leq}/D_P^+(x)) > 1-\tilde{l})). \tag{33}$$

In the above definition, the relative cardinality ES is regarded as a fuzzy number, and $\pi(\cdot)$ returns the possibility of the comparison statement between two fuzzy numbers based on the extension principle. For example, the possibility of a fuzzy number $\tilde{l}_1$ being greater than another fuzzy number $\tilde{l}_2$ is defined as $\pi(\tilde{l}_1 > \tilde{l}_2) = \sup_{v_1 > v_2} \min(\mu_{\tilde{l}_1}(v_1), \mu_{\tilde{l}_2}(v_2))$, where v_1 and v_2 are ranged over the domain of the fuzzy numbers (i.e. $v_1, v_2 \in [0, 0.5) \cap \mathbb{Q}$). In addition, the membership function of the fuzzy number $1 - \tilde{l}$ is defined as $\mu_{1-\tilde{l}}(v) = \mu_{\tilde{l}}(1 - v)$ for any $v \in [0, 0.5) \cap \mathbb{Q}$.

Third, if the consistency level is a $(0.5, 1] \cap \mathbb{Q}$-valued random variable, we can use the relative cardinality ER to measure the probability of an object's consistency. Let us overload the notation $ER(G/F)$ to denote a $[0, 1] \cap \mathbb{Q}$-valued random variable whose probability mass function is defined as $Pr(ER(G/F) = q) = ER(G/F, q)$ for any fuzzy sets F and G, and let $\hat{l}$ be any $(0.5, 1] \cap \mathbb{Q}$-valued random variable. Then, the membership functions for the P-lower and P-upper approximations of $Cl_t^{\geq}$ and $Cl_t^{\leq}$ are defined as follows:

$$\underline{P}^{\hat{l}}(Cl_t^{\geq})(x) = Cl_t^{\geq}(x) \otimes Pr(ER(Cl_t^{\geq}/D_P^{+}(x)) \geq \hat{l}), \tag{34}$$

$$\underline{P}^{\hat{l}}(Cl_t^{\leq})(x) = Cl_t^{\leq}(x) \otimes Pr(ER(Cl_t^{\leq}/D_P^{-}(x)) \geq \hat{l}), \tag{35}$$

$$\overline{P}^{\hat{l}}(Cl_t^{\geq})(x) = Cl_t^{\geq}(x) \oplus (Cl_t^{\leq}(x) \otimes Pr(ER(Cl_t^{\geq}/D_P^{-}(x)) > 1 - \hat{l})), \tag{36}$$

$$\overline{P}^{\hat{l}}(Cl_t^{\leq})(x) = Cl_t^{\leq}(x) \oplus (Cl_t^{\geq}(x) \otimes Pr(ER(Cl_t^{\leq}/D_P^{+}(x)) > 1 - \hat{l})). \tag{37}$$

The three types of variable consistency models for DFRSA are called VC1-DFRSA, VC2-DFRSA, and VC3-DFRSA, respectively. As a scalar l can be regarded as a single-point (possibility or probability) distribution, all three types of models are applicable when the consistency level is a scalar. A typical application of VC2-DFRSA is when the consistency level is given by a linguistic term. For example, it may be required that the consistency level is moderately high. On the other hand, VC3-DFRSA may be applied when the consistency level is set as a sub-interval of $(0.5, 1] \cap \mathbb{Q}$. In this case, the consistency level is regarded as a uniform distribution on the sub-interval, so it is actually a random variable.

PROPOSITION 4.2

(1) *VC1-DFRSA is reduced to VC-DRSA for the analysis of POIS.*
(2) *VC1-DFRSA is a special case of both VC2-DFRSA and VC3-DFRSA for the analysis of POIS.*

Proof

(1) In the case of POIS, the sets $D_P^{+}(x)$ and $D_P^{-}(x)$ for any $x \in U$ and $Cl_t^{\geq}$ and $Cl_t^{\leq}$ for any $t \in V_d$ are all crisp sets. Thus, by Lemma 4.1, the relative Σ count of two crisp sets is simply the classical relative cardinality. Consequently, the definitions in (26), (27), (28), (29) reduce to those given in Greco et al. (2000).
(2) By Lemma 4.1, the values of ES in (30), (31), (32), (33) are single-point possibility distributions and the values of ED in (34), (35), (36), (37) are single-point probability distributions. Thus, when the consistency level of the VC2-DFRSA (resp. VC3-DFRSA) is the special case of a single-point possibility (resp. probability) distribution, definitions (30), (31), (32), (33) (resp. (34), (35), (36), (37)) are equivalent to (26)-(29), respectively. For example, if $\Sigma\sharp(Cl_t^{\geq}/D_P^{+}(x)) = q$, then $ES(Cl_t^{\geq}/D_P^{+}(x), q) = 1$ and $ES(Cl_t^{\geq}/D_P^{+}(x), x) = 0$ for $x \neq q$. Let the consistency level $\tilde{l}$ be defined such that $\mu_{\tilde{l}}(l) = 1$ and $\mu_{\tilde{l}}(v) = 0$ if $v \neq l$. Then, $\pi(ES(Cl_t^{\geq}/D_P^{+}(x)) \geq \tilde{l}) = 1$ iff $q \geq l$ and $\pi(ES(Cl_t^{\geq}/D_P^{+}(x)) \geq \tilde{l}) = 0$ otherwise. Thus, (30) is equivalent to (26). □

4.3. Variable precision models of DFRSA

The variable precision DRSA model can be understood from an evidence-based perspective. In VP-DRSA, each object can be regarded as a piece of evidence about whether a target object should be included in a given class. For example, when we consider whether an object x should be included in the upward union of classes $Cl_t^{\geq}$, each object could be

(1) a piece of positive evidence if it is P-dominated by x and belongs to $Cl_t^{\geq}$ at the same time;
(2) a piece of negative evidence if it P-dominates x, but belongs to $Cl_{t-1}^{\leq}$; or
(3) a piece of neutral evidence if it is neither positive nor negative.

Thus, an object supports the hypothesis $x \in Cl_t^{\geq}$ if it is a piece of positive evidence, and rejects the hypothesis if it is a piece of negative evidence. Furthermore, an object is irrelevant to the hypothesis if it is a piece of neutral evidence. Consequently, the degree of confirmation for the hypothesis is the ratio of supporting objects over relevant objects. The VP-DRSA model considers that x belongs to the lower approximation of $Cl_t^{\geq}$ if the degree of confirmation for the hypothesis $x \in Cl_t^{\geq}$ is not less than a precision level in (0.5, 1] (Inuiguchi et al. 2009).

As in the case of VC-DFRSA, we can consider three types of variable precision models for DFRSA, since the precision level may be a scalar, a fuzzy number, or a random variable in DFRSA. However, because the ratio of supporting objects over relevant objects in DFRSA is not directly equal to the relative cardinality between two fuzzy sets, we have to change the definitions of $\Sigma\sharp$, ES, and ER slightly, although we still use the same notation. For two fuzzy subsets F and G of the same universe, we define

$$\Sigma\sharp(G\dagger F) = \frac{\Sigma\sharp(G)}{\Sigma\sharp(G) + \Sigma\sharp(F)}; \tag{38}$$

$$ES(G\dagger F, q) = \sup\left\{\alpha \in \Lambda(G) \cup \Lambda(F) \mid \frac{|G_\alpha|}{|G_\alpha| + |F_\alpha|} = q\right\} \tag{39}$$

for any $q \in \mathbb{Q} \cap [0, 1]$; and

$$ER(G\dagger F, q) = \sum_{\frac{|G_{\alpha_i}|}{|F_{\alpha_i}|+|G_{\alpha_i}|}=q} (\alpha_i - \alpha_{i+1}) \tag{40}$$

for any $q \in \mathbb{Q} \cap [0, 1]$, where α_i is the ith largest value of $\Lambda(G) \cup \Lambda(F)$ (recall that $\Lambda(F)$ and $\Lambda(G)$ are the level sets of F and G, respectively, i.e. $\Lambda(F) = \{\mu_F(x) \mid x \in U\}$). Analogous to lemma 4.1, we have the following results for crisp sets.

LEMMA 4.3 *Let F and G be crisp sets and $\frac{|G|}{|F|+|G|} = q$. Then,*

(1) $\Sigma\sharp(G\dagger F) = q$;
(2) $ES(G\dagger F, q) = 1$ *and* $ES(G\dagger F, v) = 0$ *if* $v \neq q$;
(3) $ER(G\dagger F, q) = 1$ *and* $ER(G\dagger F, v) = 0$ *if* $v \neq q$.

To simplify the presentation, for each object $x \in U$ and each class label $t \in V_d$, we denote the fuzzy subsets of objects that support $x \in Cl_t^{\geq}$ and $x \in Cl_t^{\leq}$ by $S_x^{\geq t} = D_P^-(x) \cap Cl_t^{\geq}$ and $S_x^{\leq t} = D_P^+(x) \cap Cl_t^{\leq}$, respectively. Following the terminology of Inuiguchi et al. (2009), we call the lower and upper approximations of VP-DFRSA positive and nonnegative regions, respectively. Hence, when the precision level is a scalar $l \in (0.5, 1]$, we can define the P-positive and P-nonnegative regions of $Cl_t^{\geq}$ and $Cl_t^{\leq}$ as follows:

$$POS_P^l(Cl_t^{\geq}) = \left\{x \in U \mid \Sigma\sharp\left(S_x^{\geq t}\dagger S_x^{\leq t-1}\right) \geq l\right\}, \tag{41}$$

$$POS^l_P(Cl^{\leq}_t) = \left\{x \in U \mid \Sigma\sharp\left(S^{\leq t}_x \dagger S^{\geq t+1}_x\right) \geq l\right\}, \tag{42}$$

$$NNG^l_P(Cl^{\geq}_t) = \left\{x \in U \mid \Sigma\sharp\left(S^{\geq t}_x \dagger S^{\leq t-1}_x\right) > 1 - l\right\}, \tag{43}$$

$$NNG^l_P(Cl^{\leq}_t) = \left\{x \in U \mid \Sigma\sharp\left(S^{\leq t}_x \dagger S^{\geq t+1}_x\right) > 1 - l\right\}. \tag{44}$$

The model based on these definitions is called VP1-DFRSA. Note that, in the definitions, an object that supports $x \in Cl^{\geq}_{t+1}$ could be regarded as a piece of negative evidence for $x \in Cl^{\leq}_t$, and an object that supports $x \in Cl^{\leq}_{t-1}$ could be regarded as a piece of negative evidence for $x \in Cl^{\geq}_t$. In contrast to VC1-DFRSA, the positive and nonnegative regions in VP1-DFRSA are crisp sets. Furthermore, the membership of an object in the positive region or the nonnegative region of a union of classes does not depend on its membership degree in that union. Thus, $POS^l_P(Cl^{\geq}_t) \subseteq Cl^{\geq}_t$, $Cl^{\geq}_t \subseteq NNG^l_P(Cl^{\geq}_t)$, $POS^l_P(Cl^{\leq}_t) \subseteq Cl^{\leq}_t$, and $Cl^{\leq}_t \subseteq NNG^l_P(Cl^{\leq}_t)$ are not always valid in the VP-DFRSA model.

If the precision level is a fuzzy number $\tilde{l}$ over (0.5, 1], we can formulate the VP2-DFRSA model by using the ES measure. In this case, the P-positive and P-nonnegative regions of $Cl^{\geq}_t$ and $Cl^{\leq}_t$ are fuzzy sets with the following membership functions:

$$POS^{\tilde{l}}_P(Cl^{\geq}_t)(x) = \pi\Big(ES(S^{\geq t}_x \dagger S^{\leq t-1}_x) \geq \tilde{l}\Big), \tag{45}$$

$$POS^{\tilde{l}}_P(Cl^{\leq}_t)(x) = \pi\Big(ES(S^{\leq t}_x \dagger S^{\geq t+1}_x) \geq \tilde{l}\Big), \tag{46}$$

$$NNG^{\tilde{l}}_P(Cl^{\geq}_t)(x) = \pi\Big(ES(S^{\geq t}_x \dagger S^{\leq t-1}_x) > 1 - \tilde{l}\Big), \tag{47}$$

$$NNG^{\tilde{l}}_P(Cl^{\leq}_t)(x) = \pi\Big(ES(S^{\leq t}_x \dagger S^{\geq t+1}_x) > 1 - \tilde{l}\Big), \tag{48}$$

where $ES(\cdot\dagger\cdot)$ is regarded as a fuzzy number and the possibility of the comparison between fuzzy numbers is obtained based on the extension principle, as in the case of VC2-DFRSA.

Finally, if the precision level is a $(0.5, 1] \cap \mathbb{Q}$-valued random variable $\hat{l}$, then we can formulate the VP3-DFRSA model by using the ER measure defined above. In this case, the P-positive and P-nonnegative regions of $Cl^{\geq}_t$ and $Cl^{\leq}_t$ are fuzzy sets with the following membership functions:

$$POS^{\hat{l}}_P(Cl^{\geq}_t)(x) = Pr\Big(ER(S^{\geq t}_x \dagger S^{\leq t-1}_x) \geq \hat{l}\Big), \tag{49}$$

$$POS^{\hat{l}}_P(Cl^{\leq}_t)(x) = Pr\Big(ER(S^{\leq t}_x \dagger S^{\geq t+1}_x) \geq \hat{l}\Big), \tag{50}$$

$$NNG^{\hat{l}}_P(Cl^{\geq}_t)(x) = Pr\Big(ER(S^{\geq t}_x \dagger S^{\leq t-1}_x) > 1 - \hat{l}\Big), \tag{51}$$

$$NNG^{\hat{l}}_P(Cl^{\leq}_t)(x) = Pr\Big(ER(S^{\leq t}_x \dagger S^{\geq t+1}_x) > 1 - \hat{l}\Big), \tag{52}$$

where $ER(G\dagger F)$ is regarded as a random variable with the distribution function $Pr(ER(G\dagger F) = q) = ER(G\dagger F, q)$.

Proposition 4.4

(1) *VP1-DFRSA is reduced to VP-DRSA for the analysis of POIS.*
(2) *VP1-DFRSA is a special case of both VP2-DFRSA and VP3-DFRSA for the analysis of POIS.*

Proof The proof is similar to that of Proposition 4.2 and we omit the details. □

4.4. Attribute reduction and rule generation

The main objective of attribute reduction is to remove superfluous condition attributes so that we can find condition attributes related to the decision attribute. A number of approaches have been proposed for attribute reduction in DFSA. Susmaga et al. (2000) presented an approach based on the quality of sorting, while Inuiguchi et al. (2009) suggested several attribute reduction methods that preserve the positive or nonnegative regions. Our definition of reduct for DFRSA based on the degree of adherence to the dominance principle is similar to that proposed by Susmaga et al. (2000). Since different formulations of the lower and upper approximations in VC-DFRSA and VP-DFRSA do not influence the degree of adherence to the dominance principle, our attribute reduction approach for DFRSA can be applied uniformly to all VC-DFRSA and VP-DFRSA models. It is noteworthy that the methods proposed by Inuiguchi et al. (2009) can also be adapted to VC-DFRSA and VP-DFRSA models in a straightforward way.

As in the case of DFRSA, we can also derive fuzzy decision rules for the VC-DFRSA and VP-DFRSA models. To observe how these rules can be induced, let us take a closer look at the process of generating rules in DFRSA. Each rule in DFRSA is regarded as a *universal rule*, which means that all objects satisfying the antecedent should also satisfy the consequent. However, due to the fuzziness of the rule, it is not possible to make a crisp decision about whether an object satisfies the antecedent or the consequent. Thus, a confidence is attached to each rule to reflect its truth degree. The confidence is simply the membership degree of an object in the lower approximation, as shown in (15) and (16). Nevertheless, in VC-DFRSA and VP-DFRSA, the membership degree of an object in the lower approximation does not represent the confidence of the corresponding rule. Instead, it represents the extent that the confidence reaches the desired consistency or precision level. In other words, the membership degree represents a kind of second-order uncertainty about the rule because the confidence can be seen as the first-order uncertainty (or the degree of certainty) that the rule holds, whereas the membership degree specifies the uncertainty about the confidence value. Hence, there are two ways to derive decision rules in VC-DFRSA and VP-DFRSA. The first is to derive universal rules with relative cardinalities as their confidences, and the second is to derive certainty-quantified rules with membership degrees as their confidences. To simplify the formulation of these rules, we use $\varphi_x^{\geq}$ and $\varphi_x^{\leq}$ to denote $\bigwedge_{i\in P}(\geq_i, f_i(x))$ and $\bigwedge_{i\in P}(\leq_i, f_i(x))$, respectively. In addition, we let L denote the consistency or precision level in the respective models, and let $rc_x^{\geq t}$ and $rc_x^{\leq t}$ denote the relative cardinality in the respective lower approximations. For example, in the VP2-DFRSA model, L denotes a fuzzy number $\tilde{l}$ and $rc_x^{\geq t}$ denotes the relative cardinality $ES(S_x^{\geq t} \dagger S_x^{\leq t-1})$.

Let P denote a reduct of a POPIS and let $t \in V_d$. Then, by using the first kind of rule induction method, for each object x such that $\underline{P}^L(Cl_t^{\geq})(x) > 0$ or $POS_P^L(Cl_t^{\geq})(x) > 0$ (or above some pre-determined threshold), we can derive the $D_{\geq}$-decision rule

$$\varphi_x^{\geq} \longrightarrow_{rc_x^{\geq t}} (\geq_d, t); \tag{53}$$

and for each object x such that $\underline{P}^L(Cl_t^{\leq})(x) > 0$ or $POS_P^L(Cl_t^{\leq})(x) > 0$ (or above some pre-determined threshold), we can derive the $D_{\leq}$-decision rule

$$\varphi_x^{\leq} \longrightarrow_{rc_x^{\leq t}} (\leq_d, t). \tag{54}$$

The forms of the rules derived in this way are similar to those derived in DFRSA, so the process of applying the rules to new decision cases is essentially the same as that described in Section 3.4. However, because the confidences of these rules (i.e. the relative cardinalities) may be fuzzy numbers or random variables, we have to defuzzify them before the rules are applied. For example, a fuzzy number is replaced by its centroid and a random variable is replaced by its expectation.

On the other hand, by using the second kind of rule induction method, the $D_{\geq}$-decision rule and the $D_{\leq}$-decision rule take the following forms:

$$\varphi_x^{\geq} \longrightarrow {}_c^L (\geq_d, t) \text{ with } c = \underline{P}^L(Cl_t^{\geq})(x); \tag{55}$$

$$\varphi_x^{\leq} \longrightarrow {}_c^L (\leq_d, t) \text{ with } c = \underline{P}^L(Cl_t^{\leq})(x), \tag{56}$$

where $\underline{P}^L$ is replaced with POS_P^L in the VP-DFRSA model, and $\longrightarrow^L$ is a certainty-quantified implication operation between the antecedent and the consequent of the rules. The certainty-quantified implication can represent less certain rules as well as universal rules. In the three models for VC-DFRSA and VP-DFRSA, the intuitive meaning of a rule $\varphi \longrightarrow {}^L\psi$ may be

(1) "At least $L \cdot 100\%$ of φ's are ψ's" if L is a scalar;
(2) "Most φ's are ψ's" if L is a fuzzy number representing the fuzzy quantifier "most"; and
(3) "It is very likely that φ's are ψ's" if L is a random variable corresponding to "very likely."

Therefore, it is more complicated to apply the certainty-quantified rules to a new decision case, since it is still unclear how the formal semantics of the rules can be defined. Here, we suggest a simplified rule application process. First, we defuzzify the parameter L into a scalar and then merge it with the confidence of the rule to approximately transform the rule into a basic DFRSA rule. In other words, a rule $\varphi \longrightarrow {}_c^L\psi$ in VC-DFRSA or VP-DFRSA is transformed into a basic rule $\varphi \longrightarrow_{c\otimes l} \psi$, where l is the result of defuzzifying L. Then, the resultant rule can be applied to the new decision case in the standard way. However, we note that the approach may be over-simplified because it does not necessarily reflect the intended semantics of the certainty-quantified rules. The formal treatment of these rules based on the generalized theory of quantifiers (Glöckner 2004) is left for future research.

4.5. The illustrative example

In Section 3.5, we derive the rule $(\geq_2, L_5) \wedge (\geq_3, L_5) \longrightarrow_1 (\geq_d, 8)$ based on the membership of the object x_1 in $\underline{P}(Cl_8^{\geq})$. However, we cannot derive a rule $(\geq_2, L_5) \wedge (\geq_3, L_5) \longrightarrow_c (\geq_d, 10)$ with any $c > 0$, because there are three objects (x_2, x_3, and x_5) that are partially better (with degree $\frac{1}{4}$) than x_1 with respect to the two criteria but do not belong to $Cl_{10}^{\geq}$ at all (i.e. the membership degree is 0). On the other hand, since the membership degree of x_1 in $Cl_{10}^{\geq}$ is 1, we should expect that the rule $(\geq_2, L_5) \wedge (\geq_3, L_5) \longrightarrow (\geq_d, 10)$ is at least plausible to some extent. In this subsection, we show that VC-DFRSA or VP-DFRSA models can alleviate the problem by

Table 5. The fuzzy sets $Cl_{10}^{\geq}$, $Cl_9^{\leq}$, $D_P^+(x_1)$, $D_P^-(x_1)$, $S_{x_1}^{\geq 10}$, and $S_{x_1}^{\leq 9}$.

	$x_1 4$	x_2	x_3	x_4	x_5	x_6	x_7	x_8
$Cl_{10}^{\geq}$	1	0	0	0	0	0	0	0
$Cl_9^{\leq}$	$\frac{1}{2}$	1	1	1	1	1	1	1
$D_P^+(x_1)$	1	$\frac{1}{4}$	$\frac{1}{4}$	0	$\frac{1}{4}$	0	0	0
$D_P^-(x_1)$	1	1	1	1	1	1	1	1
$S_{x_1}^{\geq 10}$	1	0	0	0	0	0	0	0
$S_{x_1}^{\leq 9}$	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{4}$	0	$\frac{1}{4}$	0	0	0

inducing the rule with some confidence. To do that, we show in Table 5 the fuzzy sets $Cl^{\geq}_{10}$, $Cl^{\leq}_{9}$, $D^{+}_{P}(x_1)$, $D^{-}_{P}(x_1)$, $S^{\geq 10}_{x_1}$, and $S^{\leq 9}_{x_1}$ that are needed in the rule induction process of VC-DFRSA or VP-DFRSA models.

4.5.1. VC-DFRSA models

For VC-DFRSA models with three kinds of relative cardinalities, instantiating (18), (21), and (25) with the data in Table 5, we obtain

$$\Sigma\sharp(Cl^{\geq}_{10}/D^{+}_{P}(x_1)) = \frac{1}{1+3\cdot\frac{1}{4}} = \frac{4}{7},$$

$$ES(Cl^{\geq}_{10}/D^{+}_{P}(x_1))(q) = \begin{cases} 1, & \text{if } q = 1; \\ \frac{1}{4}, & \text{if } q = \frac{1}{4}; \\ 0, & \text{otherwise,} \end{cases}$$

$$ER(Cl^{\geq}_{10}/D^{+}_{P}(x_1))(q) = \begin{cases} \frac{3}{4}, & \text{if } q = 1; \\ \frac{1}{4}, & \text{if } q = \frac{1}{4}; \\ 0, & \text{otherwise.} \end{cases}$$

The relative cardinalities are then applied to the rule generation process based on the three kinds of VC-DFRSA models. For VC1-DFRSA, if the consistency level is a scalar $l \in \left(\frac{1}{2}, \frac{4}{7}\right]$, then $\underline{P}^{l}(Cl^{\geq}_{10})(x_1) = Cl^{\geq}_{10}(x_1) = 1$. Therefore, by using (53), we can derive

$$(\geq_2, L_5) \wedge (\geq_3, L_5) \longrightarrow_{\frac{4}{7}} (\geq_d, 10);$$

or by using (55), we can derive

$$(\geq_2, L_5) \wedge (\geq_3, L_5) \longrightarrow^{l}_{1} (\geq_d, 10).$$

Note that the confidence of the rule is at most $\frac{4}{7} \approx 0.57$. Thus, these rules are not quite strong.

For VC2-DFRSA, let the consistency level be a fuzzy number $\tilde{l}$ over $(0.5, 1] \cap \mathbb{Q}$. Then,

$$\pi(ES(Cl^{\geq}_{10}/D^{+}_{P}(x_1)) \geq \tilde{l}) = \max\left(\sup_{1\geq x}\min(1, \mu_{\tilde{l}}(x)), \sup_{\frac{1}{4}\geq x}\min\left(\frac{1}{4}, \mu_{\tilde{l}}(x)\right)\right) = 1.$$

Thus, $\underline{P}^{\tilde{l}}(Cl^{\geq}_{10})(x_1) = 1$, and by using (53), we can derive

$$(\geq_2, L_5) \wedge (\geq_3, L_5) \longrightarrow_{c} (\geq_d, 10),$$

where $c = ES(Cl^{\geq}_{10}/D^{+}_{P}(x_1))$, or by using (55), we can derive

$$(\geq_2, L_5) \wedge (\geq_3, L_5) \longrightarrow^{\tilde{l}}_{1} (\geq_d, 10).$$

By using the center-of-gravity method to defuzzify $ES(Cl^{\geq}_{10}/D^{+}_{P}(x_1))$, the confidence of the first rule is $\left(1\cdot 1 + \frac{1}{4}\cdot\frac{1}{4}\right) \Big/ \left(1+\frac{1}{4}\right) = \frac{17}{20} = 0.85$, which is stronger than the rules derived by VC1-DFRSA. We can even have stronger rules by defuzzifying $\tilde{l}$ in the second rule if $\tilde{l}$ is a sufficiently large fuzzy number. In the extreme case, if $\mu_{\tilde{l}}(1) = 1$ and $\mu_{\tilde{l}}(x) = 0$ for $x \in (0.5, 1) \cap \mathbb{Q}$ (i.e. $\tilde{l}$ degenerates into the scalar 1), then the confidence of the second rule is 1 and it becomes a fully certain rule.

For VC3-DFRSA, the consistency level is a $(0.5, 1] \cap \mathbb{Q}$-valued random variable $\hat{l}$. Hence, $Pr(ER(Cl^{\geq}_{10}/D^{+}_{P}(x_1)) \geq \hat{l}) = \frac{3}{4} \cdot Pr(\hat{l} \leq 1) + \frac{1}{4} \cdot Pr\left(\hat{l} \leq \frac{1}{4}\right) = \frac{3}{4}$ and $\underline{P}^{\hat{l}}(Cl^{\geq}_{10})(x_1) = \frac{3}{4}$. Therefore, by using (53), we can derive

$$(\geq_2, L_5) \wedge (\geq_3, L_5) \longrightarrow_c (\geq_d, 10),$$

where $c = ER(Cl^{\geq}_{10}/D^{+}_{P}(x_1))$; or by using (55), we can derive

$$(\geq_2, L_5) \wedge (\geq_3, L_5) \longrightarrow^{\hat{l}}_{\frac{3}{4}} (\geq_d, 10).$$

The expectation of the probability distribution $ER(Cl^{\geq}_{10}/D^{+}_{P}(x_1))$ is $1 \cdot \frac{3}{4} + \frac{1}{4} \cdot \frac{1}{4} = \frac{13}{16} \approx 0.81$. Thus, the first rule has the similar confidence with the first rule of VC2-DFRSA above. On the other hand, the confidence of the second rule is $\frac{3}{4} \cdot E(\hat{l})$, which is at most $\frac{3}{4}$ since the expectation of a $(0.5, 1] \cap \mathbb{Q}$-valued random variable $\hat{l})$ is at most 1. Hence, the second rule is weaker than the first one, although it is still stronger than the rules derived from VC1-DFRSA.

4.5.2. VP-DFRSA models

For VP-DFRSA models with three kinds of relative cardinalities, instantiating (38), (39), (40) with the data in Table 5, we have

$$\Sigma\sharp\left(S^{\geq 10}_{x_1} \dagger S^{\leq 9}_{x_1}\right) = \frac{1}{\frac{1}{2} + 3 \cdot \frac{1}{4}} = \frac{4}{5},$$

$$ES\left(S^{\geq 10}_{x_1} \dagger S^{\leq 9}_{x_1}\right)(q) = \begin{cases} 1, & \text{if } q = 1; \\ \frac{1}{2}, & \text{if } q = \frac{1}{2}; \\ \frac{1}{4}, & \text{if } q = \frac{1}{5}; \\ 0, & \text{otherwise,} \end{cases}$$

$$ER\left(S^{\geq 10}_{x_1} \dagger S^{\leq 9}_{x_1}\right)(q) = \begin{cases} \frac{1}{2}, & \text{if } q = 1; \\ \frac{1}{4}, & \text{if } q = \frac{1}{2}; \\ \frac{1}{4}, & \text{if } q = \frac{1}{5}; \\ 0, & \text{otherwise.} \end{cases}$$

The relative cardinalities are applied to the rule generation process based on the three kinds of VP-DFRSA models. For VP1-DFRSA, if the precision level is a scalar $l \in \left(\frac{1}{2}, \frac{4}{5}\right]$, then $x_1 \in POS^{l}_{P}(Cl^{\geq}_{10})$ since $\Sigma\sharp\left(S^{\geq 10}_{x_1} \dagger S^{\leq 9}_{x_1}\right) \geq l$. Thus, by using (53), we can derive

$$(\geq_2, L_5) \wedge (\geq_3, L_5) \longrightarrow_{\frac{4}{5}} (\geq_d, 10);$$

or by using (55), we can derive

$$(\geq_2, L_5) \wedge (\geq_3, L_5) \longrightarrow^{l}_{1} (\geq_d, 10).$$

Note that the confidence of the rule is at most $\frac{4}{5}$, which is similar to that of the first rule derived from VC3-DFRSA.

For VP2-DFRSA, the precision level is a fuzzy number $\tilde{l}$ over $(0.5, 1] \cap \mathbb{Q}$. Obviously, $POS^{\tilde{l}}_{P}(Cl^{\geq}_{10})(x_1) = \pi\left(ES\left(S^{\geq 10}_{x_1} \dagger S^{\leq 9}_{x_1}\right) \geq \tilde{l}\right) = 1$. Thus, by using (53), we can derive

$$(\geq_2, L_5) \wedge (\geq_3, L_5) \longrightarrow_c (\geq_d, 10),$$

where $c = ES\left(S_{x_1}^{\geq 10} \dagger S_{x_1}^{\leq 9}\right)$; or by using (55), we can derive

$$(\geq_2, L_5) \wedge (\geq_3, L_5) \longrightarrow_1^{\tilde{l}} (\geq_d, 10).$$

By using the center-of-gravity method to defuzzify $ES\left(S_{x_1}^{\geq 10} \dagger S_{x_1}^{\leq 9}\right)$, the confidence of the first rule is $\left(1 \cdot 1 + \frac{1}{2} \cdot \frac{1}{2} + \frac{1}{4} \cdot \frac{1}{5}\right) \Big/ \left(1 + \frac{1}{2} + \frac{1}{4}\right) = \frac{26}{35} \approx 0.74$, which is near to that of the second rule derived by VC3-DFRSA. On the other hand, the second rule is completely the same as the second rule derived by VC2-DFRSA. Thus, we can obtain a fully certain rule as in the case of VC2-DFRSA.

For VP3-DFRSA, the precision level is a $(0.5, 1] \cap \mathbb{Q}$-valued random variable $\hat{l}$. Then, we have $POS_P^{\hat{l}}\left(Cl_{10}^{\geq}\right)(x_1) = Pr\left(ER\left(S_{x_1}^{\geq 10} \dagger S_{x_1}^{\leq 9}\right) \geq \hat{l}\right) = \frac{1}{2} \cdot Pr(\hat{l} \leq 1) + \frac{1}{4} \cdot Pr\left(\hat{l} \leq \frac{1}{2}\right) + \frac{1}{4} \cdot Pr\left(\hat{l} \leq \frac{1}{5}\right) = \frac{1}{2}$ and $\underline{P}^{\hat{l}}\left(Cl_{10}^{\geq}\right)(x_1) = \frac{1}{2}$. Thus, by using (53), we can derive

$$(\geq_2, L_5) \wedge (\geq_3, L_5) \longrightarrow_c (\geq_d, 10),$$

where $c = ER\left(S_{x_1}^{\geq 10} \dagger S_{x_1}^{\leq 9}\right)$; or by using (55), we can derive

$$(\geq_2, L_5) \wedge (\geq_3, L_5) \longrightarrow_{\frac{1}{2}}^{\hat{l}} (\geq_d, 10).$$

The expectation of the probability distribution $ER\left(S_{x_1}^{\geq 10} \dagger S_{x_1}^{\leq 9}\right)$ is $1 \cdot \frac{1}{2} + \frac{1}{2} \cdot \frac{1}{4} + \frac{1}{5} \cdot \frac{1}{4} = \frac{27}{40} \approx 0.675$. On the other hand, the confidence of the second rule is $\frac{1}{2} \cdot E(\hat{l})$, which is the weakest rule and can almost be dismissed.

4.6. A case study

In this subsection, we present a comparative study of the proposed models based on a small data-set of student evaluations. In the data-set, every student's passing status (PS) is determined by his/her grades on three academic courses mathematics ($Math$), computer science (CS), and literature (Lit). Each course is graded according to a grading scale that ranges from 4 (Excellent) to 1 (Failing) and the passing status is either 1 (pass) or 0 (fail). Thus, the data-set is represented as a POIS $T = (U, A, \{(V_i, \succeq_i) \mid i \in A\}, \{f_i \mid i \in A\})$, where U is the set of students; $A = \{Math, CS, Lit, PS\}$; $V_i = \{1, 2, 3, 4\}$ for $i = Math, CS$, and Lit and $V_d = \{0, 1\}$ for $d = PS$ with $4 \succ_i 3 \succ_i 2 \succ_i 1$ and $1 \succ_d 0$. The POIS is shown in Table 6. For the purpose of comparison, we intentionally include several noisy data items (i.e. $x_6, x_{10}, x_{18}, x_{23}$) in the table. Although this POIS is simply an artificially synthetic data-set, its basic characteristics could exist in many real datasets.

In this example, the set of condition attributes is $P = \{Math, CS, Lit\}$. Because T is a POIS, $D_P^+(x)$ and $D_P^-(x)$ for every $x \in U$ and $Cl_1^{\geq}$ and $Cl_0^{\leq}$ are all crisp sets. From the POIS, we can derive

$$Cl_1^{\geq} = \{x_1, x_2, x_3, x_4, x_5, x_7, x_8, x_9, x_{11}, x_{12}, x_{13}, x_{18}, x_{23}\}$$

$$Cl_0^{\leq} = \{x_6, x_{10}, x_{14}, x_{15}, x_{16}, x_{17}, x_{19}, x_{20}, x_{21}, x_{22}, x_{24}, x_{25}\}$$

and the P-dominating and P-dominated sets of every $x \in U$ are specified in Table 7. By Lemmas 4.1 and 4.3, the functions $\Sigma\sharp$, ES, and ER depend only on the cardinalities of $D_P^+(x)$, $D_P^-(x)$, $D_P^+(x) \cap Cl_1^{\geq}$, $D_P^-(x) \cap Cl_0^{\leq}$, $S_x^{\leq 0} = D_P^+(x) \cap Cl_0^{\leq}$, and $S_x^{\geq 1} = D_P^-(x) \cap Cl_1^{\geq}$. The cardinalities of these sets are shown in Table 8. Since $|D_P^+(x)| = |D_P^+(x) \cap Cl_1^{\geq}| + |D_P^+(x) \cap Cl_0^{\leq}|$ and $|D_P^-(x)| = |D_P^-(x) \cap Cl_0^{\leq}| + |D_P^-(x) \cap Cl_1^{\geq}|$, we do not explicitly specify them in the table. Then, we can derive the lower and upper approximations of the decision classes according to different models. For the simplicity of the presentation, we only consider the lower approximation since

Table 6. A POIS of student evaluation.

$U \setminus A$	*Math*	*CS*	*Lit*	*PS*
x_1	4	4	4	1
x_2	4	4	4	1
x_3	4	4	4	1
x_4	4	3	4	1
x_5	4	4	3	1
x_6	4	4	3	0
x_7	4	3	4	1
x_8	4	4	3	1
x_9	3	4	4	1
x_{10}	4	3	3	0
x_{11}	4	3	3	1
x_{12}	3	4	4	1
x_{13}	3	3	3	1
x_{14}	3	3	2	0
x_{15}	2	3	2	0
x_{16}	2	2	3	0
x_{17}	2	3	3	0
x_{18}	2	2	2	1
x_{19}	2	1	2	0
x_{20}	1	2	2	0
x_{21}	1	1	3	0
x_{22}	1	3	1	0
x_{23}	1	1	2	1
x_{24}	1	2	1	0
x_{25}	1	1	1	0

it is more important for the process of rule induction. To calculate the lower approximations in different models, we need the following four quantities for every $x \in U$:

$$q_1(x) = \frac{|D_P^+(x)|}{|D_P^+(x) \cap Cl_1^{\geq}|}, \quad q_2(x) = \frac{|D_P^-(x)|}{|D_P^-(x) \cap Cl_0^{\leq}|},$$

$$q_3(x) = \frac{|S_x^{\geq 1}| + |S_x^{\leq 0}|}{|S_x^{\geq 1}|}, \quad q_4(x) = \frac{|S_x^{\geq 1}| + |S_x^{\leq 0}|}{|S_x^{\leq 0}|}.$$

Note that $q_3(x) + q_4(x) = 1$ for any $x \in U$. The values of $q_i(x)$ for $1 \leq i \leq 4$ are shown in Table 9.

As the baseline of the comparison, we start with the DRSA model. By using Table 9, we can find that

$$\underline{P}(Cl_1^{\geq}) = \{x_1, x_2, x_3, x_4, x_7, x_9, x_{12}\} \text{ and } \underline{P}(Cl_0^{\leq}) = \{x_{22}\}.$$

That is, $x \in \underline{P}(Cl_1^{\geq})$ iff $q_1(x) = 1$ and $x \in \underline{P}(Cl_0^{\leq})$ iff $q_2(x) = 1$. Each object in the lower approximations can induce a decision rule. For example, x_7 can induce the rule

$$(\geq_{Math}, 4) \wedge (\geq_{CS}, 3) \wedge (\geq_{Lit}, 4) \longrightarrow (\geq_{PS}, 1),$$

which means that if a student's grade of mathematics is excellent, grade of computer science is at least good, and grade of literature is excellent, then his/her passing status is "pass." As expected, the noisy items block many useful rules. For examples, the $D_{\geq}$ decision rules based on x_5, x_8, x_{11} or $D_{\leq}$ decision rules based on $x_{19}, x_{20}, x_{21}, x_{24}, x_{25}$ cannot be derived by using the DRSA model.

Table 7. The dominating and dominated sets.

U	$D_P^+(x)$	$D_P^-(x)$
x_1	$\{x_1, x_2, x_3\}$	U
x_2	$\{x_1, x_2, x_3\}$	U
x_3	$\{x_1, x_2, x_3\}$	U
x_4	$\{x_1, x_2, x_3, x_4, x_7\}$	$\{x_4, x_7, x_{10}, x_{11}, x_{13}, \ldots, x_{25}\}$
x_5	$\{x_1, x_2, x_3, x_5, x_6, x_8\}$	$\{x_5, x_6, x_8, x_{10}, x_{11}, x_{13}, \ldots, x_{25}\}$
x_6	$\{x_1, x_2, x_3, x_5, x_6, x_8\}$	$\{x_5, x_6, x_8, x_{10}, x_{11}, x_{13}, \ldots, x_{25}\}$
x_7	$\{x_1, x_2, x_3, x_4, x_7\}$	$\{x_4, x_7, x_{10}, x_{11}, x_{13}, \ldots, x_{25}\}$
x_8	$\{x_1, x_2, x_3, x_5, x_6, x_8\}$	$\{x_5, x_6, x_8, x_{10}, x_{11}, x_{13}, \ldots, x_{25}\}$
x_9	$\{x_1, x_2, x_3, x_9, x_{12}\}$	$\{x_9, x_{12}, \ldots, x_{25}\}$
x_{10}	$\{x_1, \ldots, x_8, x_{10}, x_{11}\}$	$\{x_{10}, x_{11}, x_{13}, \ldots, x_{25}\}$
x_{11}	$\{x_1, \ldots, x_8, x_{10}, x_{11}\}$	$\{x_{10}, x_{11}, x_{13}, \ldots, x_{25}\}$
x_{12}	$\{x_1, x_2, x_3, x_9, x_{12}\}$	$\{x_9, x_{12}, \ldots, x_{25}\}$
x_{13}	$\{x_1, \ldots, x_{13}\}$	$\{x_{13}, \ldots, x_{25}\}$
x_{14}	$\{x_1, \ldots, x_{14}\}$	$\{x_{14}, x_{15}, x_{18}, x_{19}, x_{20}, x_{22}, x_{23}, x_{24}, x_{25}\}$
x_{15}	$\{x_1, \ldots, x_{15}, x_{17}\}$	$\{x_{15}, x_{18}, x_{19}, x_{20}, x_{22}, x_{23}, x_{24}, x_{25}\}$
x_{16}	$\{x_1, \ldots, x_{13}, x_{16}, x_{17}\}$	$\{x_{16}, x_{18}, x_{19}, x_{20}, x_{21}, x_{23}, x_{24}, x_{25}\}$
x_{17}	$\{x_1, \ldots, x_{13}, x_{17}\}$	$\{x_{15}, \ldots, x_{25}\}$
x_{18}	$\{x_1, \ldots, x_{18}\}$	$\{x_{18}, x_{19}, x_{20}, x_{23}, x_{24}, x_{25}\}$
x_{19}	$\{x_1, \ldots, x_{19}\}$	$\{x_{19}, x_{23}, x_{25}\}$
x_{20}	$\{x_1, \ldots, x_{18}, x_{20}\}$	$\{x_{20}, x_{23}, x_{24}, x_{25}\}$
x_{21}	$\{x_1, \ldots, x_{13}, x_{16}, x_{17}, x_{21}\}$	$\{x_{21}, x_{23}, x_{25}\}$
x_{22}	$\{x_1, \ldots, x_{15}, x_{17}, x_{22}\}$	$\{x_{22}, x_{24}, x_{25}\}$
x_{23}	$\{x_1, \ldots, x_{21}, x_{23}\}$	$\{x_{23}, x_{25}\}$
x_{24}	$\{x_1, \ldots, x_{18}, x_{20}, x_{22}, x_{24}\}$	$\{x_{24}, x_{25}\}$
x_{25}	U	$\{x_{25}\}$

VC-DRSA and VP-DRSA are two proposal that can alleviate the problem to some extent. We have seen that VC1-DFRSA and VP1-DFRSA are, respectively, equivalent to VC-DRSA and VP-DRSA in the case of crisp information systems. In these two models, we have to specify a scalar parameter l as the level of consistency or precision. If $l = 0.8$, then for VC1-DFRSA (i.e. VC-DRSA), we have $x \in \underline{P^{0.8}}(Cl_1^{\geq})$ iff $q_1(x) \geq 0.8$ and $x \in \underline{P^{0.8}}(Cl_0^{\leq})$ iff $q_2(x) \geq 0.8$, thus

$$\underline{P^{0.8}}(Cl_1^{\geq}) = \{x_1, \cdots, x_5, x_7, x_8, x_9, x_{11}, x_{12}, x_{13}\} \text{ and } \underline{P^{0.8}}(Cl_0^{\leq}) = \{x_{17}, x_{22}\};$$

and for VP1-DFRSA (i.e. VP-DRSA), we have $x \in POS_P^{0.8}(Cl_1^{\geq})$ iff $q_3(x) \geq 0.8$ and $x \in POS_P^{0.8}(Cl_0^{\leq})$ iff $q_4(x) \geq 0.8$, thus

$$POS_P^{0.8}(Cl_1^{\geq}) = \{x_1, \cdots, x_9, x_{12}\} \text{ and } POS_P^{0.8}(Cl_0^{\leq}) = \{x_{19}, \cdots, x_{25}\}.$$

Now, several rules blocked by noisy data in DRSA are derivable in VC1-DFRSA and VP1-DFRSA. For example, by using x_5, we have the $D_{\geq}$ decision rule

$$(\geq_{Math}, 4) \wedge (\geq_{CS}, 4) \wedge (\geq_{Lit}, 3) \longrightarrow^{0.8} (\geq_{PS}, 1),$$

which means that at least 80% of students who satisfy the antecedent of the rule will pass.

We can see that there is little difference between $\underline{P^{0.8}}(Cl_1^{\geq})$ and $POS_P^{0.8}(Cl_1^{\geq})$ and it seems that all expected $D_{\geq}$ decision rules can be induced in this case. However, it seems that VP-DRSA

Table 8. The cardinalities of $D_P^+(x) \cap Cl_1^{\geq}$, $S_x^{\leq 0}$, $D_P^-(x) \cap Cl_0^{\leq}$, and $S_x^{\geq 1}$.

U	$\lvert D_P^+(x) \cap Cl_1^{\geq}\rvert$	$\lvert D_P^+(x) \cap Cl_0^{\leq}\rvert$	$\lvert D_P^-(x) \cap Cl_0^{\leq}\rvert$	$\lvert D_P^-(x) \cap Cl_1^{\geq}\rvert$
x_1	3	0	12	13
x_2	3	0	12	13
x_3	3	0	12	13
x_4	5	0	11	6
x_5	5	1	12	6
x_6	5	1	12	6
x_7	5	0	11	6
x_8	5	1	12	6
x_9	5	0	10	5
x_{10}	8	2	11	4
x_{11}	8	2	11	4
x_{12}	5	0	10	5
x_{13}	11	2	10	3
x_{14}	11	3	7	2
x_{15}	11	5	6	2
x_{16}	11	4	6	2
x_{17}	11	3	9	2
x_{18}	12	6	4	2
x_{19}	12	7	2	1
x_{20}	12	7	3	1
x_{21}	11	8	2	1
x_{22}	11	6	3	0
x_{23}	13	9	1	1
x_{24}	12	9	1	1
x_{25}	13	12	1	1

significantly outperforms VC-DRSA in the derivation of $D_{\leq}$ decision rules since $POS_P^{0.8}(Cl_0^{\leq})$ include all expected objects, whereas $\underline{P^{0.8}}(Cl_0^{\leq})$ has only one more element (i.e. x_{17}) than $\underline{P}(Cl_0^{\leq})$ above. In addition, x_{17} seems a boundary case. Hence, it is not a definite advantage to include x_{17} in the lower approximation.

Although VC-DRSA and VP-DRSA can alleviate the effect of noisy data, it is not clear how to choose the parameter l appropriately. In the current case, $l = 0.8$ seems an optimal choice. However, it is not clear why we should not choose $l = 0.85$ or $l = 0.9$. Moreover, a small perturbation on l may cause abrupt change on the results. For example, if $l = 0.9$, then $\underline{P^{0.9}}(Cl_1^{\geq})$ collapses into $\underline{P}(Cl_1^{\geq})$. The situation has been observed in Slezak (2005) and Slezak and Ziarko (2005) in the context of VPRS and the Bayesian rough set (BRS) model is proposed to address the issue. Instead of using a fixed but arbitrarily given parameter, BRS sets the consistency or precision level as the prior probability of the occurrence of the decision class in the general population. In the current case, the prior probabilities of $Cl_1^{\geq}$ and $Cl_0^{\leq}$ are $\frac{13}{25}$ and $\frac{12}{25}$, respectively, which seems too loose for the rule induction process.

On the other hand, VC2-DRSA and VP2-DRSA do not completely free the users from the burden of choosing the parameters. However, these models allow a more flexible way to specify their consistency or precision levels with soft parameters. Instead of a precise parameter, it is allowed to use a linguistic term like "moderately possible," "highly possible," "about 0.8," etc. For instance, let the parameter be vaguely specified as $\tilde{l}$ = "about 0.9" and assume that $\tilde{l}$ is characterized by the bell-shaped membership function:

Table 9. The values of $q_1(x), q_2(x), q_3(x), q_4(x)$.

U	$q_1(x)4$	$q_2(x)$	$q_3(x)$	$q_4(x)$
x_1	1	$\frac{12}{25}$	1	0
x_2	1	$\frac{12}{25}$	1	0
x_3	1	$\frac{12}{25}$	1	0
x_4	1	$\frac{11}{17}$	1	0
x_5	$\frac{5}{6}$	$\frac{2}{3}$	$\frac{6}{7}$	$\frac{1}{7}$
x_6	$\frac{5}{6}$	$\frac{2}{3}$	$\frac{6}{7}$	$\frac{1}{7}$
x_7	1	$\frac{11}{17}$	1	0
x_8	$\frac{5}{6}$	$\frac{2}{3}$	$\frac{6}{7}$	$\frac{1}{7}$
x_9	1	$\frac{2}{3}$	1	0
x_{10}	$\frac{4}{5}$	$\frac{11}{15}$	$\frac{2}{3}$	$\frac{1}{3}$
x_{11}	$\frac{4}{5}$	$\frac{11}{15}$	$\frac{2}{3}$	$\frac{1}{3}$
x_{12}	1	$\frac{2}{3}$	1	0
x_{13}	$\frac{11}{13}$	$\frac{10}{13}$	$\frac{3}{5}$	$\frac{2}{5}$
x_{14}	$\frac{11}{14}$	$\frac{7}{9}$	$\frac{2}{5}$	$\frac{3}{5}$
x_{15}	$\frac{11}{16}$	$\frac{3}{4}$	$\frac{2}{7}$	$\frac{5}{7}$
x_{16}	$\frac{11}{15}$	$\frac{3}{4}$	$\frac{1}{3}$	$\frac{2}{3}$
x_{17}	$\frac{11}{14}$	$\frac{9}{11}$	$\frac{2}{5}$	$\frac{3}{5}$
x_{18}	$\frac{2}{3}$	$\frac{2}{3}$	$\frac{1}{4}$	$\frac{3}{4}$
x_{19}	$\frac{12}{19}$	$\frac{2}{3}$	$\frac{1}{8}$	$\frac{7}{8}$
x_{20}	$\frac{12}{19}$	$\frac{3}{4}$	$\frac{1}{8}$	$\frac{7}{8}$
x_{21}	$\frac{11}{19}$	$\frac{2}{3}$	$\frac{1}{9}$	$\frac{8}{9}$
x_{22}	$\frac{11}{17}$	1	0	1
x_{23}	$\frac{13}{22}$	$\frac{1}{2}$	$\frac{1}{10}$	$\frac{9}{10}$
x_{24}	$\frac{12}{21}$	$\frac{1}{2}$	$\frac{1}{10}$	$\frac{9}{10}$
x_{25}	$\frac{13}{25}$	$\frac{1}{2}$	$\frac{1}{13}$	$\frac{12}{13}$

$$\mu_{\tilde{l}}(v) = \frac{1}{1 + 2|v - 0.9|}.$$

Then, according to the definition, for $1 \leq i \leq 4$,

$$\pi(q_i(x) \geq \tilde{l}) = \sup_{0.5 < v \leq q_i(x)} \mu_{\tilde{l}}(v) = \begin{cases} 1, & \text{if } q_i(x) \geq 0.9, \\ \frac{1}{1+2|q_i(x)-0.9|}, & \text{if } q_i(x) < 0.9. \end{cases}$$

Hence, the lower approximations in VC2-DFRSA and the positive regions in VP2-DFRSA can be computed by using (30),(31),(45), and (46). For example, for any $x \in Cl_1^{\geq}$, we have $\underline{P^{\tilde{l}}}(Cl_1^{\geq})(x) = 1$ if $q_1(x) \geq 0.9$ and $\underline{P^{\tilde{l}}}(Cl_1^{\geq})(x) = \mu_{\tilde{l}}(q_1(x))$ if $q_1(x) < 0.9$. Hence, although the core[5] of $\underline{P^{\tilde{l}}}(Cl_1^{\geq})$ still collapses into $\underline{P}(Cl_1^{\geq})$, it contains other elements of $\underline{P^{0.8}}(Cl_1^{\geq})$ to some extent. For example, x_5 and x_8 belong to $\underline{P}(Cl_1^{\geq})$ with membership degree $\frac{15}{17} \approx 0.88$. By using (55) with these two objects, we can derive a rule

$$(\geq_{Math}, 4) \wedge (\geq_{CS}, 4) \wedge (\geq_{Lit}, 3) \longrightarrow_{0.88}^{\text{about } 0.9} (\geq_{PS}, 1),$$

which means that it is 88% certain that nearly 90% of students who satisfy the antecedent of the rule will pass.

VC3-DRSA and VP3-DRSA are similar to VC2-DRSA and VP2-DRSA on the aspect of flexibility. The main difference is that in VC3-DRSA and VP3-DRSA, the parameters have a probabilistic interpretation, whereas in VC2-DRSA and VP2-DRSA, they are interpreted in a possibilistic way. For instance, let the parameter be specified as $\hat{l}$ =“likely” and assume that $\hat{l}$ is a uniform random variable on (0.5, 1], i.e. its cumulative distribution function is defined as $F_{\hat{l}}(v) = 2v - 1$. Then, the membership functions of the lower approximations in VC3-DFRSA and the positive regions in VP3-DFRSA can be determined by $Pr(q_i(x) \geq \hat{l}) = 2q_i(x) - 1$ for $1 \leq i \leq 4$. Let us take $\underline{P}^{\hat{l}}(Cl_1^{\geq})$ as an example. Its core is still equal to $\underline{P}(Cl_1^{\geq})$. However, it also contains other elements to some extent. For example, x_5 and x_8 belong to $\underline{P}(Cl_1^{\geq})$ with membership degree $2 * \frac{5}{6} - 1 \approx 0.67$. By using (55) with these two objects, we can derive a rule

$$(\geq_{Math}, 4) \wedge (\geq_{CS}, 4) \wedge (\geq_{Lit}, 3) \longrightarrow_{0.67}^{\text{likely}} (\geq_{PS}, 1),$$

which means that it is 67% certain that students who satisfy the antecedent of the rule are likely to pass. Apparently, the rules derived from the probabilistic interpretation are more conservative than those derived from the possibilistic interpretation.

From the examples above, it may be considered that the membership degrees of x in a lower approximation or a positive region is simply the rescaling of the corresponding quantity $q_i(x)$. However, this viewpoint may be misleading because it only holds for the special case of POIS. In a proper POPIS, the relative cardinalities in VC2-DRSA, VC3-DRSA, VP2-DRSA, and VP3-DRSA are no longer single-point distributions and cannot be reduced to a quantity like $q_i(x)$.

5. Concluding remarks

In this paper, we review the dominance-based fuzzy rough set approach (DFRSA) framework, which can be applied to the reduction of criteria and the induction of rules for decision analysis in a POPIS. We show that the generation of some intuitively justified rules may be blocked by the partial violation of the rules caused by the fuzzy dominance relation. To cope with the problem, we propose the variable consistency models and the variable precision models of the DFRSA framework based on fuzzy cardinalities.

In contrast to other approaches that deal with imprecise evaluations and assignments, DFRSA induces fuzzy rules instead of qualitative rules. Thus, it would be worthwhile comparing DFRSA with other extensions of DRSA for handling uncertain information systems, e.g. those proposed by Greco et al. (2007) and Sakai et al. (2011).

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.

As shown in Fan (2011a), the computational complexity of the naive implementation of DFRSA prevents us from 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 a theoretical framework to deal with rough set analysis of possibilistic data. 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 used real examples to illustrate how the rules can be induced and how the notions of fuzzy cardinalities are used to alleviate the problem of the partial violation of rules. 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.

Acknowledgements

This work was partially supported by the National Science Council of Taiwan under grants NSC 98-2221-E-001-013-MY3 and NSC 100-2410-H-346-002. We wish to thank the guest editors and three anonymous referees for their constructive suggestions.

Notes

1. Also called knowledge representation systems, data tables, or attribute-value systems
2. For the properties of these operations, see a standard reference on fuzzy logic, such as Hájek (1998)
3. That is, $a \to b = 1$ if $a \leq b$ and $a \to b = b$ if $a > b$ for any $a, b \in [0, 1]$.
4. In this section, we use the notations E, ES, ER, and ED introduced in Delgado et al. (2002) to denote fuzzy cardinalities. Unfortunately, it is not explicitly mentioned in Delgado et al. (2002) what the acronyms stand for.
5. A core of a fuzzy set F is defined as $\{x \in U \mid \mu_F(x) = 1\}$.

Notes on contributors

Tuan-Fang Fan received a BS in material science and an MS in computer science from National Cheng-Kung University, Tainan, Taiwan, in 1983 and 1989, respectively, and a PhD in information management from National Chiao-Tung University, Hsinchu, Taiwan, in 2008. She is currently an associate professor in the Department of Computer Science and Information Engineering, National Penghu University of Science and Technology, Penghu, Taiwan. Her research interests include database, data mining, decision science, and rough set theory.

Churn-Jung Liau received a BS, an MS and a PhD in computer science and information engineering from National Taiwan University, Taipei, Taiwan, in 1985, 1987 and 1992, respectively. He then joined the Institute of Information Science, Academia Sinica, Taipei, Taiwan and is currently a tenured full research fellow. His research interests include artificial intelligence, uncertainty reasoning, and logic.

Dr. Duen-Ren Liu is a professor of the Institute of Information Management at the National Chiao Tung University of Taiwan. He received the BS and MS degrees in Computer Science from the National Taiwan University and his PhD in Computer Science from the University of Minnesota. His research interests include information systems, knowledge engineering and management, data mining, electronic commerce and recommender systems.

References

de Luca, A., and S. Termini. 1972. "A Definition of a Nonprobabilistic Entropy in the Setting of Fuzzy Sets Theory." *Information and Control* 20 (4): 301–312.

Delgado, M., D. Sánchez, M. Martín-Bautista, and M. Vila. 2002. "A Probabilistic Definition of a Nonconvex Fuzzy Cardinality." *Fuzzy Sets and Systems* 126 (2): 177–190.

Dubois, D., and H. Prade. 1990. "Rough fuzzy Sets and Fuzzy Rough Sets." *International Journal of General Systems* 17 : 191–209.

Dubois, D., and H. Prade. 2001. "Possibility theory, probability theory and multiplevalued logics: A clarification." *Annals of Mathematics and Artificial Intelligence* 32 (1–4): 35–66.

Fan, T., C. Liau, and D. Liu. 2009. "Dominance-Based Rough Set Analysis of Uncertain Data Table." In *Proceedings of the International Fuzzy Systems Association (IFSA) World Congress and the European Society for Fuzzy Logic and Technology (EUSFLAT) Conference*, edited by J. Carvalho, D. Dubois, U. Kaymak, and J. Sousa, 294–299. Lisbon, Portugal.

Fan, T., C. Liau, and D. Liu. 2011. "Dominance-based Fuzzy Rough Set Analysis of Uncertain and Possibilistic Data Tables." *International Journal of Approximate Reasoning* 52 : 1283–1297.

Fan, T., C. Liau, and D. Liu. 2011b. "Dominance-Based Rough Set Approach for Possibilistic Information systems." In *Proceedings of the 13th International Conference on Rough Sets, Fuzzy Sets, Data Mining and Granular Computing*. LNCS. Vol. 6743, edited by S. O. Kuznetsov, D. Slezak, D. H. Hepting, and B. Mirkin, 119–126. Moscow, Russia: Springer-Verlag.

Glöckner, I. 2004. "Evaluation of Quantified Propositions in Generalized Models of Fuzzy Quantification." *International Journal of Approximate Reasoning* 37 (2): 93–126.

Greco, S., B. Matarazzo, and R. Słowiński. 2001. "Rough Set Theory for Multicriteria Decision Analysis." *European Journal of Operational Research* 129 (1): 1–47.

Greco, S., B. Matarazzo, and R. Słowiński. 2002. "Rough Sets Methodology for Sorting Problems in Presence of Multiple Attributes and Criteria." *European Journal of Operational Research* 138 (2): 247–259.

Greco, S., B. Matarazzo, and R. Słowiński. 2004. "Axiomatic Characterization of a General Utility Ffunction and its Particular Cases in Terms of Conjoint Measurement and Rough-set Decision Rules." *European Journal of Operational Research* 158 (2): 271–292.

Greco, S., B. Matarazzo, and R. Słowiński. 2007. "Dominance-based Rough Set Approach as a Proper Way of Handling Graduality in Rough Set Theory." *Transactions on Rough Sets* VII : 36–52.

Greco, S., B. Matarazzo, R. Słowiński, and J. Stefanowski. 2000. "Variable Consistency Model of Dominance Based Rough Sets Approach." In *Proceedings of the 2nd International Conference on Rough Sets and Current Trends in Computing*, LNAI 2005, edited by W. Ziarko, and Y. Y. Yao, 170–181. Banff, Canada: Springer-Verlag.

Hájek, P. 1998. *Metamathematics of Fuzzy Logic*. Kluwer Academic Publishers.

Inuiguchi, M. 2009. "Rough Set Approach to Rule Induction from Imprecise Decision Tables." *Proceedings of the 8th International Workshop on Fuzzy Logic and Applications*, LNAI, Vol. 5571, edited by V. Gesú, S. Pal, and A. Petrosino, 68–76. Palermo, Italy: Springer-Verlag.

Inuiguchi, M., Y. Yoshioka, and Y. Kusunoki. 2009. "Variable-precision dominance based rough set approach and attribute reduction." *International Journal of Approximate Reasoning* 50 (8): 1199–1214.

Kryszkiewicz, M. 1998. "Properties of Incomplete Information Systems in the Framework of Rough Sets." In *Rough Sets in Knowledge Discovery*, edited by L. Polkowski, and A. Skowron, 422–450. Heidelberg: Physica-Verlag.

Lipski, W. 1981. "On databases with incomplete information." *Journal of the ACM* 28 (1): 41–70.

Myszkorowski, K. 2011. "Multiargument Relationships in Fuzzy Databases with Attributes Represented by Interval-Valued Possibility Distributions." In *Proceedings of the 13th International Conference on Rough Sets, Fuzzy Sets, DataMining and Granular Computing*. LNCS, Vol. 6743, edited by S. Kuznetsov, D. Slezak, D. Hepting, and B. Mirkin, 199–206. Moscow, Russia: Springer-Verlag.

Nguyen, H., and D. Slezak. 1999. "Approximate Reducts and Association Rules-Correspondence and Complexity Results." In *Proceedings of the 7th International Workshop on Rough Sets, Data Mining, and Granular-Soft Computing*. LNCS, Vol. 1711, edited by N. Zhong, A. Skowron, and S. Ohsuga, 137–145. Yamaguchi, Japan: Springer-Verlag.

Pawlak, Z. 1982. "Rough Sets." *International Journal of Computer and Information Sciences* 11 (15): 341–356.

Pawlak, Z. 1991. *Rough Sets – Theoretical Aspects of Reasoning about Data*. Dordrecht: Kluwer Academic Publishers.

Radzikowska, A., and E. Kerre. 2002. "A Comparative Study of Fuzzy Rough Sets." *Fuzzy Sets and Systems* 126 (2): 137–155.

Sakai, H., M. Nakata, and Slezak, D. 2011. "A Prototype System for Rule Generation in Lipskis Incomplete Information Databases." In *Proceedings of the 13th International Conference on Rough Sets, Fuzzy Sets, DataMining and Granular Computing*. LNCS, Vol. 6743, edited by Kuznetsov, S., D. Slezak, D. Hepting, and B. Mirkin, 175–182. Moscow, Russia: Springer-Verlag.

Slezak, D. 2002. "Approximate Entropy Reducts." *Fundamenta Informaticae* 53 (3–4): 365–390.

Slezak, D. 2005. "Rough sets and Bayes factor." *Transactions on Rough Sets* III : 202–229.

Slezak, D., and W. Ziarko. 2005. "The Investigation of the Bayesian Rough Set Model." *International Journal of Approximate Reasoning* 40 (1–2): 81–91.

Słowiński, R., S. Greco, and B. Matarazzo. 2002. "Rough Set Analysis of Preferenceordered Data." In *Proceedings of the 3rd International Conference on Rough Sets and Current Trends in Computing*. LNAI, Vol. 2475, edited by Alpigini, J., J. Peters, A. Skowron, and N. Zhong, 44–59. Malvern, PA, USA: Springer-Verlag.

Susmaga, R., R. Słowiński, S. Greco, and B. Matarazzo. 2000. "Generation of Reducts and Rules in Multi-attribute and Multi-criteria Classification." *Control and Cybernetics* 29 (4): 970–988.

Yao, Y., and Q. Liu. 1999. "A Generalized Decision Logic in Interval-set-valued Information Tables." In *New Directions in Rough Sets, Data Mining, and Granular-Soft Computing*. LNAI, Vol. 1711, edited by Zhong, N., A. Skowron, and S. Ohsuga, 285–293. Yamaguchi, Japan: Springer-Verlag.

Zadeh, L. 1975. "The Concept of a Linguistic Variable and its Applications in Approximate Reasoning." *Information Sciences* 8 : 199–251.

Zadeh, L. 1979. "A Theory of Approximate Reasoning." In *Machine Intelligence*. Vol. 9, edited by Hayes, J., D. Mitchie, and L. Mikulich, 149–194. New York: Halstead Press.

Zadeh, L. 1983. "A Computational Approach to Fuzzy Quantifiers in Natural Languages." *Computers and Mathematics with Applications* 9 (1): 149–184.