Rough Set Strategies to Data with Missing Attribute Values

Jerzy W. Grzymala-Busse
Department of Electrical Engineering and Computer Science
University of Kansas, Lawrence, KS 66045, USA
Jerzy@ku.edu
and
Institute of Computer Science
Polish Academy of Sciences, 01-237 Warsaw, Poland

Abstract

In this paper we assume that a data set is presented in the form of the incompletely specified decision table, i.e., some attribute values are missing. Our next basic assumption is that some of the missing attribute values are lost (e.g., erased) and some are "do not care" conditions (i.e., they were redundant or not necessary to make a decision or to classify a case). Incompletely specified decision tables are described by characteristic relations, which for completely specified decision tables are reduced to the indiscernibility relation. It is shown how to compute characteristic relations using an idea of block of attribute-value pairs, used in some rule induction algorithms, such as LEM2. Moreover, the set of all characteristic relations for a class of congruent incompletely specified decision tables, defined in the paper, is a lattice. Three definitions of lower and upper approximations are introduced. Finally, it is shown that the presented approach to missing attribute values may be used for other kind of missing attribute values than lost values and "do not care" conditions.

1. Introduction

Usually all ideas of rough set theory are explored using decision tables as a starting point [10, 11]. The decision table describes cases (also called examples or objects) using attribute values and a decision. Attributes are independent variables while the decision is a dependent variable. In the majority of papers on rough set theory it is assumed that the information is complete, i.e., that for all cases all attribute values and decision values are specified. Such a decision table is said to be completely specified.

In practice, however, input data, presented as decision tables, may have missing attribute and decision values, i.e., decision tables are incompletely specified. Since our main concern is learning from examples, and an example with a missing decision value, (i.e., not classified) is useless, we will assume that only attribute values may be missing.

There are two main reasons why an attribute value is missing: either the value was lost (e.g., was erased) or the value was not important. In the former case attribute value was useful but currently we have no access to it. In the latter case the value does not matter, so such values are also called "do not care" conditions. In practice it means that originally the case was classified (the decision value was assigned) in spite of the fact that the attribute value was not given, since the remaining attribute values were sufficient for such a classification or to make a decision. For example, a test, represented by that attribute, was redundant.

The first rough set approach to missing attribute values, when all missing values were lost, was described in 1997 in [7], where two algorithms for rule induction, LEM1 and LEM2, modified to deal with such missing attribute values, were presented. In 1999 this approach was extensively described in [13], together with a modification of the original idea in the form of a valued tolerance based on a fuzzy set approach.

The second rough set approach to missing attribute values, in which the missing attribute value is interpreted as a "do not care" condition, was used for the first time in 1991 [4]. A method for rule induction was introduced in which each missing attribute value was replaced by all possible values. This idea was further developed and furnished with theoretical properties in 1995 [8].

In this paper a more general rough set approach to missing attribute values is presented. In this approach, in the same decision table, some missing attribute values are assumed to be lost and some are "do not care" conditions. A simple method for computing a characteristic relation describing the decision table with missing attribute values of either of these two types is presented. The characteristic relation for a completely specified decision table is reduced to the ordinary indiscernibility relation. It is shown that

the set of all characteristic relations, defined by all possible decision tables with missing attribute values being one of the two types, together with two defined operations on relations, forms a lattice.

Furthermore, three different definitions of lower and upper approximations are introduced. Some of these definitions are better suited for rule induction. Examples of rules induced from incompletely specified decision tables are provided. The paper ends up with a discussion of other approaches to missing attribute values.

2. Missing attribute values and characteristic relations

In the sequel we will assume that information about any phenomena are presented by *decision tables*. Rows of the table are labeled by *cases* and columns by variables. The set of all cases will be denoted by U. Independent variables are called *attributes* and a dependent variable is called a *decision* and is denoted by d. The set of all attributes will be denoted by A. An example of such a decision table is presented in Table 1.

Table 1. An example of a completely specified decision table

	Attributes			Decision
	Location	Basement	Fireplace	Value
1	good	yes	yes	high
2	bad	no	no	small
3	good	no	no	medium
4	bad	yes	no	medium
5	good	no	yes	medium

Obviously, any decision table defines a function ρ that maps the set of ordered pairs (case, attribute) into the set of all values. For example, $\rho(1, \text{Location}) = \text{good}$.

Rough set theory is based on the idea of an indiscernibility relation. Let B be a nonempty subset of the set A of all attributes. The indiscernibility relation IND(B) is a relation on U defined for $x, y \in U$ as follows

$$(x, y) \in \text{IND}(B) \text{ if and only if } \rho(x, a) = \rho(y, a)$$
$$\text{for all } a \in B.$$

For completely specified decision tables the indiscernibility relation IND(B) is an equivalence relation. Equivalence classes of IND(B) are called *elementary sets* of B. For example, for Table 1, elementary sets of IND({Location, Basement}) are {1}, {2}, {3, 5} and {4}.

Function ρ describing Table 1 is completely specified (total). In practice, input data for data mining are frequently affected by missing attribute values. In other words, the corresponding function ρ is incompletely specified (partial).

In the sequel we will assume that all decision values are specified, i.e., are not missing. Also, we will assume that all missing attribute values are denoted either by "?" or by "*", lost values will be denoted by "?", "do not care" conditions will be denoted by "*". Additionally, we will assume that for each case at least one attribute value is specified.

Incompletely specified tables are described by characteristic relations instead of indiscernibility relations. An example of an incompletely specified table is presented in Table 2, where all missing attribute values are lost.

Table 2. An example of an incompletely specified decision table, in which all missing attribute values are lost

	Attributes			Decision
	Location	Basement	Fireplace	Value
1	good	yes	yes	high
2	bad	?	no	small
3	good	no	?	medium
4	bad	yes	no	medium
5	?	?	yes	medium

For decision tables, in which all missing attribute values are lost, a special characteristic relation was defined by J. Stefanowski and A. Tsoukias in [13], see also [12, 14]. In this paper that characteristic relation will be denoted by LV(B), where B is a nonempty subset of the set A of all attributes. For $x, y \in U$ characteristic relation LV(B) is defined as follows:

$$(x, y) \in \text{LV}(B) \text{ if and only if } \rho(x, a) = \rho(y, a)$$
$$\text{for all } a \in B \text{ such that } \rho(x, a) \neq ?.$$

For any case x, the characteristic relation LV(B) may be presented by the *characteristic set* $I_B(x)$, where

$$I_B(x) = \{y \mid (x, y) \in \text{LV}(B)\}.$$

For Table 2, characteristic sets $I_A(x)$, where $x \in U$, are the following sets:

$$I_A(1) = \{1\},$$
$$I_A(2) = \{2, 4\},$$
$$I_A(3) = \{3\},$$
$$I_A(4) = \{4\}, \text{ and}$$

$$I_A(5) = \{1, 5\}.$$

For any decision table in which all missing attribute values are lost, characteristic relation $LV(B)$ is reflexive, but—in general—does not need to be symmetric or transitive.

Another example of a decision table with all missing attribute values, this time with only "do not care" conditions, is presented in Table 3.

Table 3. An example of an incompletely specified decision table, in which all missing attribute values are "do not care" conditions

	Attributes			Decision
	Location	Basement	Fireplace	Value
1	good	yes	yes	high
2	bad	*	no	small
3	good	no	*	medium
4	bad	yes	no	medium
5	*	*	yes	medium

Table 4. An example of an incompletely specified decision table, in which some missing attribute values are lost and some are "do not care" conditions

	Attributes			Decision
	Location	Basement	Fireplace	Value
1	good	yes	yes	high
2	bad	?	no	small
3	good	no	?	medium
4	bad	yes	no	medium
5	*	*	yes	medium

For decision tables where all missing attribute values are "do not care" conditions a special characteristic relation, in this paper denoted by $DCC(B)$, was defined by M. Kryszkiewicz in [8], see also, e.g., [9]. For $x, y \in U$ characteristic relation $DCC(B)$ is defined as follows:

$(x, y) \in DCC(B)$ if and only if $\rho(x, a) = \rho(y, a)$ or $\rho(x, a) = *$ or $\rho(y, a) = *$ for all $a \in B$.

Similarly, for a case x, the characteristic relation $DCC(B)$ may be presented by the characteristic set $J_B(x)$, where

$$J_B(x) = \{y \mid (x, y) \in DCC(B)\}.$$

For Table 3, characteristic sets $J_A(x)$, where $x \in U$, are the following sets:

$$J_A(1) = \{1, 5\},$$
$$J_A(2) = \{2, 4\},$$
$$J_A(3) = \{3, 5\},$$
$$J_A(4) = \{2, 4\}, \text{ and}$$
$$J_A(5) = \{1, 3, 5\}.$$

Relation $DCC(B)$ is reflexive and symmetric but—in general—not transitive.

Table 4 presents a more general case, a decision table with missing attribute values of both types: lost values and "do not care" conditions.

In a similar way we may define a characteristic relation $R(B)$ on U for an incompletely specified decision table with both types of missing attribute values: lost values and "do not care" conditions:

$(x, y) \in R(B)$ if and only if $\rho(x, a) = \rho(y, a)$ or $\rho(x, a) = *$ or $\rho(y, a) = *$ for all $a \in B$ such that $\rho(x, a) \neq ?$,

where $x, y \in U$ and B is a nonempty subset of the set A of all attributes. For a case x, the characteristic relation $R(B)$ may be also presented by its characteristic set $K_B(x)$, where

$$K_B(x) = \{y \mid (x, y) \in R(B)\}.$$

For Table 4, characteristic sets $K_A(x)$, where $x \in U$, are the following sets:

$$K_A(1) = \{1, 5\},$$
$$K_A(2) = \{2, 4\},$$
$$K_A(3) = \{3, 5\},$$
$$K_A(4) = \{4\}, \text{ and}$$
$$K_A(5) = \{1, 5\}.$$

Obviously, characteristic relations $LV(B)$ and $DCC(B)$ are special cases of the characteristic relation $R(B)$. For a completely specified decision table, the characteristic relation $R(B)$ is reduced to $IND(B)$. The characteristic relation $R(B)$ is reflexive but—in general—does not need to be symmetric or transitive.

3. Computing characteristic relations

The characteristic relation $R(B)$ is known if we know characteristic sets $K(x)$ for all $x \in U$. Thus we may concentrate on computing characteristic sets $K(x)$. We need a few definitions. For completely specified decision tables if $t = (a, v)$ is an attribute-value pair then a *block* of t, denoted $[t]$, is a set of all cases from U that for attribute a have value v [1, 5]. For incompletely specified decision

tables the definition of a block of an attribute-value pair must be modified. If an attribute a there exists a case x such that $\rho(x, a) = ?$, i.e., the corresponding value is lost, then the case x is not included in the block $[(a, v)]$ for any value v of attribute a. If for an attribute a there exists a case x such that the corresponding value is a "do not care" condition, i.e., $\rho(x, a) = *$, then the corresponding case x should be included in blocks $[(a, v)]$ for all values v of attribute a. The characteristic set $K_B(x)$ is the intersection of blocks of attribute-value pairs (a, v) for all attributes a from B for which $\rho(x, a)$ is specified and $\rho(x, a) = v$.

For decision table from Table 4,

$$[(\text{Location, good})] = \{1, 3, 5\},$$

$$[(\text{Location, bad})] = \{2, 4, 5\},$$

$$[(\text{Basement, yes})] = \{1, 4, 5\},$$

$$[(\text{Basement, no})] = \{3, 5\},$$

$$[(\text{Fireplace, yes})] = \{1, 5\}, \text{ and}$$

$$[(\text{Fireplace, no})] = \{2, 4\}.$$

Thus

$$K_A(1) = \{1, 3, 5\} \cap \{1, 4, 5\} \cap \{1, 5\} = \{1, 5\},$$

$$K_A(2) = \{2, 4, 5\} \cap \{2, 4\} = \{2, 4\},$$

$$K_A(3) = \{1, 3, 5\} \cap \{3, 5\} = \{3, 5\},$$

$$K_A(4) = \{2, 4, 5\} \cap \{1, 4, 5\} \cap \{2, 4\} = \{4\}, \text{ and}$$

$$K_A(5) = \{1, 5\}.$$

Table 5. Decision table indistinguishable from decision table presented in Table 6

	Attributes			Decision
	Location	Basement	Fireplace	Value
1	good	yes	yes	high
2	bad	*	no	small
3	good	no	*	medium
4	bad	yes	no	medium
5	?	*	yes	medium

4. Lattice of characteristic relations

For the sake of simplicity, in this section all characteristic relations will be defined for the entire set A of attributes instead of its subset B and we will write R instead of $R(A)$. By the same token, in characteristic sets $K_A(x)$, the subscript A will be omitted.

Two decision tables with the same set U of all cases, the same attribute set A, the same decision d, and the same specified attribute values will be called *congruent*. Thus, two congruent decision tables may differ only by missing attribute values * and ?. Decision tables from Tables 2, 3, and 4 are all pairwise congruent.

Two congruent decision tables that have the same characteristic relations will be called *indistinguishable*. For example, decision tables, presented in Tables 5 and 6 are indistinguishable, both have the same characteristic relation with the following characteristic sets:

$$K(1) = \{1\},$$

$$K(2) = \{2, 4\},$$

$$K(3) = \{3\},$$

$$K(4) = \{2, 4\}, \text{ and}$$

$$K(5) = \{1, 3, 5\}.$$

Table 6. Decision table indistinguishable from decision table presented in Table 5

	Attributes			Decision
	Location	Basement	Fireplace	Value
1	good	yes	yes	high
2	bad	*	no	small
3	good	no	*	medium
4	bad	yes	no	medium
5	*	?	yes	medium

On the other hand, if the characteristic relations for two congruent decision tables are different, the decision tables will be called *distinguishable*. Obviously, there is 2^n congruent decision tables, where n is the total number of all missing attribute values in a decision table.

Let D_1 and D_2 be two congruent decision tables, let R_1 and R_2 be their characteristic relations, and let $K_1(x)$ and $K_2(x)$ be their characteristic sets for some $x \in U$, respectively. We say that $R_1 \leq R_2$ if and only if $K_1(x) \subseteq K_2(x)$ for all $x \in U$. We will use also notation that $D_1 \leq D_2$.

For two congruent decision tables D_1 and D_2, $D_1 \leq D_2$ if for every missing attribute value "?" in D_2, say $\rho_2(x, a)$, the missing attribute value for D_1 is also "?", i.e., $\rho_1(x, a)$, where ρ_1 and ρ_2 are functions defined by D_1 and D_2, respectively.

Two subsets of the set of all congruent decision tables are special: set E of n decision tables such that every decision table from E has exactly one missing attribute

value "?" and all remaining attribute values equal to "*" and the set F of n decision tables such that every decision table from E has exactly one missing attribute value "*" and all remaining attribute values equal to "?". In our example, decision tables presented in Tables 5 and 6 belong to the set E.
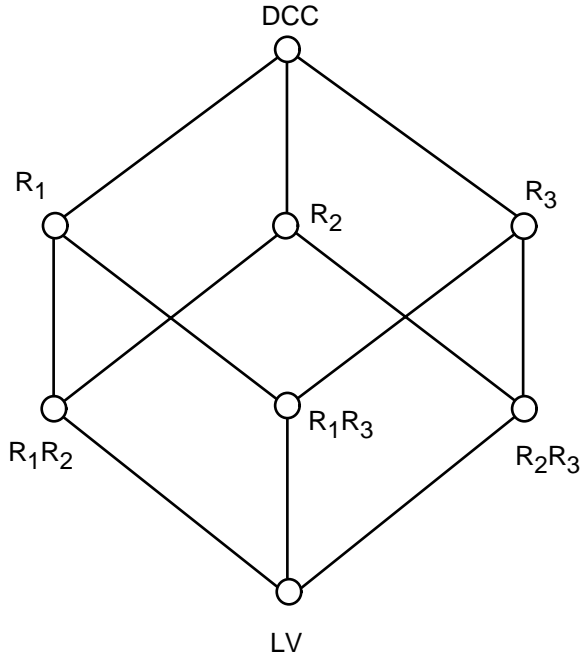

Figure 1. Diagram of the lattice of all characteristic relations

Let G be the set of all characteristic relations associated with the set E and let H be the set of all characteristic relations associated with the set F. In our example, the set G has three elements, say R_1, R_2, and R_3, defined by the following family of characteristic sets K_1, K_2, and K_3, respectively:

$K_1(1) = \{1\}$, $K_2(1) = \{1, 5\}$, $K_3(1) = \{1, 5\}$,
$K_1(2) = \{2, 4\}$, $K_2(2) = \{2, 4\}$, $K_3(2) = \{2, 4\}$,
$K_1(3) = \{3\}$, $K_2(3) = \{3, 5\}$, $K_3(3) = \{3, 5\}$,
$K_1(4) = \{2, 4\}$, $K_2(4) = \{4\}$, $K_3(4) = \{2, 4\}$,
$K_1(5) = \{1, 3, 5\}$, $K_2(5) = \{1, 3, 5\}$, $K_3(5) = \{1, 5\}$,

where R_1 is the characteristic relation of the decision table D_1 from Table 5, R_2 is the characteristic relation of the decision table D_2 congruent with D_1 and with $\rho_2(2,$ Basement$) = ?$ and all remaining missing attribute values equal to "*", and R_3 is the characteristic relation of the decision table D_3 congruent with D_1 and with $\rho_3(3,$ Fireplace$) = ?$ and all remaining missing attribute values equal to "*".

Let D and D' be two congruent decision tables with characteristic relations R and R', and with characteristic sets $K(x)$ and $K'(x)$, respectively, where $x \in U$. We define a characteristic relation $R + R'$ as defined by characteristic sets $K(x) \cup K'(x)$, for $x \in U$, and a characteristic relation $R \cdot R'$ as defined by characteristic sets $K(x) \cap K'(x)$. The set of all characteristic relations for the set of all congruent tables, together with operations $+$ and $\cdot$, is a lattice L (i.e., operations $+$ and $\cdot$ satisfy the four postulates of idempotent, commutativity, associativity, and absorption laws [2]).

Each characteristic relation from L can be represented (using the lattice operations $+$ and $\cdot$) in terms of characteristic relations from G (and, similarly for H). Thus G and H are sets of generators of L. In our example, set G, together with the operation $\cdot$, generates all characteristic relations from L, except for DCC, which may be computed as $R_1 + R_2$, for any two distinct characteristic relations R_1 and R_2 from G. Similarly, set $H = \{R_1 \cdot R_2, R_1 \cdot R_3, R_2 \cdot R_3\}$, together with the operation $+$, generates all characteristic relations from L, except for LV, which may be computed as $R_1 \cdot R_2$, for any two distinct characteristic relations R_1 and R_2 from H.

A characteristic relation R_1 covers another characteristic relation R_2 if and only if $R_1 \neq R_2$, $R_1 \geq R_2$, and there is no R with $R_1 \neq R \neq R_2$ and $R_1 \geq R \geq R_2$. A *diagram* of the lattice L represents elements of L by circles; the characteristic relation R_1 will be placed higher than R_2 if and only if $R_1 \geq R_2$, the circles represented by R_1 and R_2 are connected by a straight line if and only if R_1 covers R_2. The diagram of the lattice of all characteristic relations for our example is presented by Figure 1.

5. Lower and upper approximations

For completely specified decision tables lower and upper approximations are defined on the basis of the indiscernibility relation. An equivalence class of $IND(B)$ containing x is denoted by $[x]_B$. Any finite union of elementary sets of B is called a *B-definable set*. Let U be the set of all cases, called an *universe*. Let X be any subset of U. The set X is called *concept* and is usually defined as the set of all cases defined by specific value of the decision. In general, X is not a B-definable set. However, set X may be approximated by two B-definable sets, the first one is called a *B-lower approximation of X*, denoted by $\underline{B}X$ and defined as follows

$$\{x \in U \mid [x]_B \subseteq X \}.$$

The second set is called a *B-upper approximation of X*, denoted by $\overline{B}X$ and defined as follows

$$\{x \in U \mid [x]_B \cap X \neq \emptyset \}.$$

The B-lower approximation of X is the greatest B-definable set, contained in X. The B-upper approximation of X is the least B-definable set containing X.

For incompletely specified decision tables lower and upper approximations may be defined in a few different ways. In this paper we suggest three different definitions. Again, let X be a concept, let B be a subset of the set A of all attributes, and let $R(B)$ be the characteristic relation of the incompletely specified decision table with characteristic sets $K(x)$, where $x \in U$. Our first definition uses a similar idea as in the previous articles on incompletely specified decision tables [8, 9, 12–14], i.e., lower and upper approximations are sets of singletons from the universe U satisfying some properties. We will call these definitions *singleton*. A singleton B-lower approximation of X is defined as follows:

$$\underline{B}X = \{x \in U \mid K_B(x) \subseteq X \}.$$

A singleton B-upper approximation of X is

$$\overline{B}X = \{x \in U \mid K_B(x) \cap X \neq \emptyset \}.$$

In our example of the decision presented in Table 2 let us say that $B = A$, hence $R(A) = \mathrm{LV}(A)$. Then the singleton A-lower and A-upper approximations are:

$$\underline{A}\{1\} = \{1\},$$
$$\underline{A}\{2\} = \emptyset,$$
$$\underline{A}\{3, 4, 5\} = \{3, 4\},$$
$$\overline{A}\{1\} = \{1, 5\},$$
$$\overline{A}\{2\} = \{2\},$$
$$\overline{A}\{3, 4, 5\} = \{2, 3, 4, 5\}.$$

In our example of the decision presented in Table 3 let us say that $B = A$, hence $R(A) = \mathrm{DCC}(A)$. Then the singleton A-lower and A-upper approximations are:

$$\underline{A}\{1\} = \emptyset,$$
$$\underline{A}\{2\} = \emptyset,$$
$$\underline{A}\{3, 4, 5\} = \{3\}$$
$$\overline{A}\{1\} = \{1, 5\},$$
$$\overline{A}\{2\} = \{2, 4\},$$
$$\overline{A}\{3, 4, 5\} = \{1, 2, 3, 4, 5\} = U.$$

The second definition uses another idea: lower and upper approximations are unions of characteristic sets, subsets of U. We will call these definitions *subset*. A subset B-lower approximation of X is defined as follows:

$$\underline{B}X = \cup\{K_B(x) \mid x \in U, K_B(x) \subseteq X \}.$$

A subset B-upper approximation of X is

$$\overline{B}X = \cup\{K_B(x) \mid x \in U, K_B(x) \cap X \neq \emptyset \}.$$

In our example of the decision table presented in Table 2 and $R(A) = \mathrm{LV}(A)$, the subset A-lower and A-upper approximations are

$$\underline{A}\{1\} = \{1\},$$
$$\underline{A}\{2\} = \emptyset,$$
$$\underline{A}\{3, 4, 5\} = \{3, 4\},$$
$$\overline{A}\{1\} = \{1, 5\},$$
$$\overline{A}\{2\} = \{2, 4\},$$
$$\overline{A}\{3, 4, 5\} = \{1, 2, 3, 4, 5\} = U.$$

In our example of the decision table presented in Table 3 and $R(A) = \mathrm{DCC}(A)$, the subset A-lower and A-upper approximations are

$$\underline{A}\{1\} = \emptyset,$$
$$\underline{A}\{2\} = \emptyset,$$
$$\underline{A}\{3, 4, 5\} = \{3, 5\}$$
$$\overline{A}\{1\} = \{1, 3, 5\},$$
$$\overline{A}\{2\} = \{2, 4\},$$
$$\overline{A}\{3, 4, 5\} = \{1, 2, 3, 4, 5\} = U.$$

The next possibility is to modify the subset definition of upper approximation by replacing the universe U from the previous definition by a concept X. A *concept B-lower approximation* of the concept X is defined as follows:

$$\underline{B}X = \cup\{K_B(x) \mid x \in X, K_B(x) \subseteq X \}.$$

Obviously, the subset B-lower approximation of X is the same set as the concept B-lower approximation of X. A *concept B-upper approximation* of the concept X is defined as follows:

$$\overline{B}X = \cup\{K_B(x) \mid x \in X, K_B(x) \cap X \neq \emptyset \}.$$

In our example of the decision presented in Table 2 and $R(A) = \mathrm{LV}(A)$, the concept A-upper approximations are

$$\overline{A}\{1\} = \{1\},$$
$$\overline{A}\{2\} = \{2, 4\},$$
$$\overline{A}\{3, 4, 5\} = \{1, 3, 4, 5\}.$$

In our example of the decision presented in Table 3 and $R(A) = \text{DCC}(A)$, the concept A-upper approximations are

$$\overline{A}\{1\} = \{1, 5\},$$

$$\overline{A}\{2\} = \{2, 4\},$$

$$\overline{A}\{3, 4, 5\} = \{1, 2, 3, 4, 5\} = U.$$

Note that for completely specified decision tables, all three definitions of lower approximations coalesce to the same definition. Also, for completely specified decision tables, all three definitions of upper approximations coalesce to the same definition. This is not true for incompletely specified decision tables, as the example shows. Since any characteristic relation $R(B)$ is reflexive, singleton lower and upper approximations are subsets of subset lower and upper approximations, respectively.

Also, note that using characteristic relation $\text{LV}(A)$, even if we are going to use all three attributes to describe case 2, we cannot describe this case not describing, at the same time, case 4. Thus, the set of rules describing only $\{2\}$ is the empty set. In some rule induction systems the expectation is that the set of all possible rules, induced from an upper approximation cannot be the empty set, so such a system may encounter an infinite loop. This situation cannot happen to the subset or concept definitions of upper approximation. Besides, the concept definition of upper approximation is a subset of the subset definition of upper approximation, so the concept definition of upper approximation is better suited for rule induction. Moreover, it better fits into the idea that the upper approximation should be the *smallest* set containing the concept.

Furthermore, some properties that hold for singleton lower and upper approximations do not hold—in general—for subset lower and upper approximations and for concept lower and upper approximations. For example, as noted in [13], for singleton lower and upper approximations

$$\{x \in U \mid I_B(x) \subseteq X\} \supseteq \{x \in U \mid J_B(x) \subseteq X\}$$

and

$$\{x \in U \mid I_B(x) \cap X \neq \emptyset\} \subseteq \{x \in U \mid J_B(x) \cap X \neq \emptyset\},$$

where $I_B(x)$ is a characteristic set of $\text{LV}(B)$ and $J_B(X)$ is a characteristic set of $\text{DCC}(B)$.

In our example, for the subset definition of A-lower approximation, $X = \{3, 4, 5\}$, and the characteristic relation $\text{LV}(A)$ (see Table 2)

$$\cup\{I_B(x) \mid I_B(x) \subseteq X\} = \{3, 4\}$$

while for the subset definition of A-lower approximation, $X = \{3, 4, 5\}$, and the characteristic relation $\text{DCC}(A)$ (see Table 3)

$$\cup\{J_B(x) \mid J_B(x) \subseteq X\} = \{3, 5\},$$

so neither the former set is a subset of the latter nor vice versa.

6. Rule induction

Since all characteristic sets $K(x)$, where $x \in U$, are intersections of blocks of attribute-value pairs, for attributes from B, and for subset and concept definitions of lower and upper approximations, lower and upper approximations are unions of sets of the type $K(x)$, it is natural for rule induction to use an algorithm based on blocks of attribute-value pairs, such as LEM2 [1, 5].

For example, for Table 2, i.e., for the characteristic relation $\text{LV}(A)$, the certain rules [3], induced from the concept lower A-approximations are

(Location, good) & (Basement, yes) -> (Value, high),
(Basement, no) -> (Value, medium),
(Location, bad) & (Basement, yes) -> (value, medium).

The possible rules [3], induced from the concept upper A-approximations, for the same characteristic relation $\text{LV}(A)$ are

(Location, good) & (Basement, yes) -> (Value, high),
(Location, bad) -> (Value, small),
(Location, good) -> (Value, medium),
(Basement, yes) -> (Value, medium),
(Fireplace, yes) -> (Value, medium).

7. Other approaches to missing attribute values

In this paper two basic approaches to missing attribute values, based on interpretation of a missing attribute value as lost or a "do not care" condition were discussed. Even though the suggested definitions cover the situation in which in the same decision table some missing attribute values are considered to be lost and other are "do not care" conditions, there exist many other possibilities to interpret missing attribute values.

For example, for the attribute *Basement* from our example, we may introduce a special, new value, say *maybe*, for case 2 and we may consider that the missing attribute value for case 5 should be *no*. Neither of these two cases falls into the category of lost values or "do not care" conditions. Nevertheless, such approaches may be

studied using the same idea of blocks of attribute-value pairs. More specifically, for attribute *Basement*, the blocks will be

$$[(Basement, maybe)] = \{2\},$$
$$[(Basement, yes)] = \{1, 3\}, \text{ and}$$
$$[(Basement, no)] = \{3, 5\}.$$

Then we may compute a new characteristic relation, using the technique from Section 3 and define lower and upper approximations using one of the possibilities of Section 5, preferable concept lower and upper approximations, and, eventually, induce certain and possible rules.

8. Conclusions

This paper discussed data with missing attribute values using rough set theory as the main research tool. The existing two approaches to missing attribute values, interpreted as a lost value or as a "do not care" condition are generalized by interpreting every missing attribute value separately as a lost value or as a "do not care" condition. Characteristic relations are introduced to describe incompletely specified decision tables. For completely specified decision tables any characteristic relation is reduced to an indiscernibility relation. It is shown that the basic rough set idea of lower and upper approximations for incompletely specified decision tables may be defined in a variety of different ways. Some of these definitions should have preference of use for rule induction because there is a guarantee that all necessary cases will be described by rules. Again, for completely specified decision tables, all of these definitions of lower and upper approximations are reduced to the standard definition of lower and upper approximations.

References

[1] C.-C. Chan and J. W. Grzymala-Busse. On the attribute redundancy and the learning programs ID3, PRISM, and LEM2. Department of Computer Science, University of Kansas, TR-91-14, December 1991, 20 pp.

[2] G. Birkhoff. *Lattice Theory*. American Mathematical Society, Providence, RI, 1940.

[3] J. W. Grzymala-Busse. Knowledge acquisition under uncertainty—A rough set approach. *Journal of Intelligent & Robotic Systems* 1, 1 (1988), 3–16.

[4] J. W. Grzymala-Busse. On the unknown attribute values in learning from examples. Proc. of the ISMIS-91, 6th International Symposium on Methodologies for Intelligent Systems, Charlotte, North Carolina, October 16–19, 1991, 368–377, *Lecture Notes in Artificial Intelligence*, vol. 542, Springer-Verlag, Berlin, Heidelberg, New York, 1991.

[5] J. W. Grzymala-Busse. LERS—A system for learning from examples based on rough sets. In *Intelligent Decision Support. Handbook of Applications and Advances of the Rough Sets Theory*, ed. by R. Slowinski, Kluwer Academic Publishers, Dordrecht, Boston, London, 1992, 3–18.

[6] J. W. Grzymala-Busse and M. Hu. A comparison of several approaches to missing attribute values in data mining. Proceedings of the Second International Conference on Rough Sets and Current Trends in Computing RSCTC'2000, October 16–19, 2000, Banff, Canada, 340–347.

[7] J. W. Grzymala-Busse and A. Y. Wang. Modified algorithms LEM1 and LEM2 for rule induction from data with missing attribute values. Proc. of the Fifth International Workshop on Rough Sets and Soft Computing (RSSC'97) at the Third Joint Conference on Information Sciences (JCIS'97), Research Triangle Park, NC, March 2–5, 1997, 69–72.

[8] M. Kryszkiewicz. Rough set approach to incomplete information systems. Proceedings of the Second Annual Joint Conference on Information Sciences, September 28–October 1, 1995, Wrightsville Beach, NC, 194–197.

[9] M. Kryszkiewicz. Rules in incomplete information systems. *Information Sciences* 113 (1999) 271–292.

[10] Z. Pawlak. Rough Sets. *International Journal of Computer and Information Sciences*, 11 (1982) 341–356.

[11] Z. Pawlak. *Rough Sets. Theoretical Aspects of Reasoning about Data*. Kluwer Academic Publishers, Dordrecht, Boston, London, 1991.

[12] J. Stefanowski. *Algorithms of Decision Rule Induction in Data Mining*. Poznan University of Technology Press, Poznan, Poland, 2001.

[13] J. Stefanowski and A. Tsoukias. On the extension of rough sets under incomplete information. Proceedings of the 7th International Workshop on New Directions in Rough Sets, Data Mining, and Granular-Soft Computing, RSFDGrC'1999, Yamaguchi, Japan, 73–81.

[14] J. Stefanowski and A. Tsoukias. Incomplete information tables and rough classification. *Computational Intelligence* 17 (2001) 545–566.