

Incomplete information tables and rough classification

Jerzy Stefanowski (*), Alexis Tsoukiàs (+)

(*) Institute of Computing Science
Poznan University of Technology, 60-965, Poznan, Poland
e-mail: Jerzy.Stefanowski@cs.put.poznan.pl
(+) LAMSADE - CNRS,
Université Paris Dauphine, 75775 Paris Cédex 16, France
e-mail: tsoukias@lamsade.dauphine.fr

Abstract. The rough set theory, based on the original definition of the indiscernibility relation, is not useful for analysing incomplete information tables where some values of attributes are unknown. In this paper we distinguish two different semantics for incomplete information: the “missing value” semantics and the “absent value” semantics. The already known approaches, e.g. based on the tolerance relations, deal with the missing value case. We introduce two generalisations of the rough sets theory to handle these situations. The first generalisation introduces the use of a non symmetric similarity relation in order to formalise the idea of absent value semantics. The second proposal is based on the use of valued tolerance relations. A logical analysis and the computational experiments show that for the valued tolerance approach it is possible to obtain more informative approximations and decision rules than using the approach based on the simple tolerance relation.

Keywords: incomplete information, rough sets, fuzzy sets, similarity relation, valued tolerance relation, decision rules

1 Introduction

Rough sets theory has been developed since Pawlak’s paper [Pawlak, 1981] (see also [Pawlak, 1991, Komorowski et al., 1999]) as a tool for analysing vague descriptions of objects. The starting point of this theory is an observation that objects having the same description are indiscernible (or similar) with respect to available information. In other words, it may happen that objects considered as distinct could 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.

However, an explicit hypothesis done in the rough sets theory is that all available objects are completely described by the set of attributes. Denoting the set of objects as $A = \{a_1, \dots, a_n\}$ and the set of attributes as $C = \{c_1, \dots, c_m\}$ it is assumed that $\forall a_j \in A, c_i \in C$, the attribute value always exists, i.e. $c_i(a_j) \neq \emptyset$.

Such a hypothesis, although sound, contrasts with several empirical situations where information concerning the set A is only partial, either because it has not been possible to record the attribute values (for instance, if the set A are patients and the attributes are clinical examinations, some results may not be available in a given moment) or because it is definitely impossible to get a value on a given attribute for certain object (for instance,

this patient could be allergic to a specific examination). A problem arises 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. Therefore, it is necessary to develop a theory which may enable to classify objects in presence of partial information.

The problem has already been faced in literature. [Słowiński and Stefanowski, 1989] and [Grzymala, 1991] proposed two different approaches to replace unknown value of attribute by specific subsets of values. Grzymala [Grzymala et al., 1999, Grzymala et al. 2000] performed computational studies on the medical data, where unknown values of attributes were replaced using probabilistic techniques. Kryszkiewicz introduced a new tolerance relation to handle incomplete information tables without any changes in their content [Kryszkiewicz, 1995, Kryszkiewicz, 1998]. Some properties of incomplete information were also studied in [Toumi, 1996]. Recently, Greco and Slowinski used a specific definition of the indiscernibility relation to analyse unknown attributes values for multi-criteria decision problems [Greco et al.,1999a, Greco et al.,1999b]. In their approach the indiscernibility relation between a pair of objects is considered as a directional statement where a subject object is compared to a referent object that cannot have any missing values.

Our paper enhances such works by distinguishing two different semantics for incomplete information:

- the “missing value” semantics (unknown values of attributes allow any comparison),
- the “absent value ” semantics (unknown values of attributes do not allow any comparison).

Three different formalisms to handle incomplete information tables are explored: tolerance relation, non symmetric similarity relations and valued tolerance relation. In all cases particular attention is paid to create approximations of sets and to induce decision rules from the incomplete information table.

The paper is organised as follows. In section 2, a brief reminder of rough sets is presented. Then, an example of incomplete information table is introduced which will be used all along the paper in order to help the understanding of the different approaches. In section 3, we present and discuss the tolerance relation approach proposed by Kryszkiewicz [Kryszkiewicz, 1995]. In section 4, an approach based on non symmetric similarity relations is introduced using some results obtained by Słowiński and Vanderpooten [Słowiński and Vanderpooten, 1997, Słowiński and Vanderpooten, 2000]. We also demonstrate that the non symmetric similarity approach refines the results obtained using the tolerance relation approach. In section 5, a valued tolerance approach is introduced through the concepts of lower and upper approximability. In section 6, we show how approximations can be computed and how decision rules can be induced using such a valued tolerance relation. Since in this approach the user can control the process of calculating approximations and decision rules by some thresholds referring to his acceptance levels, we performed a computational study on several data sets. Results are summarised in section 7. Further research directions are included in the conclusions.

2 Rough Sets

Rough sets have been introduced by Pawlak [Pawlak, 1981] as a new approach for analysing vague information and more precisely the capability of distinguishing objects described in an ambiguous way.

Following Pawlak, an information table IT is a pair (A, C) where $A \neq \emptyset$ is a non-empty set of objects and C is a non empty set of attributes, such that $\forall c \in C : U \rightarrow V_c$ where V_c is a domain of c . Each subset of attributes $B \subseteq C$ defines an *indiscernibility* relation $IND(B)$ as:

$$IND(B) = \{(x, y) \in A \times A : \forall c_i \in B \ c_i(x) = c_i(y)\}$$

Let $A/IND(B)$, or shortly A/B , be the family of all the equivalence classes of the equivalence relation $IND(B)$ on A .

Given an information table $IT = (A, C)$ and a non-empty subset of attributes $B \subseteq C$, with each subset of objects $\Phi \subseteq A$ we associate two sets:

$$\Phi_B = \bigcup \{Y \in A/B : Y \subseteq \Phi\}$$

$$\Phi^B = \bigcup \{Y \in A/B : Y \cap \Phi \neq \emptyset\}$$

called the *B-lower* and *B-upper approximation* of Φ , respectively. In other words, if it is difficult to describe a set of objects Φ (some classes of $IND(B)$ may contain at the same time objects of Φ and of its complement), then it is possible to approximate it using two sets: the lower approximation and the upper approximation. The lower approximation contains objects of A which are certainly assigned to Φ while the upper approximation contains these objects which possibly belong to Φ , following the partition A/B .

$Bn(\Phi) = \Phi^B \setminus \Phi_B$ is the *B-boundary region* of Φ , i.e. it is the set of elements for which there is a doubt about their assigning to Φ .

The definition of approximation of a subset $\Phi \in A$ can be extended to a classification of all objects, i.e. partition $\mathcal{Y} = \{\Phi_1, \Phi_2, \dots, \Phi_n\}$ of A . By *B-lower* and *B-upper approximations* of \mathcal{Y} we understand sets $\mathcal{Y}_B = \{\Phi_{1B}, \Phi_{2B}, \dots, \Phi_{nB}\}$ and $\mathcal{Y}^B = \{\Phi_1^B, \Phi_2^B, \dots, \Phi_n^B\}$, respectively.

The approximations of \mathcal{Y} by the set of attributes $B \subseteq C$ can be numerically characterised by the coefficient called a *quality of the approximation of classification* γ . It is defined as ratio

$$\gamma_B(\mathcal{Y}) = \frac{(\sum_{i=1}^n |\Phi_{iB}|)}{|A|}$$

where $|\Phi|$ denotes the cardinality of a set.

The basic concepts of rough approximations can be used for further operations, i.e. reduction of attributes and derivation of decision rules from the decision table. Reduction of the information table allows to find smaller subsets of attributes that preserve the quality of approximation of classification for all objects in the table. It is done on the basis of a *reduct* of the subset of attributes $B \subseteq C$.

The subset $B' \subseteq B$ is a reduct of B with respect to the approximation of partition \mathcal{Y} on A (notation $RED_{\mathcal{Y}}(B)$) if it is a minimal subset of B which keeps the quality of

classification unchanged, i.e. $\gamma_{B'}(\mathcal{Y}) = \gamma_B(\mathcal{Y})$. More than one \mathcal{Y} -reduct of B can exist in an information table.

A decision table DT is an information table $(A, C \cup \{d\})$ where $d \notin C$ is a distinguished attribute called *decision* and elements of C are called *condition* attributes. The decision attribute induces a partition of A into decision classes in a way which is independent from the condition attributes of C . If we underline the functional dependencies between condition and decision attributes, a decision table may also be seen as a set of *decision rules*.

A decision rule ρ is represented in the following form:

$$\bigwedge_{c_j \in B} (c_j(a_i) = v) \rightarrow (d(a_i) = w)$$

where $c_j \in B \subseteq C$, v is a value of condition attribute c_j and w is a value of decision attribute d .

The elementary condition is usually represented as an attribute value pair ($c_j = v$) and decision as a pair ($d = w$). We denote by $s = \bigwedge_j (c_j = v)$ and by $t = (d = w)$ the condition and decision part of a rule, respectively. Let $[s]$ be the set of objects in DT satisfying s and $[t]$ be the set of objects satisfying t . Objects which satisfy both expressions s and t are called objects *supporting* a rule. A decision rule $s \rightarrow t$ is *certain* if $[s] \subseteq [t]$. If the decision table contains inconsistencies, possible decision rules can be induced from upper approximations of decision classes. In such a case $[s]$ should be a subset of the B -upper approximation of a decision class expressed in t .

Certain and possible decision rules should have *non-redundant* condition parts, i.e. no other certain (possible) rule can be constructed from a proper subset of elementary conditions occurring in the given rule. Induction of decision rules from decision tables is a complex task and a number of algorithms have been already proposed (see e.g. [Grzymala, 1992, Skowron, 1993, Stefanowski, 1998, Komorowski et al., 1999, Słowiński et al., 2000]).

Let us introduce 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 their comparisons.

Example 2.1 *Suppose the following information table is given*

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

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

Suppose that we are trying to induce some decision rules for sets Φ or Ψ . Then, following the original definition of the indiscernibility relation (wrt. to $c_1 \cdots c_4$) we get a result which is of very little help. In the following sections we will present three different approaches in order to handle such a situation.

3 The tolerance relation

In the following section, we briefly present the idea introduced by [Kryszkiewicz, 1995] (see also [Kryszkiewicz, 1998]). The readers can refer to the quoted papers for more details.

In our point of view the key concept introduced in this approach is to interpret an unknown value of the attribute as similar to all other possible values for this attribute. 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 it is just missed for the moment.

Given an information table $IT = (A, C)$, a subset of attributes $B \subseteq C$, denoting the missing attribute value by $*$, $\forall x, y \in A \times A$ the following binary relation T is defined:

$$T_B(x, y) \Leftrightarrow \forall c_j \in B, \quad c_j(x) = c_j(y) \text{ or } c_j(x) = * \text{ or } c_j(y) = *$$

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 $I_B(x)$ the set of objects y for which $T_B(x, y)$ holds. In other words, the set $I_B(x)$ can be seen as the set of objects similar to x taking into account attributes B . The set $I_B(x)$ is called the tolerance class of x .

The tolerance classes are the basis for redefining the concept of lower and upper approximation of a set Φ using the set of attributes $B \subseteq C$. The B -lower approximation and the B -upper approximation of Φ are:

$$\Phi_B = \{x \in A \mid I_B(x) \subseteq \Phi\}$$

$$\Phi^B = \{x \in A \mid I_B(x) \cap \Phi \neq \emptyset\}$$

It is easy to observe that $\Phi^B = \bigcup \{I(x) \mid x \in \Phi\}$. The usual properties of lower and upper approximations apply in this case [Kryszkiewicz, 1995].

Continuation of Example 2.1 Using the example introduced in section 1 we have the following results for all attributes C (notice that $\Psi = \Phi^c$, the complement of Φ):

$$\begin{aligned} I_C(a_1) &= \{a_1, a_{11}, a_{12}\}, \\ I_C(a_2) &= \{a_2, a_3\}, \\ I_C(a_3) &= \{a_2, a_3\}, \\ I_C(a_4) &= \{a_4, a_5, a_{10}, a_{11}, a_{12}\}, \end{aligned}$$

$$\begin{aligned}
I_C(a_5) &= \{a_4, a_5, a_{10}, a_{11}, a_{12}\}, \\
I_C(a_6) &= \{a_6\}, \\
I_C(a_7) &= \{a_7, a_8, a_9, a_{11}, a_{12}\}, \\
I_C(a_8) &= \{a_7, a_8, a_{10}\}, \\
I_C(a_9) &= \{a_7, a_9, a_{11}, a_{12}\}, \\
I_C(a_{10}) &= \{a_4, a_5, a_8, a_{10}, a_{11}\}, \\
I_C(a_{11}) &= \{a_1, a_4, a_5, a_7, a_9, a_{10}, a_{11}, a_{12}\}, \\
I_C(a_{12}) &= \{a_1, a_4, a_5, a_7, a_9, a_{11}, a_{12}\},
\end{aligned}$$

From which we can deduce that:

$$\begin{aligned}
\Phi_C &= \emptyset, \Phi^C = \{a_1, a_2, a_3, a_4, a_5, a_7, a_8, a_9, a_{10}, a_{11}, a_{12}\}, \\
\Psi_C &= \{a_6\}, \Psi^C = A
\end{aligned}$$

The results are quite poor. Moreover there exist objects which intuitively could be classified in Φ or in Ψ , while they are not. Take for instance object a_1 . We have its complete description and intuitively there is no other object perceived as very similar to it. However, a_1 is not included into the lower approximation of Φ . This is due to missing attribute values of objects a_{11} and a_{12} which enables them to be considered as “similar” to a_1 . Of course, this is a “safe” approach because potentially these two objects could come up with exactly the same values of a_1 . Same type of reasoning may apply to objects a_7, a_8, a_9, a_{10} .

Approximations of sets by tolerance relation are then used to define a reduct of the information table. A reduct is defined similarly as in the “classical” rough set, i.e. it is a minimal subset of attributes that preserves the same lower approximations of object classification as for all attributes. In Example 2.1, the set of attributes $\{c_1, c_2, c_4\}$ is the only reduct.

[Kryszkiewicz, 1995] discussed the generation of decision rules from incomplete information tables. She considered mainly generalised decision rules of the form $\bigwedge_j (c_j = v) \rightarrow \bigvee (d = w)$. If the decision part contains one disjunct only, the rule is certain. Let B be a set of condition attributes which occur in a condition 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) :

$$(c_1 = 2) \wedge (c_2 = 3) \wedge (c_4 = 1) \rightarrow (d = \Psi).$$

4 Similarity Relations

We introduce now a new approach based on the concept of a not necessarily symmetric similarity relation. Such a concept has been introduced in rough sets in order to enhance the concept of indiscernibility relation (see [Słowiński and Vanderpooten, 1997, Słowiński and Vanderpooten, 2000]).

Although at a first glance a non symmetric similarity relation may appear odd, we have several intuitive examples where such situation may occur. We always say that a child is similar to a parent but we do not claim that the parent is similar to his child. A copy of a painting is similar to the original, but non necessarily the inverse holds too.

We first introduce what we call the “absent values semantics” for incomplete information tables. In this approach we assume that objects may be partially described not only because of our imperfect knowledge, but also because it is definitely impossible to describe them on all the attributes. For instance, it is impossible to give a colour to a transparent glass, or to make an examination of a patient suffering a specific allergy. Therefore, we do not consider the unknown attribute values as uncertain, but as “non existing” and we do not allow to compare unknown values.

Using this interpretation each object may have a more or less complete description, depending on how many attributes have been applied. From this point of view an object x can be considered as similar to another object y only if they have the same *known* attribute values. Such a relation will not be symmetric since if one object has a more complete description than the other, the inverse relation will not hold. More formally, denoting the unknown attribute value as $*$, given an information table $IT = (A, C)$ and a subset of attributes $B \subseteq C$ we introduce $\forall x, y \in A \times A$ a similarity relation S as follows:

$$S(x, y) \Leftrightarrow \forall c_j \in B \text{ such that } c_j(x) \neq *, \quad c_j(x) = c_j(y)$$

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:

$$R(x) = \{y \in A \mid S(y, x)\} \text{ the set of objects similar to } x,$$

$$R^{-1}(x) = \{y \in A \mid S(x, y)\} \text{ the set of objects to which } x \text{ is similar.}$$

Clearly $R(x)$ and $R^{-1}(x)$ are two different sets. We can now introduce our definitions for the B -lower and B -upper approximation of a set Φ (denoted by Φ_B and Φ^B) as follows:

$$\Phi_B = \{x \in A \mid R^{-1}(x) \subseteq \Phi\}$$

$$\Phi^B = \bigcup \{R(x) \mid x \in \Phi\}$$

In other words, the B -lower approximation of a set Φ contains all objects which have objects similar to them belonging to Φ . On the other hand, the B -upper approximation contains all objects which are similar to an object in Φ .

Comparing our approach with the tolerance relation based approach we can state the following result:

Theorem 4.1 *Given an information table $IT = (A, C)$ and a set Φ , the upper and lower approximations of Φ obtained using the similarity relation are a refinement of the ones obtained using the tolerance relation.*

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

By definition of relations S and T we have that the conditions for which the relation S holds are a subset of the conditions for which the relation T holds. Thus, if $(x, y) \in S$ then $(x, y) \in T$ or $\forall x, y \ S(x, y) \rightarrow T(x, y)$. Therefore: $\forall x \ R(x) \subseteq I(x)$ and $R^{-1}(x) \subseteq I(x)$.

1. $\Phi_B^T \subseteq \Phi_B^S$. By definition $\Phi_B^T = \{x \in A | I(x) \subseteq \Phi\}$ and $\Phi_B^S = \{x \in A | R^{-1}(x) \subseteq \Phi\}$. Therefore if an object x belongs to Φ_B^T we have that $I_B(x) \subseteq \Phi$ 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_B^S . The inverse is not always true. Thus 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. $\Phi_S^B \subseteq \Phi_T^B$. By definition $\Phi_S^B = \cup_{x \in \Phi} R(x)$ and $\Phi_T^B = \cup_{x \in \Phi} I(x)$ 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. ■

Intuitively speaking, since the relation S is a subset of relation T , tolerance classes of elements in A will be “wider” than the respective similarity classes. Thus, approximations based on relation T are less precise than approximations based on relation S .

Continuation of Example 2.1 Let us come back to the example introduced in section 2. Using the complete set of attributes C we have the following results:

$$\begin{array}{ll}
 R^{-1}(a_1) = \{a_1\} & R(a_1) = \{a_1, a_{11}, a_{12}\}, \\
 R^{-1}(a_2) = \{a_2, a_3\} & R(a_2) = \{a_2, a_3\}, \\
 R^{-1}(a_3) = \{a_2, a_3\} & R(a_3) = \{a_2, a_3\}, \\
 R^{-1}(a_4) = \{a_4, a_5\} & R(a_4) = \{a_4, a_5, a_{11}\}, \\
 R^{-1}(a_5) = \{a_4, a_5\} & R(a_5) = \{a_4, a_5, a_{11}\}, \\
 R^{-1}(a_6) = \{a_6\} & R(a_6) = \{a_6\}, \\
 R^{-1}(a_7) = \{a_7, a_9\} & R(a_7) = \{a_7\}, \\
 R^{-1}(a_8) = \{a_8\} & R(a_8) = \{a_8\}, \\
 R^{-1}(a_9) = \{a_9\} & R(a_9) = \{a_7, a_9, a_{11}, a_{12}\}, \\
 R^{-1}(a_{10}) = \{a_{10}\} & R(a_{10}) = \{a_{10}\}, \\
 R^{-1}(a_{11}) = \{a_1, a_4, a_5, a_9, a_{11}, a_{12}\} & R(a_{11}) = \{a_{11}\}, \\
 R^{-1}(a_{12}) = \{a_1, a_9, a_{12}\} & R(a_{12}) = \{a_{11}, a_{12}\}
 \end{array}$$

From which we can deduce that:

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

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 were expecting to be there. Obviously such an approach is less “safe” than the tolerance based one, since objects can be classified as belonging to Φ_B although very little is known about them (e.g. the object a_{10}). However, within the “absent values” semantic, we do not consider a partially described object as “little known”, but as “known” just on few attributes. From this point of view there are no objects in A which could be similar to the object a_{10} because there is no other object with the same known values. Therefore, we can classify a_{10} into the B -lower approximation of Φ . Absent values semantics allows us to make a kind of “non monotonic classification”, in the sense that the classification is defeasible when new information is added, such that an object will have a more complete description (resulting in a different class).

Let us consider the concept of a reduct using the similarity relation. The subset B' of B is a reduct with the respect of classification \mathcal{Y} if it is a minimal subset of attributes B that keeps the same lower approximation of the classification \mathcal{Y} . Let us observe that according to definition of the relation, an object “totally unknown” (i.e. having unknown values for all attributes) 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 therefore conclude that all such attributes have to be in the reducts. In Example 1 there is one reduct $\{c_1, c_2, c_4\}$ - 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 $s = \bigwedge_j (c_j = v)$ and $t = (d = w)$). The certain rule is true if for each object x satisfying s its class $R(x) \subseteq [t]$. The condition part cannot contain redundant conditions. This way of defining the decision rule follows the semantics of the rules discussed in the original version of similarity approach [Słowiński and Vanderpooten, 1997, Słowiński and Vanderpooten, 2000]. Moreover, it is consistent with our idea of “non-monotonic classification”, i.e. we classify objects which are as similar as possible to a given rule although this might not be the safest conclusion.

The following certain decision rules can be generated from the example of the information table:

$$\begin{aligned} (c_1 = 1) &\rightarrow (d = \Phi) \\ (c_3 = 1) \wedge (c_4 = 0) &\rightarrow (d = \Phi) \\ (c_1 = 3) \wedge (c_4 = 0) &\rightarrow (d = \Phi) \\ (c_2 = 3) \wedge (c_4 = 1) &\rightarrow (d = \Psi) \\ (c_2 = 0) &\rightarrow (d = \Psi) \\ (c_3 = 0) &\rightarrow (d = \Psi) \end{aligned}$$

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. We will present such an approach in the next section.

5 Valued tolerance relations

5.1 Motivations

Let us consider the objects a_1 , a_{11} and a_{12} in the illustrative example. For both the tolerance (T) relation approach and the non symmetric (S) 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 attribute 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 different types of valued tolerance (or similarity) can be defined using different comparison rules. Moreover a valued tolerance (or similarity) relation can be defined also for complete information tables. In fact, 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 (see also [Stefanowski and Tsoukiàs 2000]).

5.2 Lower and upper approximability

Given a valued tolerance relation, for each element of A we can define a “tolerance class”, that is a fuzzy set with membership function being the “tolerance degree” to the reference object. It is easy to observe that if we associate the value 1 to the non zero tolerance degree we obtain the tolerance classes introduced in section 3.

The open problem is how to define the concepts of upper and lower approximation of a set Φ . The approach we will adopt in this paper considers an approximation as a *continuous valuation*. Given a set Φ 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 equivalent of the concepts of lower and upper approximation.

Some researchers [Dubois and Prade, 1990, Greco et al. 1998, Greco et al. 2000] made similar considerations and explored the idea of combining fuzzy and rough sets. The basic idea in such approaches is to consider lower and upper approximations as fuzzy sets to which elements from the universe of discourse may more or less belong. However, our approach is different. We consider the power set 2^A as a fuzzy set both as far as the lower and upper approximation is concerned. Each subset of A may be a lower or upper approximation of Φ , but to a different degree which we denote as “*lower (upper) approximability*” (such an approach has been inspired by the work of [Kitainik, 1993]).

Therefore, we need to define for each subset Z of A its degree of lower (upper) approximation for a given set Φ with respect to a set of attributes B . For this purpose we need to translate in a functional representation the usual logical connectives of negation, conjunction etc. (in the following x, y represent 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 definitions of the properties that such a function may satisfy do not make the unanimity. Two basic properties which could be required are:

- 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 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 these two properties. In few cases where it happens other properties are not satisfied (for an excellent discussion see [Dubois et al., 1991]).

Given a set of objects $Z \subseteq A$, a subset of attributes $B \subseteq C$ and a set $\Phi \subset A$ the usual definitions of lower and upper approximations 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 “indiscernibility (tolerance, similarity etc.)” class of element z created on the basis of the subset of attributes B .

The functional translation of such definitions is straightforward. Considering that,

$$\forall x \phi(x) =_{def} T_x \phi(x);$$

$$\exists x \phi(x) =_{def} S_x \phi(x);$$

$$\Phi \subseteq \Psi =_{def} T_x(I(\mu_\Phi(x), \mu_\Psi(x)));$$

$$\Phi \cap \Psi \neq \emptyset =_{def} \exists x \phi(x) \wedge \psi(x) =_{def} S_x(T(\mu_\Phi(x), \mu_\Psi(x))) \text{ we get:}$$

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 B -lower approximability of Φ by set Z ;

$\mu_{\Phi^B}(Z)$ is the B -upper approximability of Φ by set Z ;

$\Theta_B(z)$ is the tolerance class of element z ;

T, S, I are the functions previously defined;

$R_B(z, x)$ is the membership degree of element x in the tolerance class of z (at the same time it 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 ($\hat{x} \in \{0, 1\}$).

An interesting property of lower approximability is that its value decreases as the number of attributes considered decreases. More formally:

Proposition 5.1 $\forall Z \subset A, \hat{B} \subset B \Rightarrow \mu_{\Phi_{\hat{B}}}(Z) \leq \mu_{\Phi_B}(Z)$.

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. ■

The choice of the appropriate functions for the operators N, S, T and I mainly depends on the semantics we associate to the valued tolerance relation. If the tolerance relation is valued because we associate a possibility distribution to this relation, then we will choose operators generated by the min T-norm. If we associate a probability distribution, then we will use the product T-norm, etc.

6 On the use of valued tolerance

6.1 Defining approximations

Continuation of Example 2.1 Assuming 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 (i.e., all values have the same probability to be associated to an element $x \in A$). Such an hypothesis (although sound) is arbitrary, but helps the presentation of our approach. Consider an attribute c_j in an information table $IT = (A, C)$ and associate to it the set $E_j = \{e_j^1, \dots, e_j^m\}$ of all the possible values of the attribute. Given an object $x \in A$, the probability that $c_j(x) = e_j^i$ is equal to $1/|E_j|$. Therefore, for any two objects $x, y \in A$ and an attribute c_j , if $c_j(y) = e_j^i$, the probability $R_{c_j}(x, y)$ that x is similar to y on the attribute c_j is $1/|E_j|$. Moreover, if both values are unknown, then the probability $R_{c_j}(x, y)$ that x is similar to y on the attribute c_j is $1/|E_j|^2$. Obviously, if both values are known and different, then the probability that x and y are similar is 0.

On this basis we can compute the probability that two objects are similar to each other on a set of attributes C as the joint probability R_C that the values of the two elements are the same on all such attributes:

$$R_C(x, y) = \prod_{c_j \in C} R_{c_j}(x, y)$$

It is easy to observe that if the two elements have different known values on at least one attribute such a probability is 0.

Let us consider the example 2.1 where all attributes have the same set of possible values $\{0, 1, 2, 3\}$. The probability that a_{12} is similar to a_1 is $1/4$ resulting from: $R_{c_1}(a_{12}, a_1) = 1, R_{c_2}(a_{12}, a_1) = 1, R_{c_3}(a_{12}, a_1) = 1, R_{c_4}(a_{12}, a_1) = 1/4$ The probability that a_{11} is similar to a_1 is $1/64$ resulting from: $R_{c_1}(a_{11}, a_1) = 1/4, R_{c_2}(a_{11}, a_1) = 1, R_{c_3}(a_{11}, a_1) = 1/4, R_{c_4}(a_{11}, a_1) = 1/4$. The probability that a_{11} is similar to a_{12} is $1/256$ resulting from: $R_{c_1}(a_{11}, a_{12}) = 1/4, R_{c_2}(a_{11}, a_{12}) = 1, R_{c_3}(a_{11}, a_{12}) = 1/4, R_{c_4}(a_{11}, a_{12}) = 1/16$. Applying this rule to the complete set of examples we obtain the following table 1 representing 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 object a_1 and the set of attributes C , the valued tolerance relation $R_C(a_1, x)$, $x \in U$ 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 $\Theta_C(a_1)$ of object a_1 . The reader may notice that the crisp tolerance class of object 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 $(\Theta_B(z))$ 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 degree $\mu_{\Phi_B}(Z)$ will collapse to $\{0, 1\}$ (i.e. to the usual lower approximation).

We obtain (for attributes $B \subseteq C$):

$$\mu_{\Phi_B}(Z) = \prod_{z \in Z} \prod_{x \in \Theta_B(z)} (1 - R_B(z, x) + R_B(z, x)\hat{x})$$

$$\mu_{\Phi_B}(Z) = \prod_{z \in Z} (1 - \prod_{x \in \Theta_B(z)} (1 - R_B(z, x)\hat{x}))$$

In order to explain computations in our example, consider the set Φ and as the set Z consider the object a_1 . We are going to compute the degree by which a_1 is a lower (upper) approximation of Φ . We have ($B = C$):

$$\mu_{\Phi_C}(a_1) = \prod_{x \in \Theta_C(a_1)} (1 - R_C(a_1, x) + R_C(a_1, x)\hat{x})$$

$$\mu_{\Phi_C}(a_1) = (1 - \prod_{x \in \Theta_C(a_1)} (1 - R_C(a_1, x)\hat{x}))$$

where $R_C(a_1, x)$ was previously introduced and \hat{x} takes the values 1, 1, 0, 1, 0, 0, 1, 0, 0, 1, 0, 1 (membership to set Φ).

We obtain $\mu_{\Phi_C}(a_1) = (1 - 1 - 1 \cdot 1) \cdot (1 - 1/64 + (1/64) \cdot 0) \cdot (1 - 1/4 + 1/4 \cdot 1) = 0.984$ and $\mu_{\Phi^C}(a_1) = (1 - (1 - 1 \cdot 1) \cdot (1 - (1/64) \cdot 0) \cdot (1 - (1/4) \cdot 1)) = 1$. The fact that lower approximability of Φ by a_1 is less than 1 (although a_1 belongs to Φ) is due to the presence of element a_{11} (not belonging to Φ) which could be similar to a_1 . However, since the probability that the two objects have the same values is very low (1/64), as we expect, the influence of such an element is “marginal”. In the same way we may compute $\mu_{\Phi_C}(a_{12}) = 0.73$ and $\mu_{\Phi^C}(a_{12}) = 1$, from which we can obtain $\mu_{\Phi_C}(a_1, a_{12}) = 0.984 \cdot 0.73 = 0.71$ and $\mu_{\Phi^C}(a_1, a_{12}) = 1$ and so on for any subset of U . The degrees by which single objects approximate sets Φ and Ψ are presented in Table 2.

	$\mu_{\Phi_C}(z)$	$\mu_{\Phi^C}(x)$	$\mu_{\Psi_C}(z)$	$\mu_{\Psi^C}(x)$
a_1	0.984	1	0	0.016
a_2	0	1	0	1
a_3	0	1	0	1
a_4	0.995	1	0	0.005
a_5	0	0.020	0.98	1
a_6	0	0	1	1
a_7	0.933	1	0	0.067
a_8	0	0.005	0.995	1
a_9	0	0.297	0.703	1
a_{10}	0.998	1	0	0.002
a_{11}	0	0.022	0.978	1
a_{12}	0.735	1	0	0.265

Table 2. Lower and upper approximability of Φ and Ψ by each element of A .

Let us notice that in the valued tolerance relation, unlike the previous approaches, each subset of objects Z could be a lower (upper) approximation of the given set but with a different degree. For instance, degrees for one element subsets of A to be C -lower approximation of Φ are presented in Table 2., examples of two element subsets are $\{a_1, a_4\}$ with degree $\mu_{\Phi_C}(a_1, a_4) = 0.984 \cdot 0.995 = 0.979$, $\{a_1, a_7\}$ with $\mu_{\Phi_C}(a_1, a_7) = 0.984 \cdot 0.933 = 0.918$, $\{a_1, a_{10}\}$ with $\mu_{\Phi_C}(a_1, a_{10}) = 0.984 \cdot 0.998 = 0.928$ $\{a_4, a_{10}\}$ with $\mu_{\Phi_C}(a_4, a_{10}) = 0.995 \cdot 0.998 = 0.993$. Then, an example of three element C -lower approximation of Φ is $\{a_1, a_4, a_{10}\}$ with degree $\mu_{\Phi_C}(a_1, a_4, a_{10}) = 0.984 \cdot 0.995 \cdot 0.998 = 0.974$, an example of four element approximation is $\{a_1, a_4, a_7, a_{10}\}$ with degree $\mu_{\Phi_C}(a_1, a_4, a_7, a_{10}) = 0.984 \cdot 0.995 \cdot 0.933 \cdot 0.998 = 0.909$, five element is $\{a_1, a_4, a_7, a_{10}, a_{12}\}$ with degree 0.670 , etc.

In practice, the user can be interested only in a part of these subsets which gives him a possibility to approximate the set Φ with high enough degrees. So, (s)he could express (her)his level of acceptance using a threshold λ . Such a threshold should represent the minimum value of the degree for any subset in order to be an approximation of the given set. This value generally, depends on the specificity of the analysed problem and the experience of the user.

It is possible to choose a set Z as a lower (upper) approximation of set Φ as follows:

Step 1 take all objects z for which $\mu(\Theta(z) \rightarrow \Phi) = 1$ ($\mu(\Theta(z) \cap \Phi) = 1$); intuitively these are objects which have $\mu_{\Phi_B}(z) = 1$ and $\mu_{\Phi^B}(z) = 1$;

Step 2 then add objects in a way such that $\mu(\Theta(z) \rightarrow \Phi) > k$ ($\mu(\Theta(z) \cap \Phi) > k$), (for decreasing values of k , let us say 0.99, 0.98 etc.), thus obtaining a family of sets with decreasing lower (upper) approximability $\mu_{\Phi_B}(Z)$ ($\mu_{\Phi^B}(Z)$);

Step 3 fix a minimum level λ enabling to accept a set Z as a B -lower (upper) approximation of Φ and choose the maximal set Z such that $\mu_{\Phi_B}(Z) \geq \lambda$ ($\mu_{\Phi^B}(Z) \geq \lambda$).

Let us also comment that in some situations, the user may be interested in choosing only one subset from the family of subsets having degree over the acceptance threshold λ . For instance, in step 3 the user can choose for the lower approximation, the largest set of objects Z having degree $\mu_{\Phi_B}(Z) \geq \lambda$.

The concept of reduct can also be generalised in the valued tolerance case. Given the decision table (A, C) and the partition $\mathcal{Y} = \Phi_1, \Phi_2, \dots, \Phi_n$, the subset of attributes $B' \subseteq B \subseteq C$ is a reduct iff it does not reduce the degree of lower approximation obtained with B , i.e. if Z_1, Z_2, \dots, Z_n is a family of lower approximations of $\Phi_1, \Phi_2, \dots, \Phi_n$ then $\forall_{i=1, \dots, n} Z_i \mu_{\Phi_i B}(Z_i) \leq \mu_{\Phi_i B'}(Z_i)$.

6.2 Decision rules induction

In order to induce decision rules from the decision table we may accept rules with a “credibility degree” derived from the fact that objects may be similar to the condition part of the rule only to a certain degree, besides the fact that the implication in the decision part is also uncertain. We give the following representation for a rule ρ_i :

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

where: $B \subseteq C$ is a subset of condition attributes used in the condition part of rule ρ_i , v is the value of condition attribute c_j , w 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 extent to the condition 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 $W = \{x : d(x) = w\}$. In a case of crisp relation ρ_i is a decision rule iff:

$$\forall x \in S(\rho_i) : \Theta_B(x) \subseteq W$$

Shifting in the valued case we compute a credibility degree for any rule ρ_i calculating the truth value of the previous formula which can be rewritten as:

$\forall x, y \ s(x, \rho_i) \rightarrow (R(x, y) \rightarrow W(y))$. We get:

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

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

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 .

While inducing decision rules we need to pay attention to a number of observations:

1. When choosing objects as candidates for inducing a decision rule for a class Φ it is natural to choose objects with the maximal lower approximability of set Φ .
2. However, in presence of unknown values, such lower approximability should be computed considering only attributes where the candidate object is completely described (see also Proposition 5.1). Further on, search for non-redundant sets of conditions can be performed in a similar way as in local reducts approach (see [Skowron, 1993, Komorowski et al., 1999]).
3. Finally, the user should fix a credibility threshold τ for the induced rules in order to prevent proliferation of rules considered to be “unsafe” for the classification purposes. Of course, a sensivity analysis could be computed around such a threshold. Moreover, notice that given a credibility threshold for the acceptance of rules, objects having lower approximability below such threshold cannot be used as candidates to create conditions of rules.
4. A problem remains when rules are induced from a reduced set of conditions. From a rule ρ_i with credibility $\mu(\rho_i)$, new rules could be still generated with shortest condition part, but having lower credibility still over the allowed threshold.

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

Continuation of Example 2.1.

Consider again the example of the incomplete information table used in this paper and let us come back to the probabilistic form of a valued tolerance relation introduced before. Since we have chosen for the functional representation of implication the satisfaction of De Morgan law and for T -norms the product, we get:

$$\mu(\rho_i) = \prod_{x \in S(\rho_i)} (1 - s_B(x, \rho_i) + s_B(x, \rho_i) \prod_{y \in \Theta_B(x)} (1 - \mu_{\Theta_B(x)}(y) + \mu_{\Theta_B(x)}(y)\mu_W(y)))$$

where $s_B(x, \rho_i)$ will represent now the “support” degree of element x to the rule ρ_i .

Let us fix the value of the acceptance threshold τ equal to 0.9. According to it, for the decision class Φ and all attributes C only objects a_1, a_4, a_7, a_{10} could be taken into account (see Table 2). However, if we consider only the attributes where objects are completely defined we get $\mu_{\Phi_{c_2, c_4}}(a_4) = 0, \mu_{\Phi_{c_1, c_4}}(a_7) = 0, \mu_{\Phi_{c_1}}(a_{10}) = 0.422$) thus restricting our interest only to element a_1 with $\mu_{\Phi_C}(a_1)=0.98$. Therefore, the only candidate for the rule is: $\rho_1 : (c_1 = 3) \wedge (c_2 = 2) \wedge (c_3 = 1) \wedge (c_4 = 0) \rightarrow (d = \Phi)$

In order to compute the credibility degree of this rule we need the values of $s(x, \rho_1)$ and $I_y(\mu_{\Theta(x)}(y), \mu_{\Phi}(y))$ which are presented in the following table:

	$s(x, \rho_1)$	$I_y(\mu_{\Theta(x)}(y), \mu_{\Phi}(y))$
a_1	1	0.984
a_{11}	0.0156	0
a_{12}	0.25	0.735

The first column indicates the tolerance degree of each element with respect to the condition part of the rule, while the second column indicates the cumulative implication degree for each element in the tolerance class of the element in the line. Objects a_2, \dots, a_{10} have degree $s(x, \rho_i) = 0$ so we can omit them since they reduce to value 1. Thus, we get $\mu(\rho_1) = (1 - 1 + 1 * 0.984) * (1 - 0.156 + 0.156 * 0) * (1 - 0.25 + 0.25 * 0.735) = 0.905$

However, the condition part of rule ρ_1 is redundant as the rule could be transformed to $\rho_1 : (c_1 = 3) \wedge (c_3 = 1) \wedge (c_4 = 0) \rightarrow (d = \Phi)$ with degree $\mu(\rho_1) = 0.905$. This rule is supported by the set of objects $S(\rho_1) = \{a_1, a_{11}, a_{12}\}$. It is not possible to reduce it more as it leads to a rapid decrease of the credibility degree.

In the case of set Ψ we can start from considering objects a_5, a_6, a_8, a_{11} . Again, computing $\mu_{\Psi_B}(x)$, for B being the set of attributes where x is completely described, we restrict our interest only to object a_6 . So we take it as a candidate for the rule and we get: $\rho_2 : (c_1 = 2) \wedge (c_2 = 3) \wedge (c_3 = 2) \wedge (c_4 = 1) \rightarrow (d = \Psi)$ with degree $\mu(\rho_2) = 1.0$. The condition part is not minimal and can be transformed to the following:

$\rho_2 : (c_1 = 2) \wedge (c_2 = 3) \wedge (c_4 = 1) \rightarrow (d = \Psi)$ with degree $\mu(\rho_2) = 1.0$ and supporting object a_6 .

The decision rule ρ_2 could have more reduced condition parts if one does not want to maintain strictly $\mu_{\Psi_B}(x)=1$. We can relax this requirement and look for more reduced rules if their credibility levels are not lower than a fixed threshold τ . Proceeding in this way, we can induce from the decision table more decision rules:

$\rho'_2 : (c_1 = 2) \wedge (c_4 = 1) \rightarrow (d = \Psi)$ with $\mu(\rho'_2) = 0.931$

$\rho'_3 : (c_2 = 3) \wedge (c_3 = 2) \wedge (c_4 = 1) \rightarrow (d = \Psi)$ with $\mu(\rho'_3) = 0.969$

where the new rules ρ'_2 and ρ'_3 are obtained relaxing credibility threshold for rule ρ_2 to 0.9.

Then, if the user may relax the credibility threshold to 0.87, then the third rule can be replaced by two others:

$\rho''_3 : (c_2 = 3) \wedge (c_4 = 1) \rightarrow (d = \Psi)$ with $\mu(\rho''_3) = 0.879$

$\rho''_4 : (c_2 = 3) \wedge (c_3 = 2) \rightarrow (d = \Psi)$ with $\mu(\rho''_4) = 0.879$

As one can notice we obtained more rules than in the approach based on a simple tolerance relation. On the other hand, this set is more limited than in the case of absent value semantics.

7 Computational experiments

In the valued tolerance relation approach the user can control the produced approximations and rule sets by some threshold levels. In general, such thresholds should reflect

(her)his knowledge of the problem and data. On the other hand, it is interesting to perform a computational study on some larger data sets where we can change systematically threshold values and observe their influence on lower, upper approximations and decision rules. We have undertaken such experiments.

In the case of lower and upper approximations we are using the same value of threshold λ for both approximations of each decision class at the same moment. Moreover we choose always the largest sets of elements Z as lower approximations.

We used 4 real life public data sets coming from machine learning repository at University of California [Blake and Merz, 1998]. They contained missing values - some data were slightly modified to obtain non-numerical attributes. One additional real-life data set, named *nbuses*, was coming from known rough sets application [Słowiński et al., 1996]. Finally, we created an artificial data set *Art1* consisting of two classes with non empty boundary region (each part having approximately the same cardinality of objects). This data set was modified by introducing 10%, 14%, 18%, 22%, 26% and 30% randomly chosen missing values. The characteristics of used data sets are summarised in Table 3.

Data set	Number of examples	Number of attributes	Number of decision classes	Number of missing values
Art1	50	5	2	0%
Art2	50	5	2	10%
Art3	50	5	2	14%
Art4	50	5	2	18%
Art5	50	5	2	22%
Art6	50	5	2	26%
Art7	50	5	2	30%
nbuses	76	8	2	10%
bridges	108	8	6	6%
credit	139	9	2	1%
breast cancer	285	8	2	6%
hungarian	183	12	2	18%

Table 3. Data sets considered

Further on, we decided to performed a comparative study of using the valued tolerance relation and tolerance relation applied to the same data. Results are summarised in the Tables 4,5 and 6. Tables 4 and 5 show the sum of cardinalities of approximations for each decision classes (expressed in numbers of objects).

Data set	tolerance relation	valued tolerance relation				
		1.0	0.9	0.8	0.7	0.6
Art1	35	35	35	35	35	35
Art2	31	31	33	33	33	36
Art3	28	28	32	33	36	36
Art4	26	26	30	32	35	35
Art5	20	20	26	31	32	34
Art6	17	17	23	27	32	33
Art7	14	14	19	24	28	30
nbuses	73	73	74	76	76	76
bridges	60	60	68	72	78	80
credit	117	117	118	120	122	122
breast cancer	250	250	261	267	270	272
hungarian	168	168	183	183	183	183

Table 4. Cardinalities of lower approximations for each decision class; for valued tolerance relation various levels of threshold λ are listed

Data set	tolerance relation	valued tolerance relation				
		1.0	0.9	0.8	0.7	0.6
Art1	65	65	65	65	65	65
Art2	69	52	53	53	53	54
Art3	72	50	50	51	51	51
Art4	74	50	50	51	51	52
Art5	80	50	51	51	51	53
Art6	83	50	50	51	51	51
Art7	86	50	50	50	51	52
nbuses	79	76	76	76	76	76
bridges	190	133	133	134	134	134
credit	161	152	152	153	153	153
breast cancer	320	285	285	285	285	285
hungarian	198	183	183	183	183	183

Table 5. Cardinalities of upper approximations for each decision class; for valued tolerance relation various levels of threshold λ are listed

Data set	tolerance relation	valued tolerance relation					
		1.0	0.95	0.9	0.85	0.8	0.7
Art1	15	15	-	15	-	15	15
Art2	6	6	-	10	-	15	15
Art3	3	1	-	9	-	10	12
Art4	2	1	-	8	-	12	14
Art5	0	0	-	6	-	9	9
Art6	0	0	-	4	-	6	7
Art7	0	0	-	3	-	5	5
nbuses	96	142	228	242	251	-	-
bridges	110	84	120	148	170	-	-
credit	389	347	435	501	567	-	-
breast cancer	791	817	875	902	934	-	-
hungarian	594	562	598	653	676	-	-

Table 6. Number of decision rules induced (“-” stands for a non performed experiment)

Let us shortly summarise and comment the obtained results. Changing value of the threshold λ (the limit imposed in order to accept a set as a lower approximation) increased the cardinalities of lower approximations while upper approximation were unchanged (or slightly changed) in most cases. It is interesting to notice that almost for all data sets the increases of lower approximation cardinality was greatest for high levels of the threshold - intervals 1.0-0.9 or 1.0-0.8 depending on the data set. Moreover, the increase of the cardinality of the lower approximation was higher for data with a large number of missing values.

Table 6 presents results of the rule induction. For both compared approaches we induced all rules, i.e. certain rules in the case of the tolerance relation and all rules with credibility over the acceptance threshold τ . Let us comment that in the case of data *Art1- Art7* we used values of threshold τ similar to those of threshold λ . For the other data, the structure of decision rule sets was changing faster for higher values of τ , so additional threshold values 0.95 and 0.85 were added.

Decreasing the threshold τ increased the number of induced rules. However, large modifications of this threshold may produce too many rules, especially for real-life data.

The comparison of using tolerance and valued tolerance relation showed that results based on the valued relation were more informative. For thresholds lower than 1.0 more objects were assigned to lower approximations. Sums of cardinalities of upper approximations created by the valued tolerance relation were always smaller than in a case of tolerance relation. Moreover, choosing a proper acceptance level for the credibility of rules may result in increasing the number of induced rules. It seems to be particular useful for data sets with large number of missing values (e.g. *Art3 – Art7*) where tolerance relation is very restrictive. The obtained results show once again that the valued tolerance relation is more flexible and informative approach than the crisp tolerance approach.

8 Conclusions

Rough sets theory has been conceived under the implicit hypothesis that all objects in a universe can be evaluated using a given set of attributes. However, in some situations

attribute values may be unknown either because they are not available at a given time or because definitely impossible to obtain. In our paper we introduce two different semantics in order to distinguish such situations. “Missing values” semantics implies that not available information could always become available. In order to make “safe” classifications and rule induction we might consider that such missing values could be similar to every possible values of the given attribute. The tolerance relation (which is reflexive and symmetric, but not transitive) captures in a formal way such a case. “Absent values” semantics implies that unknown attribute values are not available and cannot be used in comparing objects. So, classification and rule induction should be performed with the existing information since the absent values could never become available. Similarity relations (which in this case are reflexive and transitive, but not symmetric) are introduced in our paper in order to formalise such an idea. We demonstrate that our approach always lead to more informative results than the tolerance relation based approach (although it is less safe).

A third approach is based on the use of the valued tolerance relation. The valued relation could appear for several reasons not only because of the non available information. In fact, the approach presented here has a more general validity. However, in this paper we limit ourselves to discussing case of the missing attribute values. The functional extension of the concepts of the upper and lower approximation is introduced so that a degree of lower (upper) approximation can be associated to any subset of the universe. Further on, such a functional extension enables to compute a credibility degree for any induced rule. The logical analysis and computational experiments show that with such an approach it is possible to generate more decision rules (and/or simpler ones) than for the simple tolerance approach. Fixing a credibility threshold, the user can prevent the proliferation of “unsafe” decision rules.

Further research directions include, but are not limited to:

- a further analysis of rules properties and induction algorithms for the non symmetric similarity relation based approach,
- an analysis of the non-monotonic behaviour of the classification obtained for the non symmetric similarity relation based approach (e.g. what could happen 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,

Acknowledgements

Research on this paper was partly supported from the KBN grant no. 8T11F 006 19 and the POLONIUM project. An anonymous referee comments helped to improve the paper.

References

- [Blake and Merz, 1998] Blake, C.L., Merz, C.J., UCI Repository of machine learning databases [<http://www.ics.uci.edu/~mlearn/MLRepository.html>]. Irvine, CA: University of California, Department of Information and Computer Science, 1998.

- [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.,1999a] Greco S., Matarazzo B., Slowinski R., Zanakis S., Rough set analysis of data tables with missing values. in *Proc. 5th International Conference of the Decision Sciences Institute*, Athens, Greece, July 4-7, 1999, 1359-1362.
- [Greco et al.,1999b] Greco S., Matarazzo B., Slowinski R., Handling missing values in rough set analysis of multi-attribute and multi-criteria decision problems, in A. Skowron and N. Zhong (eds.), *Proc. 7th Workshop New Directions in Rough Sets, Data Mining and Granular-Soft Computing* Yamaguchi, Japan, LNAI 1711, Springer-Verlag, Berlin, 146-157.
- [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.
- [Grzymala, 1992] Grzymala-Busse J. W. LERS - A system for learning from examples based on rough sets, in Słowiński R. (ed.), *Intelligent Decision Support. Handbook of Applications and Advances of the Rough Sets Theory*, Kluwer Academic Publishers, 1992, 3–18.
- [Grzymala et al., 1999] Grzymala-Busse J. W., Grzymala-Busse W., Goodwin L.K., A closest fit approach to missing attribute values, in Zhong N., Skowron A., Ohsuga S. (eds.) *Proc. 7th Workshop New Direction in Rough Sets, Data Mining, and Granular-Soft Computing* Yamaguchi, Japan, Springer Verlag LNAI 1711, 1999, 405-414.
- [Grzymala et al. 2000] Grzymala-Busse J. W., M. Hu, A Comparison of Several Approaches to Missing Attribute Values in Data Mining, in *Proceedings of the 2nd Int. Conference on Rough Sets and New Trends in Computing*, Banff 2000, 340-347.
- [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: tutorial, in Pal S.K., Skowron A. (eds.) *Rough Fuzzy Hybridization. A new trend in decision making*, Springer Verlag, Singapore, 1999, 3–98.
- [Kryszkiewicz, 1995] 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.
- [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, 1998, 422–450.
- [Pawlak, 1981] Pawlak Z., Rough sets. *Int. J. Computer and Information Sci.*, 11, 1982, 341–356.
- [Pawlak, 1991] Pawlak Z., *Rough sets. Theoretical aspects of reasoning about data*. Kluwer Acad. Publ., Dordrecht, 1991.
- [Skowron, 1993] 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.
- [Słowiński and Stefanowski, 1989] Słowiński R., Stefanowski J., Rough classification in incomplete information systems, *Math. Computing Modelling*, 12 (10/11), 1989, 1347–1357.
- [Słowiński et al., 1996] Słowiński R., Stefanowski J., Susmaga R., Rough analysis of attribute dependencies in technical diagnostics, in Tsumoto S. (ed.) *Proceedings of the Fourth International Workshop on Rough Sets, Fuzzy Sets and Machine Discovery*, Tokyo Nov. 6-8 1996, 384-291.
- [Słowiński et al., 2000] Słowiński R., Stefanowski J., Greco, S., Matarazzo B., Rough sets processing of inconsistent information. *Control and Cybernetics*, 29, 2000, no. 1, 379-404
- [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] Słowiński R., Vanderpooten D., A generalized definition of rough approximations based on similarity, *IEEE Transactions on Data and Knowledge Engineering*, 12 (2000), 331 - 336.
- [Stefanowski and Tsoukiàs 2000] Stefanowski J., Tsoukiàs A., Valued tolerance and decision rules, in *Proceedings of the 2nd Int. Conference on Rough Sets and New Trends in Computing*, Banff 2000, 180-187.
- [Stefanowski, 1998] 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.

[Toumi, 1996] Toumi Kh., Aide à la décision dans le cadre de la problématique de description: une approche inductive pour décrire et expliquer, Phd Thesis, Université Paris Dauphine, 1996.