Valued Tolerance and Decision Rules.

Jerzy Stefanowski[1] and Alexis Tsoukiàs[2]

[1] Institute of Computing Science
Poznan University of Technology, 60-965, Poznan, Poland
Jerzy.Stefanowski@cs.put.poznan.pl
[2] LAMSADE - CNRS, Université Paris Dauphine
75775 Paris Cedex 16, France
tsoukias@lamsade.dauphine.fr

Abstract. In this paper we introduce the concept of valued tolerance as an extension of the usual concept of indiscernibility (which is a crisp equivalence relation) in rough sets theory. Some specific properties of the approach are discussed in the paper. Further on the problem of inducing rules from such a knowledge is addressed. In this case a "credibility degree" is associated to each rule. Properties of such a degree are analysed in the paper and its use in classification problems is discussed.

Keywords: valued tolerance, rough sets, fuzzy sets, decision rules, classification.

1 Introduction

Rough sets were introduced by Pawlak [Pawlak 1982,1991] to deal with a vague description of objects. They operate on an information table composed by a set of objects $A = \{x_1, \cdots x_n\}$ and a set of attributes $C = \{c_1, \cdots c_m\}$. The starting point of the rough sets theory is an observation that objects having the same description are *indiscernible* (similar) with respect to the available information. The *indiscernibility* relation between objects is a mathematical basis of the rough sets theory. As a consequence of inconsistent indiscernible objects it may not be possible to specify a set of objects in terms of the available information. Therefore, the concept of *rough set* is introduced, characterised by a pair of two ordinary sets denoted *lower* and *upper approximations*. Next basic notions are: approximation of classification of objects from A, reduction of attributes from C and decision rules derived from proper approximations of subsets identified with decision classes.

Although original rough sets theory has been used to face several problems (see e.g. review in [Komorowski et al., 1999]), the use of the indiscernibility relation may be too rigid in some real situations. Therefore several generalisations of the rough sets theory have been proposed. Some of them ([Dubois and Prade, 1990,Dubois and Prade, 1992,Słowiński and Stefanowski, 1996,Yao, 1996]) extend the basic idea to a fuzzy context while others use more general similarity or tolerance relations instead of classical indiscernibility relation (see e.g. [Skowron and Stepaniuk, 1996,Stepaniuk, 1996], [Słowiński and Vanderpooten, 1997,Słowiński and Vanderpooten, 2000]). There are also combinations of both extensions by [Greco et al., 1998] and [Greco et al., 2000] where lower and upper approximations are fuzzy sets based on a fuzzy similarity relation. Properties of extended binary relations were studied in [Yao and Wang, 1999].

In this paper we introduce the concept of valued tolerance relation as a new extension of rough sets theory. A functional extension of the concepts of upper and lower approximation is introduced so that to any subset of the universe a degree of lower (upper) approximability can be associated. In other terms, any subset of the universe can be lower (upper) approximation of a given set, but to a different degree. Further on, such a functional extension enables to compute a credibility degree for any rule induced from the input information table. Such an idea first appeared in our previous work on incomplete information tables [Stefanowski and Tsoukiàs, 1999].

The paper is organised as follows. In section 2, we discuss motivations for using valued tolerance relation. Then, in section 3 we introduce formally the concept of valued tolerance. Some specific properties of this approach are also examined. In section 4, problems of inducing decision rules and computing credibility of rule are discussed. All considerations are illustrated by a didactic example. Results are summarised in section 5.

2 Why valued tolerance?

Conventional rough sets theory is based on the crucial concept of indiscernibility relation which is a crisp equivalence relation (complete, reflexive, symmetric and transitive relation valued in $\{0,1\}$). Practically speaking, two objects, described under a set of attributes, are indiscernible iff they have identical values.

However, real life suggests that this is a very strong assumption. Objects may be practically indiscernible without having identical values. The idea of substituting indiscernibility with a wider concept of similarity has already been studied in e.g. [Słowiński and Vanderpooten, 1997] or [Skowron and Stepaniuk 1995]. Moreover, it could be the case that objects can be "more or less similar" depending on the particular information available. Consider the following two examples.

Example 1. Three objects x_1, x_2, x_3 and four attributes c_1, c_2, c_3, c_4 are given, each attribute equipped with a discrete nominal scale A,B,C,D. Besides the following information table is provided:

	c_1	c_2	c_3	c_4
x_1	A	B	B	C
x_2	A	B	*	C
x_3	A	*	*	*

* representing the "unknown" value of attribute.

If any similarity is to be considered among the three objects it is easy to suggest that "is more possible that x_2 is similar to x_1 than x_3 to x_1". Conventional rough sets theory simply does not apply in this case and its usual extensions handling unknown values will consider the three objects as completely different or totally identical ([Grzymala, 1991,Kryszkiewicz, 1995,Kryszkiewicz, 1998]). However, being able to give a value to the possibility that objects are similar could open interesting operational directions (see [Stefanowski and Tsoukiàs, 1999]).

Example 2. Three objects x_1, x_2, x_3 and four attributes c_1, c_2, c_3, c_4 are given, each attribute equipped with an interval scale in the interval [0,100]. Besides, the following information table is provided:

	c_1	c_2	c_3	c_4
x_1	90	20	50	80
x_2	91	21	51	81
x_3	85	15	45	75

If any similarity is to be considered the reader might agree that is reasonable to consider that "is more possible that x_2 is similar to x_1 than x_3 to x_1", while x_2 and x_3 are not similar at all. This is an effect of the existence of a discrimination threshold. In such a model (see [Luce, 1956]) objects are different only if they have a difference of more than the established threshold (in the example such a threshold is 5). However, the threshold by itself does not solve the problem, for the same discrimination problem could be considered near the threshold: i.e. why a difference of 4 is not significant and a difference of 5 it is? It is more natural to consider that the possibility that two objects are similar decreases as the difference of value of the two objects increases. The use of a valued tolerance appears to be more appropriate in this case also.

From the above considerations it is clear (for us) the opportunity of introducing a valued tolerance when comparing objects, whatever the purpose of the comparison is. Our aim is to introduce such a concept in the domain of classification based on rough sets theory. In the following we will restrict ourselves to symmetric (possibly valued) similarity relations which we denote as (possibly valued) tolerance relations.

3 Rough sets and valued tolerance

In the following we introduce an approach already discussed for handling incomplete information tables in [Stefanowski and Tsoukiàs, 1999].

Given a valued tolerance relation on the set A we can define a "tolerance class", that is a fuzzy set with membership function the "possibility of tolerance" to a reference object $x \in A$.

The problem is to define the concepts of upper and lower approximation of a set Φ. The approach we will adopt in this paper considers, coherently with the rest, approximation as a *continuous valuation*. Given a set Φ to describe and a set $Z \subseteq A$ we will try to define the degree by which Z approximates from the top or from the bottom the set Φ. Technically we will try to give the functional correspondent of the concepts of lower and upper approximation.
[Dubois and Prade, 1990,Dubois and Prade, 1992,Greco et al., 1998,Greco et al., 2000] have similar concerns and explored the idea of combining fuzzy and rough sets, but under our perspective lower and upper approximations are not fuzzy sets to which elements from the universe of discourse may more or less belong. Each subset of A may be a lower or upper approximation of Φ, but to different degrees (such an approach has been inspired by the work of Kitainik [Kitainik, 1993]).

For this purpose we need to translate in a functional representation the usual logical connectives of negation, conjunction etc. ($x, y, z \cdots$ represent in the following membership degrees).

1. A negation is a function $N : [0,1] \mapsto [0,1]$, such that $N(0) = 1$ and $N(1) = 0$. An usual representation of the negation is $N(x) = 1 - x$.
2. A T-norm is a continuous, non decreasing function $T : [0,1]^2 \mapsto [0,1]$ such that $T(x,1) = x$. Clearly a T-norm stands for a conjunction. Usual representations of T-norms are:
   - the min: $T(x,y) = \min(x,y)$;
   - the product: $T(x,y) = xy$;
   - the Łukasiewicz T-norm: $T(x,y) = \max(x+y-1,0)$.
3. A T-conorm is a continuous, non decreasing function $S : [0,1]^2 \mapsto [0,1]$ such that $S(0,y) = y$. Clearly a T-conorm stands for a disjunction. Usual representations of T-conorms are:
   - the max: $S(x,y) = \max(x,y)$;
   - the product: $S(x,y) = x + y - xy$;
   - the Łukasiewicz T-conorm: $S(x,y) = \min(x+y,1)$.

If $S(x,y) = N(T(N(x),N(y)))$ we have the equivalent of the De Morgan law and we call the triplet $\langle N,T,S \rangle$ a De Morgan triplet. $I(x,y)$, the degree by which x may imply y is again a function $I : [0,1]^2 \mapsto [0,1]$. However, the definition of the properties that such a function may satisfy do not make the unanimity. Two basic properties may be desired:
- the first claiming that $I(x,y) = S(N(x),y)$ translating the logical equivalence $x \rightarrow y =_{def} \neg x \vee y$;
- the second claiming that whenever the membership degree x is not greater than the membership degree y, then the implication should be true ($x \leq y \Leftrightarrow I(x,y) = 1$).

It is almost impossible to satisfy both the two properties. In the very few cases where this happens other properties are not satisfied (for an excellent discussion see [Dubois et al., 1991]).

Coming back to our lower and upper approximations we know that, given a set $Z \subseteq A$, a subset of attributes $B \subseteq C$ and a set Φ, the usual definitions are:

1. $Z = \Phi_B \;\Leftrightarrow\; \forall\, z \in Z,\;\; \Theta_B(z) \subseteq \Phi$
2. $Z = \Phi^B \;\Leftrightarrow\; \forall\, z \in Z,\;\; \Theta_B(z) \cap \Phi \neq \emptyset$

where $\Theta_B(z)$ is the tolerance class of element z created on the basis of the subset of attributes B. The functional translation of such definitions is straightforward. Considering that,
$\forall\, x\; \phi(x) =_{def} T_x\phi(x)$;
$\exists\, x\; \phi(x) =_{def} S_x\phi(x)$;
$\Phi \subseteq \Psi =_{def} T_x(I(\mu_\Phi(x), \mu_\Psi(x)))$;
$\Phi \cap \Psi \neq \emptyset =_{def} \exists\, x\; \phi(x) \wedge \psi(x) =_{def} S_x(T(\mu_\Phi(x), \mu_\Psi(x)))$ we get:

1. $\mu_{\Phi_B}(Z) \;=\; T_{z \in Z}(T_{x \in \Theta_B(z)}(I(R_B(z,x), \hat{x})))$.
2. $\mu_{\Phi^B}(Z) \;=\; T_{z \in Z}(S_{x \in \Theta_B(z)}(T(R_B(z,x), \hat{x})))$.

where:
$\mu_{\Phi_B}(Z)$ is the degree for set Z to be a B-lower approximation of Φ;
$\mu_{\Phi^B}(Z)$ is the degree for set Z to be a B-upper approximation of Φ;
$\Theta_B(z)$ is the tolerance class of element z;
T, S, I are the functions previously defined; as far as $I(x,y)$ is concerned we will always choose to satisfy De Morgan law ($I(x,y) = S(N(x), y)$). This is due to the particular case of $\mu_{\Phi_B}(Z)$ where $\hat{x} \in \{0,1\}$. If we choose any other representation then lower approximability collapse to $\{0,1\}$.
$R_B(z,x)$ is the membership degree of element x in the tolerance class of z (at the same time is the valued tolerance relation between elements x and z for attribute set B; in our case $R_B(z,x) = T_{j \in B} R_j(z,x)$);
$\hat{x}$ is the membership degree of element x in the set Φ ($\hat{x} \in \{0,1\}$).

In the following we provide some formal properties that such an approach fulfill. The reader should remark that in the following Φ^c denotes the complement of set Φ with respect to the universe.

Proposition 1. *If T, S, I fulfill the De Morgan law and R_B is a valued tolerance relation then* $\forall Z \in A\;\; \mu_{\Phi_B}(Z) \leq \mu_{\Phi^B}(Z)$.

Proof.
In order to demonstrate the proposition we first observe that both the lower and the upper approximability are T-norms on the same set Z. Therefore it is sufficient to demonstrate that $\forall z \in Z$ the argument of the T-norm defining the lower approximability is less or equal to the argument of the T-norm defining the upper approximability. Thus we have to demonstrate that: $T_{x \in \Theta_B(z)}(S(1 - R_B(z,x), \hat{x})) \leq S_{x \in \Theta_B(z)}(T(R_B(z,x), \hat{x}))$

Since min is the largest T-norm and max is the smallest T-conorm is sufficient to demonstrate that: $\min_{x \in \Theta_B(z)}(\max(1 - R_B(z,x), \hat{x})) \leq \max_{x \in \Theta_B(z)}(\min(R_B(z,x), \hat{x}))$.

We distinguish two cases:

1. Consider $z = x_k \in \Phi^c$. Then $\hat{x_k} = 0$. Therefore when $x = x_k$ we have $\max(1 - R_B(x_k, x_k), \hat{x_k}) = 0$ so that $\min_{x \in \Theta_B(z)}(\max(1 - R_B(z,x), \hat{x})) = 0$. At the same time:
   $\forall x \in \Phi^c\; \hat{x} = 0$ and $\min(R_B(z,x), \hat{x}) = 0$ and
   $\forall x \in \Phi\; \hat{x} = 1$ and $\min(R_B(z,x), \hat{x}) = R_B(z,x)$
   so that $\max_{x \in \Theta_B(z)}(\min(R_B(z,x), \hat{x})) \;=\; \max_{x \in \Phi}(R_B(z,x)) \geq 0$. Therefore if $z \in \Phi^c$ we get $\mu_{\Phi_B}(z) = 0 \leq \mu_{\Phi^B}(z)$.

2. Consider $z = x_k \in \Phi$. Then $\hat{x_k} = 1$. Therefore when $x = x_k$ we have $\min(R_B(x_k, x_k), \hat{x_k}) = 1$ so that $\max_{x \in \Theta_B(z)}(\min(R_B(z,x), \hat{x})) = 1$. At the same time:
$\forall x \in \Phi\ \hat{x} = 1$ and $\max(1 - R_B(z,x), \hat{x}) = 1$ and
$\forall x \in \Phi^c\ \hat{x} = 0$ and $\max(1 - R_B(z,x), \hat{x}) = 1 - R_B(z,x) \leq 1$
so that $\min_{x \in \Theta_B(z)}(\max(1 - R_B(z,x), \hat{x})) = \min_{x \in \Phi^c}(1 - R_B(z,x)) \leq 1$. Therefore if $z \in \Phi$ we get $\mu_{\Phi_B}(z) \leq 1 = \mu_{\Phi^B}(z)$.

And this completes the proof.

■

We immediately obtain the following corollary.

Corollary 1.
If $z \in \Phi$ *then* $\mu_{\Phi_B}(z) = \min_{x \in \Phi^c}(1 - R_B(z,x)) \leq \mu_{\Phi^B(z)} = 1$
If $z \in \Phi^c$ *then* $\mu_{\Phi_B}(z) = 0 \leq \mu_{\Phi^B(z)} = \max_{x \in \Phi}(R_B(z,x))$

Proposition 2. *If* T, S, I *respect the De Morgan law and* R_B *is a valued tolerance relation then* $\forall z\ \mu_{\Phi_B}(z) = 1 - \mu_{(\Phi^c)^B}(z)$.

Proof.
Denote by $\hat{x^c}$ the membership of x to Φ^c. Clearly $\hat{x^c} = 1 - \hat{x}$. We then have:
$\mu_{\Phi_B}(z) = T_{x \in \Theta_B(z)}(S(1 - R_B(z,x), \hat{x})) =$
$T_{x \in \Theta_B(z)}(S(1 - R_B(z,x), 1 - \hat{x^c})) =$
$T_{x \in \Theta_B(z)}(1 - T(R_B(z,x), \hat{x^c})) =$
$1 - S_{x \in \Theta_B(z)}(T(R_B(z,x), \hat{x^c})) =$
$1 - \mu_{(\Phi^c)^B}(z)$.

■

We immediately obtain the following corollary.

Corollary 2.
$\mu_{\Phi_B}(Z) = 1 - S_{z \in Z}(\mu_{(\Phi^c)^B}(z)$
$\mu_{\Phi^B}(Z) = 1 - S_{z \in Z}(\mu_{(\Phi^c)_B}(z)$

The practical consequence of the above result is that in order to compute the lower (upper) approximability of any subset $\Phi \subseteq A$ is sufficient to compute the upper (lower) approximability of each single element of A of the sets Φ and Φ^c. Operationally we can fix a threshold k (l) for the lower (upper) approximability and then add elements to the empty set by decreasing order of their lower (upper) approximability.

Finally we can show the following result.

Proposition 3. $\forall Z \subset A\ \ \hat{B} \subset B \Rightarrow \mu_{\Phi_{\hat{B}}}(Z) \leq \mu_{\Phi_B}(Z)$

Proof.
Since $R_B(z,x) = T_{j \in B} R_j(z,x)$, if $\hat{B} \subseteq B$ then $\forall x\ R_{\hat{B}}(z,x) \geq R_B(z,x)$ and therefore $\forall x\ 1 - R_{\hat{B}}(z,x) \leq 1 - R_B(z,x)$. Then by definition of lower approximability the proposition holds.

■

In order to improve comprehension of our approach consider the following example (Φ and Ψ being two distinct classes to which elements of A may belong, the one being the complement of the other).

Example 3. A set of 12 objects $x_1, \cdots x_{12}$, four attributes c_1, c_2, c_3, c_4 and a decision attribute d are given, each attribute equipped with an interval scale in the interval [0,100]. Besides, the following decision table is provided:

	c_1	c_2	c_3	c_4	d
x_1	89	91	95	87	Φ
x_2	85	87	90	86	Φ
x_3	80	83	92	80	Ψ
x_4	79	80	85	77	Φ
x_5	74	76	83	71	Ψ
x_6	70	71	79	70	Ψ
x_7	68	70	74	66	Φ
x_8	63	64	70	62	Ψ
x_9	59	64	69	60	Ψ
x_{10}	57	59	57	56	Φ
x_{11}	55	56	56	54	Ψ
x_{12}	25	32	15	48	Φ

Table 1: Decision table for valued tolerance

A constant threshold of $k = 10$ applies in order to consider two objects as surely not different (similar) to any of the four attributes.

Following an approach recently introduced ([Tsoukiàs and Vincke, 2000,Tsoukiàs, 2000]) we consider for each attribute c_j a valued tolerance relation as follows:

$$R_j(x,y) = \frac{\max(0, \min(c_j(x), c_j(y)) + k - \max(c_i(x), c_j(y)))}{k}$$

where: k is the discrimination threshold. It is easy to observe that:

$\forall x, y \in A \;\; R_j(x,y) = 1 \quad$ iff $c_j(x) = c_j(y)$
$\forall x, y \in A \;\; R_j(x,y) \in]0,1[$ iff $| c_j(x) - c_j(y) | < k$
$\forall x, y \in A \;\; R_j(x,y) = 0 \quad$ iff $| c_j(x) - c_j(y) | \geq k$

Actually the valued tolerance defined above considers that when the values of two objects are identical the objects are necessarily similar. If the difference of value is more than the threshold k then the two objects are necessarily not similar. If the difference of value is between 0 and k then the valued tolerance decreases linearly as the difference increases. If $R_j(x,y)$ represents the necessity that x is similar to y then, in presence of several attributes, a way to evaluate the necessity that x is comprehensively similar to y is to take the T-norm of the different similarities. Coherently with the necessity approach we get: $R(x,y) = \min_j(R_j(x,y))$. Applying this formula to the information in Table 1 we get the following comprehensive valued relation on the set A. Clearly R is a valued tolerance.

	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}	x_{12}
x_1	1	0.5	0.1	0	0	0	0	0	0	0	0	0
x_2	0.5	1	0.4	0.1	0	0	0	0	0	0	0	0
x_3	0.1	0.4	1	0.3	0.1	0	0	0	0	0	0	0
x_4	0	0.1	0.3	1	0.4	0.1	0	0	0	0	0	0
x_5	0	0	0.1	0.4	1	0.5	0.1	0	0	0	0	0
x_6	0	0	0	0.1	0.5	1	0.5	0.1	0	0	0	0
x_7	0	0	0	0	0.1	0.5	1	0.4	0.1	0	0	0
x_8	0	0	0	0	0	0.1	0.4	1	0.6	0	0	0
x_9	0	0	0	0	0	0	0.1	0.6	1	0	0	0
x_{10}	0	0	0	0	0	0	0	0	0	1	0.7	0
x_{11}	0	0	0	0	0	0	0	0	0	0.7	1	0
x_{12}	0	0	0	0	0	0	0	0	0	0	0	1

Table 2: Valued tolerance relation for example 3

Using the above information we can compute the lower and upper approximability for each element of set A as results in Table 3.

	$\mu_{\Phi_B}(x)$	$\mu_{\Phi^B}(x)$	$\mu_{\Psi_B}(x)$	$\mu_{\Psi^B}(x)$
x_1	0.9	1	0	0.1
x_2	0.6	1	0	0.4
x_3	0	0.4	0.6	1
x_4	0.6	1	0	0.4
x_5	0	0.4	0.6	1
x_6	0	0.5	0.5	1
x_7	0.5	1	0	0.5
x_8	0	0.5	0.5	1
x_9	0	0.2	0.8	1
x_{10}	0.3	1	0	0.7
x_{11}	0	0.7	0.3	1
x_{12}	1	1	0	0

Table 3: Lower and upper approximability of each element of A.

4 Decision Rules

In order to induce classification rules from the decision table on hand we may accept now rules with a "credibility degree" derived from the fact that objects may be similar to the conditional part of the rule only to a certain degree, besides the fact the implication in the decision part is also uncertain. More formally we give the following representation for a rule ρ_i:

$$\rho_i =_{def} \bigwedge_{c_j \in B} (c_j(x) = v) \rightarrow (d = \phi)$$

where: $B \subseteq C$, v is the value of conditional attribute c_j, ϕ is the value of decision attribute d.

We may use the valued relation $s_B(x, \rho_i)$ in order to indicate that element x "supports" rule ρ_i or that, x is similar to some extend to the conditional part of rule ρ_i on attributes B. The relation s is a valued tolerance relation defined exactly as relation R. We denote as $S(\rho_i) = \{x : s_B(x, \rho_i) > 0\}$ and as $\Phi = \{x : d(x) = \phi\}$. In a case of crisp relation ρ_i is a classification rule iff:

$$\forall\, x \in S(\rho_i) : \; \Theta_B(x) \subseteq \Phi$$

Shifting in the valued case we can compute a credibility degree for any rule ρ_i calculating the credibility for the previous formula which can be rewritten as: $\forall\, x, y\; s_B(x, \rho_i) \rightarrow (R_B(x, y) \rightarrow \Phi(y))$. We get:

$$\mu(\rho_i) = T_{x \in S(\rho_i)}(I(s_B(x, \rho_i), T_{y \in \Theta_B(x)}(I(\mu_{\Theta_B(x)}(y), \mu_\Phi(y)))))$$

where: $\mu_{\Theta_B(x)}(y) = R_B(x, y)$ and $\mu_\Phi(y) \in \{0, 1\}$.

Finally it is necessary to check whether B is a non-redundant set of conditions for rule ρ_i, i.e. to look if it is possible to satisfy the condition: $\exists\, \hat{B} \subset B : \; \mu(\rho_i^{\hat{B}}) \geq \mu(\rho_i^B)$. We can equivalently state that if there is no $\hat{B}$ satisfying the condition then B is a "non redundant" set of attributes for rule ρ_i.

Before we continue the presentation of our approach in rule induction, is important to state the following result.

Proposition 4. *Consider a rule ρ_i classifying objects to a set $\Phi \subseteq A$ under a set of attributes B. If T, S, I satisfy the De Morgan law and R_B is a valued tolerance, the credibility $\mu(\rho_i)$ of the rule is upper bounded by the lower approximability of set Φ by the element x_k whose description (under attributes B) coincides with the conditional part of the rule.*

Proof.
Consider the definition of rule credibility. We know that:
$T_{y \in \Theta_B(x)}(I(\mu_{\Theta_B(x)}(y), \mu_\Phi(y))) = \mu_{\Phi_B}(x)$. Therefore considering that $I(x,y) = S(N(x), y)$ and that $s_B = R_B$ we can rewrite:

$$\mu(\rho_i) = T_{x \in S(\rho_i)}(S(1 - R_B(x, \rho_i), \mu_{\Phi_B}(x)))$$

We distinguish four cases.

1. It exists an element $x_k \in A$ whose description (under attributes B) coincides with the conditional part of rule ρ_i. We have $R_B(x, \rho_i) = 1$. Therefore $S(1 - R_B(x, \rho_i), \mu_{\Phi_B}(x)) = \mu_{\Phi_B}(x)$ in this case.
2. For all x for which $R_B(x, \rho_i) = 0$ we get $S(1 - R_B(x, \rho_i), \mu_{\Phi_B}(x)) = 1$.
3. For all x for which $1 - R_B(x, \rho_i) > \mu_{\Phi_B}(x)$ we get $S(1 - R_B(x, \rho_i), \mu_{\Phi_B}(x)) \geq 1 - R_B(x, \rho_i)$ since max is the smallest T-conorm.
4. For all x for which $1 - R_B(x, \rho_i) < \mu_{\Phi_B}(x)$ we get $S(1 - R_B(x, \rho_i), \mu_{\Phi_B}(x)) \geq \mu_{\Phi_B}(x)$ since max is the smallest T-conorm.

Denoting x_k, x_l, x_i, x_j the x for the four cases respectively we obtain:

$$\mu(\rho_i) = T_{x \in S(\rho_i)}(\mu_{\Phi_B}(x_k), \{\forall x_l : 1\}, \{\forall x_i : 1 - R_B(x_i, \rho_i)\}, \{\forall x_j : \mu_{\Phi_B}(x_j)\})$$

And since by definition we have $T(x,y) \leq \min(x,y) \leq x$, $\mu_{\Phi_B}(x_k)$ is an upper bound for $\mu(\rho_i)$.
■

Operationally, the user should fix a credibility threshold for the induced rules in order to prevent proliferation of rules considered as "unsafe" for the classification purposes. A sensitivity analysis could be performed around such a threshold to find accepted rules.

In general, elementary conditions of the induced rules are created using the description of objects in the decision table. Assuming that the user has defined a credibility threshold at level λ, it is possible to use the result of Proposition 4.1. to induce decision rules, i.e.

1. When choosing objects as candidates for inducing a classification rules for class Φ, it is sufficient to choose only objects with lower approximability of a set Φ not worse than λ. Other objects could be skipped.
2. Further on it is necessary to search for the non-reduced sets of conditions (in general it corresponds to the problem of looking for local reducts [Komorowski et al., 1999])
3. Given credibility threshold λ one can relax the previous requirements to non-redundant rule, i.e. it may be accepted that from rule ρ_i with credibility $\mu(\rho_i)$ new rules could be generated with shortest condition part but with lower credibility however still over the allowed threshold.

Let us notice that the problem of inducing all rules with accepted credibility from examples in the information table has exponential complexity in the worst case (see corresponding problems of looking for all local reducts in the classical rough sets theory [Komorowski et al., 1999]). However, fixing sufficient high value of credibility threshold may reduce the search space.

Continuation of Example 3: Consider again the example of information table used in this paper. Coherently with our previous computations we choose min as the T-norm and max as T-conorm.

When choosing objects as candidates for inducing a classification rule for class Φ or Ψ it is natural to choose objects with the maximal lower approximability of a set. Fixing a high credibility threshold we can look for rules describing class Φ on the basis of object x_{12} only (see table 3.). So, let us fix the value of credibility threshold to 0.6. For the set of all attributes C and class Φ, objects x_1, x_2, x_4, x_{12} could be taken into account.

Let us consider candidate for a decision rule based of description of the object x_1, i.e. $\rho_1 : (c_1 = 89) \wedge (c_2 = 91) \wedge (c_3 = 95) \wedge (c_4 = 87) \rightarrow (d = \Phi)$. Three objects (x_1, x_2, x_3) are similar to its condition part with $R_C(x_1, \rho_i)$=1, $R_C(x_2, \rho_i)$=0.5 and $R_C(x_3, \rho_i)$=0.1. So $S(\rho_i) = \{x_1, x_2, x_3\}$. We can compute credibility of the rule according to formula: $\mu(\rho_i) = \min_{x \in S(\rho_i)}(\max(1 - R_B(x, \rho_i), \mu_{\Phi_B}(x)))$. So, taking values of lower approximability (see Table 3.) we have $\mu(\rho_i) = \min(\max(1-1, 0.9), \max(1-0.5, 0.6), \max(1 - 0.1, 0)) = 0.6$. However, this is a rule with a redundant set of conditions. As one can check, it can be reduced to the much simpler form
$\rho_1 : (c_4 = 87) \rightarrow (d = \Phi)$
which is still supported by objects $S(\rho_i) = \{x_1, x_2, x_3\}$ with $R_{c4}(x_1, \rho_i)$=1, $R_{c4}(x_2, \rho_i)$=0.9 and $R_{c4}(x_3, \rho_i)$=0.3. It is necessary to compute the new lower approximability, $\mu_{c4}(x_1) = \min(\max(1 - 1, 1), \max(1 - 0.9, 1), \max(1 - 0.3, 0) = 0.7$, $\mu_{c4}(x_2)$=0.6 and $\mu_{c4}(x_3)$=0. The credibility degree is still $\mu(\rho_i) = \min(\max(1 - 1, 0.7), \max(1 - 0.9, 0.6), \max(1 - 0.3, 0)) = 0.6$.

Proceeding in a similar way we can induce the other decision rules for class Φ:
$\rho_2 : (c_4 = 86) \rightarrow (d = \Phi)$ with $\mu(\rho_2) = 0.6$ and $S(\rho_2) = \{x_1, x_2, x_3, x_4\}$
$\rho_3 : (c_3 = 85) \wedge (c_4 = 77) \rightarrow (d = \Phi)$ with $\mu(\rho_3) = 0.6$ and $S(\rho_3) = \{x_2, x_3, x_4, x_5, x_6\}$
$\rho_4 : (c_1 = 25) \rightarrow (d = \Phi)$ with $\mu(\rho_4) = 1.0$ and $S(\rho_4) = \{x_{12}\}$
$\rho_5 : (c_2 = 32) \rightarrow (d = \Phi)$ with $\mu(\rho_5) = 1.0$ and $S(\rho_5) = \{x_{12}\}$
$\rho_6 : (c_3 = 15) \rightarrow (d = \Phi)$ with $\mu(\rho_6) = 1.0$ and $S(\rho_6) = \{x_{12}\}$
$\rho_7 : (c_4 = 48) \rightarrow (d = \Phi)$ with $\mu(\rho_7) = 0.6$ and $S(\rho_7) = \{x_{11}, x_{12}\}$

Further on, for class Ψ the following rules could be induced:
$\rho_8 : (c_3 = 92) \wedge (c_4 = 80) \rightarrow (d = \Psi)$ with $\mu(\rho_8) = 0.6$ and $S(\rho_2) = \{x_1, x_2, x_3, x_4, x_5\}$
$\rho_9 : (c_1 = 74) \wedge (c_4 = 71) \rightarrow (d = \Psi)$ with $\mu(\rho_9) = 0.6$ and $S(\rho_9) = \{x_3, x_4, x_5, x_6, x_7\}$
$\rho_{10} : (c_3 = 83) \wedge (c_4 = 71) \rightarrow (d = \Psi)$ with $\mu(\rho_{10}) = 0.6$ and $S(\rho_{10}) = \{x_3, x_4, x_5, x_6, x_7\}$
$\rho_{11} : (c_2 = 76) \wedge (c_4 = 71) \rightarrow (d = \Psi)$ with $\mu(\rho_{11}) = 0.6$ and $S(\rho_{11}) = \{x_3, x_4, x_5, x_6, x_7\}$
$\rho_{12} : (c_2 = 64) \wedge (c_3 = 69) \rightarrow (d = \Psi)$ with $\mu(\rho_{12}) = 0.6$ and $S(\rho_{12}) = \{x_7, x_8, x_9, x_{10}\}$

Let us now consider the use of induced decision rules to classify new unclassified objects. The problem is to assign such objects to a-priori known sets (decision classes) on the basis of their tolerance to the conditional part of the already induced rules.

We have a double source of uncertainty. First, the new object will be similar to a certain extend to the conditional part of a given rule. Second the rule itself has a credibility (classification is not completely sure any more). In general a new unclassified object will be more or less similar to more than one decision rule and such rules may indicate different decision classes. Therefore an unclassified object can be assigned to several different classes. In order to choose one class the following procedure is proposed:

1. For each decision rules ρ_i in the set of induced rules (assigning objects to class Φ_i) we calculate the tolerance of the new object z to its condition part, $R_B(z, \rho_i)$.
2. Then we compute the membership degree of object z to each decision class Φ_i as $\mu_{\Phi_i}(z) = T(R_B(z, \rho_i), \mu(\rho_i))$. Then we choose the class with the maximum membership degree.
3. If a tie occurs (the same membership for different classes) choose the rule with the highest number of supporting objects $S(\rho_i)$.

Let us comment that if we compute the best matched rules independently for each considered decision class, we can show the user the distribution of possibilities of classifications z to different classes.

Continuation of Example 3: Consider again the example of information table and induced decision rules. Let us assume that the new unclassified object z has the following description $(c_1 = 28) \wedge (c_2 = 55) \wedge (c_3 = 78) \wedge (c_4 = 56)$. First its similarity to condition part of each decision rule should be calculated. One can check that the tolerance $R_B(z, \rho_i)$ is greater than zero only for three rules $\rho_4, \rho_7, \rho_{12}$, i.e. $R_B(z, \rho_4)$=0.7, $R_B(z, \rho_7)$=0.2 and $R_B(z, \rho_{12}) = \min(0.1, 0.1) = 0.1$. The two first rules indicate Φ with membership degrees $\mu_\Phi(z) = \min(0.7, 1.0) = 0.7$ and $\mu_\Phi(z) = \min(0.2, 0.6) = 0.2$, while the last indicates Ψ with degree $\mu_\Psi(z) = \min(0.1, 0.6) = 0.1$. Therefore object z should be assigned to class Φ.

Further on, let us consider a more difficult case, i.e. assume that object z has the following description $(c_1 = 85) \wedge (c_2 = 87) \wedge (c_3 = 91) \wedge (c_4 = 84)$. Four decision rules have a non-zero similarity with this object, i.e.
$R_B(z, \rho_1) = 0.7$ and $\mu_\Phi(z) = \min(0.7, 0.6) = 0.6$
$R_B(z, \rho_2) = 0.8$ and $\mu_\Phi(z) = \min(0.8, 0.6) = 0.6$
$R_B(z, \rho_3) = \min(0.4, 0.3) = 0.3$ and $\mu_\Phi(z) = \min(0.3, 0.6) = 0.3$
$R_B(z, \rho_8) = \min(0.9, 0.6) = 0.6$ and $\mu_\Psi(z) = \min(0.6, 0.6) = 0.6$.

In fact a tie occurs between rules ρ_1, ρ_2 and ρ_8. However the supports of these rules are the following: $S(\rho_1) = \{x_1, x_2, x_3\}$, $S(\rho_2) = \{x_1, x_2, x_3, x_4\}$, $S(\rho_8) = \{x_1, x_2, x_3, x_4, x_5\}$. According to the highest number of supporting objects, the rule ρ_8 should be chosen and object z could be assigned to the set Ψ.

5 Conclusion

In the paper we develop the idea that valued tolerance relations (symmetric valued similarity relations) can be more suitable when objects are compared for classification purposes. Particularly when rough sets are used, the classic indiscernibility relation (which is a crisp equivalence relation) can be a too strong assumption with respect to the available information.

The main contribution of the paper consists in considering that any subset of the universe of discourse can be considered as a lower (upper) approximation of set Φ, but to a different degree, due to the existence of a valued tolerance relation among the elements of the universe. A number of formal properties of this approach are demonstrated and discussed in the paper.

Further on, the availability of a lower (upper) approximability degree for each set with respect to a decision class Φ enables to compute classification rules equipped with a credibility degree. A significant result obtained in the paper consists in demonstrating that a rule credibility is upper bounded by the lower approximability of the set whose elements description coincides with the conditional part of the rule.

Besides increasing the number of induced rules for a given data set, the user has an improved toolkit in order to explore the security of any classification. In fact when new objects have to be classified using the induced rules two sources of uncertainty have to be considered: the first due to the fact the new object might be more or less similar to the conditional part of the rule and the second due to the fact that the rule itself is more or less safe (credible). A number of operational hints are presented in the paper in order to implement efficient rule induction algorithms and classification procedures.

References

[Dubois and Prade, 1990] Dubois D., Prade H.: Rough Fuzzy Sets and Fuzzy Rough Sets. International Journal of General Systems, **17**, (1990), 191–209.

[Dubois and Prade, 1992] Dubois D., Prade H.: Putting rough sets and fuzzy sets together. In Shi-Yu Huang (ed.), Intelligent Decicion Support, Kluwer Academic, Dordrecht, (1992), 203–232.

[Dubois et al., 1991] Dubois D., Lang J., Prade H.: Fuzzy sets in approximate reasoning. Fuzzy Sets and Systems, **40**, (1991), 203–244.

[Greco et al., 1998] Greco S., Matarazzo B. Slowinski R.: Fuzzy similarity relation as a basis for rough approximations. In Polkowski L., Skowron A. (eds.), Proc. of the First Int. Conf. on Rough Sets and Current Trends in Computing, Springer Verlag, Berlin, LNAI 1424, (1998), 283–289.

[Greco et al., 2000] Greco S., Matarazzo B. Slowinski R.: Rough set processing of vague information using fuzzy similarity relations. In Calude C.S., Paun G. (eds), Finite vs infinite: contributions to an eternal dilemma, Springer Verlag, Berlin, (2000), 149–173.

[Grzymala, 1991] Grzymala-Busse J.W.: On the unknown attribute values in learning from examples. Proc. of Int. Symp. on Methodologies for Intelligent Systems, (1991), 368–377.

[Kitainik, 1993] Kitainik L.: Fuzzy Decision Procedures with Binary Relations, Kluwer Academic, Dordrecht, (1993).

[Komorowski et al., 1999] Komorowski J., Pawlak Z., Polkowski L. Skowron A.: Rough Sets: a tutorial. In Pal S.K., Skowron A. (eds.), Rough Fuzzy Hybridization. A new trend in decision making, Springer Verlag, Berlin, (1999), 3–98.

[Kryszkiewicz, 1995] Kryszkiewicz M.: Rough set approach to incomplete information system. Information Sciences, **112**, (1998), 39–49.

[Kryszkiewicz, 1998] Kryszkiewicz M.: Properties of incomplete information systems in the framework of rough sets. In Polkowski L., Skowron A. (eds.), Rough Sets in Data Mining and Knowledge Discovery, Physica-Verlag, Heidelberg, (1998), 422–450.

[Luce, 1956] Luce R.D.: Semiorders and a theory of utility discrimination, Econometrica, **24**, (1956), 178–191.

[Skowron and Stepaniuk, 1996] Skowron A., Stepaniuk J.: Tolerance approximation spaces, Fundamenta Informaticae, **27**, (1996), 245–253.

[Słowiński and Stefanowski, 1996] Słowiński R., Stefanowski J.: Rough set reasoning about uncertain data. Fundamenta Informaticae, **27**, (1996), 229–244.

[Słowiński and Vanderpooten, 1997] Słowiński R., Vanderpooten D.: Similarity relation as a basis for rough approximations, In Wang P. (ed.), Advances in Machine Intelligence and Soft Computing, vol. IV., Duke University Press, (1997), 17–33.

[Słowiński and Vanderpooten, 2000] A generalized definition of rough approximations based on similarity. IEEE Transactions on Data and Knowledge Engineering, (2000), (to appear).

[Stefanowski and Tsoukiàs, 1999] Stefanowski J., Tsoukiàs A.: On the extension of rough sets under incomplete information, in N. Zhong, A. Skowron, S. Ohsuga, (eds.), New Directions in Rough Sets, Data Mining and Granular-Soft Computing, Springer Verlag, LNAI 1711, Berlin, (1999), 73–81.

[Stepaniuk, 1996] Stepaniuk J.: Similarity based rough sets and learning. In Tsumoto S. et al. (eds.), Proceedings of the fourth international workshop on rough sets, fuzzy sets and machine discovery, RSFD'96, (1996), 18–22.

[Tsoukiàs, 2000] Tsoukiàs A.: On the valued extension of *PQI* Interval Orders, submitted, (2000).

[Tsoukiàs and Vincke, 2000] Tsoukiàs A., Vincke Ph.: A characterization of PQI interval orders, to appear in *Discrete Applied Mathematics*, (2000).

[Yao, 1996] Yao Y.: Combination of rough sets and fuzzy sets based on α-level sets. In Lin T.Y., Cercone N. (eds.), Rough sets and data mining, Kluwer Academic, Dordrecht, (1996), 301–321.

[Yao and Wang, 1999] Yao Y., Wang T.: On rough relations: an alternative fromulation. In N. Zhong, A. Skowron, S. Ohsuga, (eds.), New Directions in Rough Sets, Data Mining and Granular-Soft Computing, Springer Verlag, LNAI 1711, Berlin, (1999), 82–90.