On the extension of rough sets under incomplete information

Jerzy Stefanowski[1] and Alexis Tsoukiàs[2]

[1] Institute of Computing Science, Poznań University of Technology, 3A Piotrowo, 60-965 Poznań, Poland, e-mail: Jerzy.Stefanowski@cs.put.poznan.pl
[2] LAMSADE - CNRS, Université Paris Dauphine, 75775 Paris Cédex 16, France e-mail: tsoukias@lamsade.dauphine.fr

Abstract. The rough set theory, based on the indiscernibility relation, is not useful for analysing incomplete information. Therefore, we introduce two generalizations of this theory, besides the well known one based on tolerance relations. The first proposal is based on non symmetric similarity relations, while the second one uses valued tolerance relation. Both approaches provide more informative results than the previously known approach employing simple tolerance relation.

Keywords: incomplete information, rough sets, fuzzy sets, decision rules

1 Introduction

Rough sets theory has been developed since Pawlak's seminal work [6] (see also [7]) as a tool enabling to classify objects which are only "roughly" described, in the sense that the available information enables only a partial discrimination among them although they are considered as different objects. In other terms, objects considered as "distinct" could happen to have the "same" or "similar" description, at least as far as a set of attributes is considered. Such a set of attributes can be viewed as the possible dimensions under which the surrounding world can be described for a given knowledge. An explicit hypothesis done in the classic rough sets theory is that all available objects are completely described by the set of available attributes. Denoting the set of objects as $A = \{a_1, \cdots a_n\}$ and the set of attributes as $C = \{c_1, \cdots c_m\}$ it is considered that $\forall a_j \in A, c_i \in C$, the attribute value always exists, i.e. $c_i(a_j) \neq \emptyset$.

Such an hypothesis, although sound, contrast with several empirical situations where the information concerning the set A is only partial either because it has not been possible to obtain the attribute values (for instance if the set A are patients and the attributes are clinical exams, not all results may be available in a given time) or because it is definitely impossible to get a value for some object on a given attribute. A problem arise when such an incomplete information table is used in order to make a classification which implies an action. It may be the case that such an action has to be undertaken while the information is still incomplete. Under such conditions it is necessary to develop a theory which may enable to induce a classification in presence of partial information.

The problem has been already faced in literature by Grzymala [2] Kryszkiewicz [4, 5], Słowiński and Stefanowski [10]. Our paper enhances such works by distinguishing two different semantics for the incomplete information: the "missing" semantics (unknown values allow any comparison) and the "absent" semantics (unknown values do not allow any comparison) and explores three different formalisms to handle incomplete information tables: tolerance relations, non symmetric similarity relations and valued tolerance relations.

Due to the limited size of the paper we will assume that the reader is at least partly familiar with basic rough set concepts, i.e. notion of information table, "classic" indiscernibility relation, approximations of ambiquous decision classes, reduct and decision rules. More information can be found e.g. in [8]; rough set based rule induction is discussed in [3, 9, 12]. The rest paper is organized as follows. In section 2 we present and discuss the tolerance approach introduced by Kryszkiewicz [4]. Moreover, we give an example of incomplete information table which will be used all along the paper in order to help the understanding of the different approaches and allow comparisons. In section 3 an approach based on non symmetric similarity relations is introduced using some results obtained by Słowiński and Vanderpooten [11]. We also demonstrate that the non symmetric similarity approach refines the results obtained using the tolerance relation approach. Finally, in section 4 a valued tolerance approach is introduced and discussed as an intermediate approach among the two previous ones. Further research directions are included in the conclusions.

2 Tolerance relations

In the following we briefly present the idea introduced by Kryszkiewicz [4]. The interested readers can refer to the quoted papers for more details.

In our point of view the key concept introduced in this approach is to associate to the unavailable values of the information table a "null" value to be considered as "everything is possible" value. Such an interpretation corresponds to the idea that such values are just "missing", but they do exist. In other words, it is our imperfect knowledge that obliges us to work with a partial information table. Each object potentially has a complete description, but we just miss it for the moment. More formally, given an information table $IT = (A, C)$ we denote the missing values by $*$ and we introduce the following binary relation T:
$\forall x, y \in A \times A \;\; T(x, y) \Leftrightarrow \forall c_j \in C \;\; c_j(x) = c_j(y)$ or $c_j(x) = *$ or $c_j(y) = *$

Clearly T is a reflexive and symmetric relation, but not necessarily transitive. We call the relation T a "tolerance relation". Further on let us denote by $I_C(x)$ the set of of objects y for which $T(x, y)$ holds. In other terms the set $I_C(x)$ can be seen as the set of objects similar to x taking into account attributes C. We call such a set the "tolerance class of x", thus allowing the definition of a set of tolerance classes of the set A. We can now use the tolerance classes as the basis for redefining the concept of lower and upper approximation of a set Φ using the set of attributes C. We have:
$\Phi_C = \{x \in A | I_C(x) \subseteq \Phi\}$ the lower approximation of Φ
$\Phi^C = \{x \in A | I_C(x) \cap \Phi \neq \emptyset\}$ the upper approximation of Φ
It is easy to observe that $\Phi^C = \bigcup\{I(x) | x \in \Phi\}$ also. The usual properties of lower and upper approximations apply in this case also. Let us introduce now an example of incomplete information table which will be used all along the paper in order to help the understanding of the different approaches and allow comparisons.

Example 1. Suppose the following information table is given

A	a_1	a_2	a_3	a_4	a_5	a_6	a_7	a_8	a_9	a_{10}	a_{11}	a_{12}
c_1	3	2	2	*	*	2	3	*	3	1	*	3
c_2	2	3	3	2	2	3	*	0	2	*	2	2
c_3	1	2	2	*	*	2	*	0	1	*	*	1
c_4	0	0	0	1	1	1	3	*	3	*	*	*
d	Φ	Φ	Ψ	Φ	Ψ	Ψ	Φ	Ψ	Ψ	Φ	Ψ	Φ

where a_1, ..., a_{12} are the available objects, c_1, ..., c_4 are four attributes which values (discrete) range from 0 to 3 and d is a decision attribute classifying objects either to the set Φ or to the set Ψ.

Using the tolerance relation approach to analyse the above example we have the following results (notice that $\Psi = \Phi^c$): $I_C(a_1) = \{a_1, a_{11}, a_{12}\}$, $I_C(a_2) = \{a_2, a_3\}$, $I_C(a_3) = \{a_2, a_3\}$, $I_C(a_4) = \{a_4, a_5, a_{10}, a_{11}, a_{12}\}$, $I_C(a_5) = \{a_4, a_5, a_{10}, a_{11}, a_{12}\}$, $I_C(a_6) = \{a_6\}$, $I_C(a_7) = \{a_7, a_8, a_9, a_{11}, a_{12}\}$, $I_C(a_8) = \{a_7, a_8, a_{10}\}$, $I_C(a_9) = \{a_7, a_9, a_{11}, a_{12}\}$, $I_C(a_{10}) = \{a_4, a_5, a_8, a_{10}, a_{11}\}$, $I_C(a_{11}) = \{a_1, a_4, a_5, a_7, a_9, a_{10}, a_{11}, a_{12}\}$,

$I_C(a_{12}) = \{a_1, a_4, a_5, a_7, a_9, a_{11}, a_{12}\}$. From which we can deduce that: $\Phi_C = \emptyset$, $\Phi^C = \{a_1, a_2, a_3, a_4, a_5, a_7, a_8, a_9, a_{10}, a_{11}, a_{12}\}$, $\Psi_C = \{a_6\}$, $\Psi^C = A$

The results are quite poor. Moreover there exist elements which intuitively could be surely classified in Φ or in Ψ, while they are not. Take for instance a_1. We have complete knowledge about it and intuitively there is no element perceived as similar to it. However, it is not in the lower approximation of Φ. This is due to "missing values" of a_{11} and a_{12} which enables them to be considered as "similar" to a_1. Of course this is "safe" because potentially the two objects could come up with exactly the same values of a_1.

A reduct is defined in a similar way as in the "classical" rough set model, i.e. it is a minimal subset of attributes that preserves lower approximations of object classification. In table from Example 1 one can notice that it is possible to reduce that table just by skipping attribute c_3 - we can built the same classes of tolerance relation and approximations as ones created using all attributes. The set of attributes $\{c_1, c_2, c_4\}$ is the only reduct in this information table. In Kryszkiewicz [4] generalized decision rules and their generation from incomplete information tables is discussed. Generalized decision rules are of the form $\wedge_i(c_i, v)\rightarrow\vee(d, w)$. If the decision part contains one disjunct only, the rule is certain. Let B be a set of condition attributes which occur in a conditional part of the rule $s \rightarrow t$. A decision rule is true if for each object x satisfying condition part s, $I_B(x) \subseteq [t]$. It is also required that the rule must have non-redundant conditional part. In our example, we can find only one certain decision rule (due to the small size of the lower approximation) : $(c_1 = 2)\wedge(c_2 = 3)\wedge(c_4 = 1)\rightarrow(d = \Psi)$.

3 Similarity Relations

We introduce now a new approach based on the concept of a not necessarily symmetric similarity relation. Such a concept has been first introduced in general rough sets theory by Słowiński and Vanderpooten [11] in order to enhance the concept of indiscernability relation. We first introduce what we call the "absent values semantics" for incomplete information tables. In this approach we consider that objects may be described "incompletely" not only because of our imperfect knowledge, but also because definitely impossible to describe them on all the attributes Therefore we do not consider the unknown values as uncertain, but as "non existing" and we do not allow to compare unknown values.

Under such a perspective each object may have a more or less complete description, depending on how many attributes has been possible to apply. From this point of view an object x can be considered similar to another object y only if they have the same *known* values. More formally, denoting as usual the unknown value as $*$ and given an information table $IT = (A, C)$ we introduce a similarity relation S as follows: $\forall x, y \;\; S(x, y) \Leftrightarrow \forall c_j \in C : \;\; c_j(x) \neq *, \;\; c_j(x) = c_j(y)$

It is easy to observe that such a relation although not symmetric is transitive. The relation S is a partial order on the set A. Actually it can be seen as a representation of the inclusion relation since we can consider that *"x is similar to y"* iff the *"the description of x"* is included in *"the description of y"*. We can

now introduce for any object $x \in A$ two sets:
$R(x) = \{y \in A | S(y, x)\}$ the set of objects similar to x
$R^{-1}(x) = \{y \in A | S(x, y)\}$ the set of objects to which x is similar

Clearly $R(x)$ and $R^{-1}(x)$ are two different sets. We can now introduce our definitions for the lower and upper approximation of a set Φ as follows:
$\Phi_C = \{x \in A | R^{-1}(x) \subseteq \Phi\}$ the lower approximation of Φ
$\Phi^C = \bigcup\{R(x) | x \in \Phi\}$ the upper approximation of Φ

In other terms we consider as surely belonging to Φ all objects which have objects similar to them belonging to Φ. On the other hand any object which is similar to an object in Φ could potentially belong to Φ. Comparing our approach with the tolerance relation based one we can state the following result.

Theorem 1. *Given an information table $IT = (A, C)$ and a set Φ, the upper and lower approximations of Φ obtained using a non symmetric similarity relation are a refinement of the ones obtained using a tolerance relation.*

Proof. Denote as Φ_C^T the lower approximation of Φ using the tolerance approach and Φ_C^S the lower approximation of Φ using the similarity approach, Φ_T^C and Φ_S^C being the upper approximations respectively. We have to demonstrate that: $\Phi_C^T \subseteq \Phi_C^S$ and $\Phi_S^C \subseteq \Phi_T^C$.

Clearly we have that: $\forall x, y \;\; S(x, y) \rightarrow T(x, y)$ since the conditions for which the relation S holds are a subset of the conditions for which the relation T holds. Then it is easy to observe that: $\forall x \;\; R(x) \subseteq I(x)$ and $R^{-1}(x) \subseteq I(x)$.

1. $\Phi_C^T \subseteq \Phi_C^S$. By definition $\Phi_C^T = \{x \in A \; |I(x) \subseteq \Phi\}$ and $\Phi_C^S = \{x \in A \; |R^{-1}(x) \subseteq \Phi\}$. Therefore if an object x belongs to Φ_C^T we have that $I_C(x) \subseteq \Phi$ and since $R^{-1}(x) \subseteq I(x)$ we have that $R^{-1}(x) \subseteq \Phi$ and therefore the same object x will belong to Φ_C^S. The inverse is not always true. Therefore the lower approximation of Φ using the non symmetric similarity relation is at least as rich as the lower approximation of Φ using the tolerance relation.
2. $\Phi_S^C \subseteq \Phi_T^C$. By definition $\Phi_S^C = \cup_{x \in \Phi} R(x)$ and $\Phi_T^C = \cup_{x \in \Phi} I(x)$ and since $R(x) \subseteq I(x)$ the union of the sets $R(x)$ will be a subset of the union of the sets $I(x)$. The inverse is not always true. Therefore the upper approximation of Φ using the non symmetric similarity relation is at most as rich as the upper approximation of Φ using the tolerance relation.

Continuation of Example 1 Let us come back to the example introduced in section 1. Using the whole set of attributes we have the following results (notice that $\Psi = \Phi^c$): $R^{-1}(a_1) = \{a_1\}$, $R(a_1) = \{a_1, a_{11}, a_{12}\}$, $R^{-1}(a_2) = \{a_2, a_3\}$, $R(a_2) = \{a_2, a_3\}$, $R^{-1}(a_3) = \{a_2, a_3\}$, $R(a_3) = \{a_2, a_3\}$, $R^{-1}(a_4) = \{a_4, a_5\}$, $R(a_4) = \{a_4, a_5, a_{11}\}$, $R^{-1}(a_5) = \{a_4, a_5\}$, $R(a_5) = \{a_4, a_5, a_{11}\}$, $R^{-1}(a_6) = \{a_6\}$, $R(a_6) = \{a_6\}$, $R^{-1}(a_7) = \{a_7, a_9\}$, $R(a_7) = \{a_7\}$, $R^{-1}(a_8) = \{a_8\}$, $R(a_8) = \{a_8\}$, $R^{-1}(a_9) = \{a_9\}$, $R(a_9) = \{a_7, a_9, a_{11}, a_{12}\}$, $R^{-1}(a_{10}) = \{a_{10}\}$, $R(a_{10}) = \{a_{10}\}$, $R^{-1}(a_{11}) = \{a_1, a_4, a_5, a_9, a_{11}, a_{12}\}$, $R(a_{11}) = \{a_{11}\}$, $R^{-1}(a_{12}) = \{a_1, a_9, a_{12}\}$, $R(a_{12}) = \{a_{11}, a_{12}\}$. From which we can deduce that:

$\Phi_C = \{a_1, a_{10}\}$, $\Phi^C = \{a_1, a_2, a_3, a_4, a_5, a_7, a_{10}, a_{11}, a_{12}\}$, $\Psi_C = \{a_6, a_8, a_9\}$, $\Psi^C = \{a_2, a_3, a_4, a_5, a_6, a_7, a_8, a_9, a_{11}, a_{12}\}$.

As expected the new approximations are more informative than the tolerance based ones. Moreover, we find now in the lower approximations of the sets Φ and Ψ some of the objects which intuitively we were expecting to be there. Obviously such an approach is less "safe" than the tolerance based one, since objects can be classified as "surely in Φ" although very little is known about them (as in our case the object a_{10}). However, under the "absent values" semantic which we introduced at the beginning of the section, we do not consider a partially described object as "little known", but as "known" just on few attributes. Absent semantics allow to make a kind of "non monotonic classification", in the sense that the classification is defeasable in the case new information is added such that an object will have a more complete description, classifying it differently from the present class.

Let us now consider the concept of a reduct using the similarity relation. The subset C' of C is a reduct with respect to a classification if it is minimal subset of attributes C that keeps the same lower approximation of such a classification. Therefore, the definition is the same as in the original rough set approach, the difference consists in using different similarity relation while building approximations. We observe that according to definition of the relation an object "totally unknown" (having in all attributes an unknown value) is not similar to any other object. If we eliminate one or more attributes which will make an object to become "totally unknown" on the remaining attributes we lose relevant information for the classification. We can conclude that all such attributes have to be in the reducts. Therefore, there is one reduct in our example $\{c_1, c_2, c_4\}$ - it leads to the same classes $R^{-1}(x)$ and $R(x)$ as using all attributes.

While defining the decision rule we employ classes $R(x)$. The decision rule is defined as $s \rightarrow t$ (where $s = \wedge_i(c_i, v)$ and $t = (d, w)$). The certain rule is true if for each object x satisfying s its class $R(x) \subseteq [t]$. The conditional part cannot contain redundant conditions. This way of defining the decison rule follows the semantics of the rules discussed in the original version of similarity approach [11]. Moreover it is consistent with our idea of "non monotonic classification": we classify objects which as similar as possible to a given rule although this might not be the safest conclusion.

Given absent value semantics and similarity relation the following certain decision rules can be generated from the example of the information table:
$(c_1 = 1) \rightarrow (d = \Phi)$, $(c_3 = 1) \wedge (c_4 = 0) \rightarrow (d = \Phi)$, $(c_3 = 3) \wedge (c_4 = 0) \rightarrow (d = \Phi)$
$(c_2 = 3) \wedge (c_4 = 1) \rightarrow (d = \Psi)$, $(c_2 = 0) \rightarrow (d = \Psi)$, $(c_3 = 0) \rightarrow (d = \Psi)$
The absent value semantics gives more informative decision rules than tolerance based approach. Nevertheless these two different approaches (the tolerance and the non symmetric similarity) appear to be two extremes, in the middle of which it could be possible to use a more flexible approach.

4 Valued tolerance relations

Going back to the example of section 2, let's consider the elements a_1, a_{11} and a_{12}. Under both the tolerance relation approach and the non symmetric similarity relation approach we have: $T(a_{11}, a_1), T(a_{12}, a_1), S(a_{11}, a_1), S(a_{12}, a_1)$

However we may desire to express the intuitive idea that a_{12} is "more similar" to a_1 than a_{11} or that a_{11} is "less similar" to a_1 than a_{12}. This is due to the fact that in the case of a_{12} only one value is unknown and the rest all are equal, while in the case of a_{11} only one value is equal and the rest are unknown. We may try to capture such a difference using a valued tolerance relation.

The reader may notice that we can define different types of valued tolerance (or similarity) using different comparison rules. Moreover a valued tolerance (or similarity) relation can be defined also for complete information tables. Actually the approach we will present is independent from the specific formula adopted for the valued tolerance and can be extended to any type of valued relation.

Given a valued tolerance relation for each element of U we can define a "tolerance class" that is a fuzzy set with membership function the "tolerance degree" to the reference object. It is easy to observe that if we associate to the non zero tolerance degree the value 1 we obtain the tolerance classes introduced in section 2. The problem is to define the concepts of upper and lower approximation of a set Φ. Given a set Φ to describe and a set $Z \subseteq U$ we will try to define the degree by which Z approximates from the top or from the bottom the set Φ. Under such a perspective, each subset of U may be a lower or upper approximation of Φ, but to different degrees. For this purpose we need to translate in a functional representation the usual logical connectives of negation, conjunction etc..:

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 usual logical equivalence $x \rightarrow y =_{def} \neg x \vee y$;

- the second claiming that whenever the truth value of x is not greater than the truth value of 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 a discussion see [1]).

Coming back to our lower and upper approximations we know that given a set $Z \subseteq U$ and a set Φ the usual definitions are:
1. $Z = \Phi_C \;\Leftrightarrow\; \forall\, z \in Z,\;\; \Theta(z) \subseteq \Phi$, 2. $Z = \Phi^C \;\Leftrightarrow\; \forall\, z \in Z,\;\; \Theta(z) \cap \Phi \neq \emptyset$
$\Theta(z)$ being the "indiscernability (tolerance, similarity etc.)" class of element z. The functional translation of such definitions is straightforward. Having:
$\forall\, x\; \phi(x) =_{def} T_x \phi(x)$; $\exists\, x\; \phi(x) =_{def} S_x \phi(x)$; $\Phi \subseteq \Psi =_{def} T_x(I(\mu_\Phi(x), \mu_\Psi(x)))$; $\Phi \cap \Psi \neq \emptyset =_{def} \exists\, x\; \phi(x) \wedge \psi(x) =_{def} S_x(T(\mu_\Phi(x), \mu_\Psi(x)))$ we get:
1. $\mu_{\Phi_C}(Z) \;=\; T_{z \in Z}(T_{x \in \Theta(z)}(I(R(z,x), \hat{x})))$.
2. $\mu_{\Phi^C}(Z) \;=\; T_{z \in Z}(S_{x \in \Theta(z)}(T(R(z,x), \hat{x})))$. where:
$\mu_{\Phi_C}(Z)$ is the degree for set Z to be a lower approximation of Φ;
$\mu_{\Phi^C}(Z)$ is the degree for set Z to be an upper approximation of Φ;
$\Theta(z)$ is the tolerance class of element z; T, S, I are the functions previously defined; $R(z,x)$ is the membership degree of element x in the tolerance class of z; $\hat{x}$ is the membership degree of element x in the set Φ ($\hat{x} \in \{0,1\}$).

Continuation of Example 1 Considering that the set of possible values on each attribute is discrete we make the hypothesis that there exists a uniform probability distribution among such values. More formally, consider c_j an attribute of an information table $IT = (A, C)$ and associate to it the set $E_j = \{e_j^1, \cdots e_j^m\}$ of all the possible values of the attribute. Given an element $x \in A$ the probability that $c_j(x) = e_j^i$ is $1/|E_j|$. Therefore given any two elements $x, y \in A$ and an attribute c_j, if $c_j(y) = e_j^i$, the probability $R_j(x,y)$ that x is similar to y on the attribute c_j is $1/|E_j|$. On this basis we can compute the probability that two elements are similar on the whole set of attributes as the joint probability that the values of the two elements are the same on all the attributes: $R(x,y) \;=\; \prod_{c_j \in C} R_j(x,y)$. Applying this rule to the whole set example we obtain the following table 1 concerning the valued tolerance relation.

	a_1	a_2	a_3	a_4	a_5	a_6	a_7	a_8	a_9	a_{10}	a_{11}	a_{12}
a_1	1	0	0	0	0	0	0	0	0	0	1/64	1/4
a_2	0	1	1	0	0	0	0	0	0	0	0	0
a_3	0	1	1	0	0	0	0	0	0	0	0	0
a_4	0	0	0	1	1/256	0	0	0	0	1/1024	1/1024	1/64
a_5	0	0	0	1/256	1	0	0	0	0	1/1024	1/1024	1/64
a_6	0	0	0	0	0	1	0	0	0	0	0	0
a_7	0	0	0	0	0	0	1	1/256	1/16	0	1/1024	1/64
a_8	0	0	0	0	0	0	1/256	1	0	1/1024	0	0
a_9	0	0	0	0	0	0	1/16	0	1	0	1/64	1/4
a_{10}	0	0	0	1/1024	1/1024	0	0	1/1024	0	1	1/4096	0
a_{11}	1/64	0	0	1/1024	1/1024	0	1/1024	0	1/64	1/4096	1	1/256
a_{12}	1/4	0	0	1/64	1/64	0	1/64	0	1/4	0	1/256	1

Table 1: Valued tolerance relation for example 2.1

If we consider element a_1, the valued tolerance relation $R(a_1, x),\; x \in U$ will result in the vector $[1,0,0,0,0,0,0,0,0,0,1/64,1/4]$ which actually represents the tolerance class $\Theta(a_1)$ of element a_1. The reader may notice that the crisp tolerance class of element a_1 was the set $\{a_1, a_{11}, a_{12}\}$ which corresponds to

the vector $[1,0,0,0,0,0,0,0,0,0,1,1]$. Following our "probabilistic approach" we may choose for T and S the product representation, while for I we will satisfy the De Morgan property thus obtaining: $T(x,y) = xy$, $S(x,y) = x + y - xy$, $I(x,y) = 1 - x + xy$. Clearly our choice of $I(x,y)$ does not satisfy the second property of implication. However, the reader may notice that in our specific case we have a peculiar implication from a fuzzy set ($\Theta(z)$) to a regular set (Φ), such that $\hat{x} \in \{0,1\}$. The application of any implication satisfying the second property will reduce the valuation to the set $\{0,1\}$ and therefore the whole degree $\mu_{\Phi_C}(Z)$ will collapse to $\{0,1\}$ and thus to the usual lower approximation. With such considerations we obtain:
$\mu_{\Phi_C}(Z) = \prod_{z\in Z}\prod_{x\in\Theta(z)}(1 - R(z,x) + R(z,x)\hat{x})$
$\mu_{\Phi^C}(Z) = \prod_{z\in Z}(1 - \prod_{x\in\Theta(z)}(1 - R(z,x)\hat{x}))$

Consider now the set Φ and as set Z consider the element a_1, where $R(a_1,x)$ was previously introduced and $\hat{x} = [1,1,0,1,0,0,1,0,0,1,0,1]$. We obtain $\mu_{\Phi_C}(a_1) = 0.98$ and $\mu_{\Phi^C}(a_1) = 1$. Operationally we could choose a set Z as lower (upper) approximation of set Φ as follows:

Step 1 take all elements for which $\mu(\Theta(z)\to\Phi) = 1$ ($\mu(\Theta(z)\cap\Phi) = 1$);
Step 2 then add elements in a way such that $\mu(\Theta(z)\to\Phi) > k$ ($\mu(\Theta(z)\cap\Phi) > k$), (for decreasing values of k, let's say 0.99, 0.98 etc.), thus obtaining a family of lower (upper) approximations with decreasing membership function $\mu_{\Phi_C}(Z)$ ($\mu_{\Phi^C}(Z)$);
Step 3 fix a minimum level λ enabling to accept a set Z as a lower (upper) approximation of Φ (thus $\mu_{\Phi_C}(Z) \geq \lambda$).

The concept of reduct and decision rules are also generalized in the valued tolerance case. Given the decision table (A,C) and the partition $\mathcal{Y} = \Phi_1, \Phi_2, \ldots \Phi_n$, the subset of attributes $C' \subset C$ is a reduct iff it does not reduce the degree of lower approximation obtained with C, i.e. if $z_1, z_2, \ldots, z_n$ is a family of lower approximations of $\Phi_1, \Phi_2, \ldots \Phi_n$ then $\forall_{i=1,\ldots,n} z_i\ \mu_{\Phi_{iC}}(z_i) \leq \mu_{\Phi_{iC'}}(z_i)$.

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: $\rho_i^J =_{def} \bigwedge_j (c_j(a_i) = v) \to (d = w)$ where: $J \subseteq C$, v is the value of attribute c_j, w is the value of attribute d.

As usual we may use relation $s(x,\rho_i)$ 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. We denote as $S(\rho_i) = \{x : s(x,\rho_i) > 0\}$ and as $W = \{x : d(x) = w\}$. Then ρ_i is a classification rule iff: $\forall\ x\ \in S(\rho_i) :\ \Theta(x) \subseteq W$

We can compute a credibility degree for any rule ρ_i calculating the truth value of the previous formula which can be rewritten as:
$\forall\ x,y\ s(x,\rho_i)\to(R(x,y)\to W(y))$. We get: $\mu(\rho_i) = T_x(I_y(s(x,\rho_i), I(\mu_{\Theta(x)}(y), \mu_W(y))))$

Finally it is necessary to check whether J is a non-redundant set of conditions for rule ρ_i, i.e. to look if it is possible to satisfy the condition: $\exists\ \hat{J} \subset J :\ \mu(\rho_i^{\hat{J}}) \geq \mu(\rho_i^J)$ or not.

Continuation of Example 1. Consider again the example of incomplete information table used in the paper and take as candidate the rule: $\rho_1 : \quad (c_1 = 3) \wedge (c_2 = 2) \wedge (c_3 = 1) \wedge (c_4 = 0) \rightarrow (d = \Phi)$ Since in the paper we have chosen for the functional representation of implication the satisfaction of De Morgan law and for T-norms the product, we get:
$\mu(\rho_i) = \prod_{x \in S(\rho_i)} (1 - s(x, \rho_i) + s(x, \rho_i) \prod_{y \in \Theta(x)} (1 - \mu_{\Theta(x)}(y) + \mu_{\Theta(x)}(y)\mu_W(y)))$
where $s(x, \rho_i)$ represents the "support" degree of element x to the rule ρ_i. We thus get that $\mu(\rho_1) = 0.905$
However, the condition part of rule ρ_1 is redundant as the rule could be transformed to: $\rho_1 : \quad (c_1 = 3) \wedge (c_3 = 1) \wedge (c_4 = 0) \rightarrow (d = \Phi)$ with degree $\mu(\rho_1) = 0.905$.

Operationally a user may first fix a threshold of credibility for the rules to accept and then could operate a sensitivity analysis on the set of rules that is possible to accept in an interval of such threshold.

Supposing that the level of credibility is fixed at 0.9 we can induce from the decision table the following set of decision rules:
$\rho_1 : \quad (c_1 = 3) \wedge (c_3 = 1) \wedge (c_4 = 0) \rightarrow (d = \Phi)$ with $\mu(\rho_1) = 0.905$
$\rho_2 : \quad (c_1 = 2) \wedge (c_4 = 1) \rightarrow (d = \Psi)$ with $\mu(\rho_1) = 0.931$
$\rho_3 : \quad (c_2 = 3) \wedge (c_3 = 2) \wedge (c_4 = 1) \rightarrow (d = \Psi)$ with $\mu(\rho_1) = 0.969$

If we reduce the level to 0.87 we can substitute the third rule by two others:
$\rho_3 : \quad (c_2 = 3) \wedge (c_4 = 1) \rightarrow (d = \Psi)$ with $\mu(\rho_1) = 0.879$
$\rho_4 : \quad (c_2 = 3) \wedge (c_3 = 2) \rightarrow (d = \Psi)$ with $\mu(\rho_1) = 0.879$

5 Conclusions

Rough sets theory has been conceived under the implicit hypothesis that all objects in a universe can be evaluated under a given set of attributes. However, it can be the case that several values are not available for various reasons. In our paper we introduce two different semantics in order to distinguish such situations. "Missing values" imply that not available information could always become available and that in order to make "safe" classifications and rules induction we might consider that such missing values are equal to everything. Tolerance relations (which are reflexive and symmetric, but not transitive) capture in a formal way such an approach. "Absent values" imply that not available information cannot be used in comparing objects and that classification and rules induction should be performed with the existing information since the absent values could never become available. Similarity relations (which in our case are reflexive and transitive, but not symmetric) are introduced in our paper in order to formalize such an idea. We demonstrate in the paper that our approach always lead to more informative results with respect to the tolerance relation based approach (although less safe).

A third approach is also introduced in the paper, as an intermediate position among the two previously presented. Such an approach is based on the use of a valued tolerance relation. A valued relation could appear for several reasons not only because of the non available information and in fact the approach presented

has a more general validity. However in this paper we limit ourselves in discussing the missing values case. A functional extension of the concepts of upper and lower approximation is introduced in the paper so that to any subset of the universe a degree of lower (upper) approximation can be associated. Further on such a functional extension enables to compute a credibility degree for any rule induced by the classification. Further research direction include, but are not limited to:
- a further analysis of rules induction properties and algorithms under the non symmetric similarity relation based approach;
- an analysis of the non-monotonic behaviour of the classification obtained under the non symmetric similarity relation based approach (what happens if an absent value becomes available);
- the introduction of other examples of valued tolerance relations, besides the probability based one introduced as an example in the paper;
- the analysis of the obtained results when the valued tolerance relation obeys to a precise fuzzy set formalism as for instance possibility distributions.

References

1. Dubois D., Lang J., Prade H., Fuzzy sets in approximate reasoning *Fuzzy Sets and Systems* 40 (1991), 203–244.
2. 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.
3. Grzymala-Busse J. W. LERS - A system for learning from examples based on rough sets. In Slowinski R. (ed.), *Intelligent Decision Support. Handbook of Applications and Advances of the Rough Sets Theory*, Kluwer Academic Publishers, 1992, 3–18.
4. Kryszkiewicz M., Rough set approach to incomplete information system. ICS Research Report 55/95, Warsaw University of Technology, also appeared in *Information Sciences* 112 (1998), 39–49.
5. 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, 1998, 422–450.
6. Pawlak Z., Rough sets. *Int. J. Computer and Information Sci.*, 11, 1982, 341–356.
7. Pawlak Z., *Rough sets. Theoretical aspects of reasoning about data.* Kluwer Acad. Publs., Dordrecht, 1991.
8. Pawlak Z. Rough set elements. In Polkowski L., Skowron A. (eds.) *Rough Sets in Data Mining and Knowledge Discovery*, Physica-Verlag, 1998, 10–31.
9. Skowron A., Boolean reasoning for decision rules generation. In Komorowski J., Ras Z. (eds.) *Methodologies for Intelligent Systems*, LNAI 689, Springer-Verlag, Berlin, 1993, 295–305.
10. Słowiński R., Stefanowski J., Rough classification in incomplete information systems. *Math. Computing Modelling*, 12 (10/11), 1989, 1347-1357.
11. 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.
12. Stefanowski J., On rough set based approaches to induction of decision rules. In Polkowski L., Skowron A. (eds.) *Rough Sets in Data Mining and Knowledge Discovery*, Physica-Verlag, 1998, 500–530.