Characteristic Relations for Incomplete Data: A Generalization of the Indiscernibility Relation

Jerzy W. Grzymala-Busse

Department of Electrical Engineering and Computer Science,
University of Kansas, Lawrence, KS 66045, USA
Institute of Computer Science,
Polish Academy of Sciences, 01-237 Warsaw, Poland
Jerzy@ku.edu
http://lightning.eecs.ku.edu/index.html

Abstract. This paper shows that attribute-value pair blocks, used for many years in rule induction, may be used as well for computing indiscernibility relations for completely specified decision tables. Much more importantly, for incompletely specified decision tables, i.e., for data with missing attribute values, the same idea of attribute-value pair blocks is a convenient tool to compute characteristic sets, a generalization of equivalence classes of the indiscernibility relation, and also characteristic relations, a generalization of the indiscernibility relation. For incompletely specified decision tables there are three different ways lower and upper approximations may be defined: singleton, subset and concept. Finally, it is shown that, for a given incomplete data set, the set of all characteristic relations for the set of all congruent decision tables is a lattice.

1 Introduction

An idea of an attribute-value pair block, used for many years in rule induction algorithms such as LEM2 [4], may be applied not only for computing indiscernibility relations for completely specified decision tables but also for computing characteristic relations for incompletely specified decision tables. A characteristic relation is a generalization of the indiscernibility relation.

Using attribute-value pair blocks for completely specified decision tables, equivalence classes of the indiscernibility relation are computed first, then the indiscernibility relation is defined from such equivalence classes. Similarly, for incompletely specified decision tables, attribute-value pair blocks, defined in a slightly modified way, are used to compute characteristic sets, then characteristic relations are computed from these sets.

Decision tables are incomplete mainly for two reasons. First, an attribute value is lost, i. e., it was recorded but currently is unavailable. Second, the original value was irrelevant and as such not recorded and the case was classified on the basis of remaining attribute values. Such missing attribute values will be called "do not care" conditions.

J.F. Peters and A. Skowron (Eds.): Transactions on Rough Sets IV, LNCS 3700, pp. 58–68, 2005.

Initially, decision tables with all missing attribute values that are lost were studied, within rough set theory, in [8], where two algorithms for rule induction from such data were presented. This approach was studied later, see, e.g., [15] and [16] where the indiscernibility relation was generalized to describe such incompletely specified decision tables.

The first attempt to study "do not care" conditions using rough set theory was presented in [3], where a method for rule induction was introduced in which missing attribute values were replaced by all values from the domain of the attribute. "Do not care" conditions were also studied later, see, e.g., [9] and [10], where the indiscernibility relation was again generalized, this time to describe incomplete decision tables with "do not care" conditions.

In this paper we will assume that the same incomplete decision table may have missing attribute values of both types—lost attribute values and "do not care" conditions.

For a given completely specified decision table and concept, the lower and upper approximations of the concept are unique, though they may be defined in a few different ways [11] and [12]. For an incomplete decision table, lower and upper approximations of the concept may be defined in a few different ways, but—in general—the approximations of different types differ. In this paper we will discuss three different lower and upper approximations, called singleton, subset, and concept approximations [5]. Singleton lower and upper approximations were studied in [9], [10], [15] and [16]. As it was observed in [4], concept lower and upper approximations should be used for data mining. Note that similar three definitions of lower and upper approximations, though not for incomplete decision tables, were studied in [2], [13], [17], [18] and [19].

The last topic of the paper is studying the class of congruent incomplete decision tables, i.e., tables with the same set of all cases, the same attribute set, the same decision, and the same corresponding specified attribute values. Two congruent decision tables may differ only by missing attribute values (some of them are lost attribute values the others are "do not care" conditions). A new idea of a signature, a vector of all missing attribute values, is introduced. There is a one-to-one correspondence between signatures and congruent decision tables. The paper includes also the Homomorphism Theorem showing that the defined operation on characteristic relations is again a characteristic relation for some congruent decision table. For a given incomplete decision table, the set of all characteristic relations for the set of all congruent decision tables is a lattice.

A preliminary version of this paper was presented at the Fourth International Conference on Rough Sets and Current Trends in Computing, Uppsala, Sweden, June 15, 2004 [6].

2 Blocks of Attribute-Value Pairs, Characteristic Sets, and Characteristic Relations

An example of a decision table is presented in Table 1. Rows of the decision table represent *cases*, while columns represent *variables*. The set of all cases is denoted by U. In Table 1, $U = \{1, 2, ..., 7\}$. 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. In Table 1, $A = \{Age,\ Hypertension,\ Complications\}$. Any decision table defines a function ρ that maps the direct product of U and A into the set of all values. For example, in Table 1, $\rho(1, Age) = 20..29$. Function ρ describing Table 1 is completely specified (total). A decision table with completely specified function ρ will be called *completely specified*, or, simpler, *complete*.

Table 1. A complete decision table

	Attributes			Decision
Case	Age	Hypertension	Complications	Delivery
1	20..29	no	none	fullterm
2	20..29	yes	obesity	preterm
3	20..29	yes	none	preterm
4	20..29	no	none	fullterm
5	30..39	yes	none	fullterm
6	30..39	yes	alcoholism	preterm
7	40..50	no	none	fullterm

Rough set theory, see, e.g., [11] and [12], is based on the idea of an indiscernibility relation, defined for complete decision tables. 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

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

The indiscernibility relation $IND(B)$ is an equivalence relation. Equivalence classes of $IND(B)$ are called *elementary sets* of B and are denoted by $[x]_B$. For example, for Table 1, elementary sets of $IND(A)$ are $\{1, 4\}$, $\{2\}$, $\{3\}$, $\{5\}$, $\{6\}$, $\{7\}$. The indiscernibility relation $IND(B)$ may be computed using the idea of blocks of attribute-value pairs. Let a be an attribute, i.e., $a \in A$ and let v be a value of a for some case. For complete 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. For Table 1,

[(Age, 20..29)] = {1, 2, 3, 4},
[(Age, 30..39)] = {5, 6},
[(Age, 40..50)] = {7},
[(Hypertension, no)] = {2, 3, 5, 6},
[(Hypertension, yes)] = {1, 4, 7},
[(Complications, none)] = {1, 3, 4, 5, 7},
[(Complications, obesity)] = {2}, and
[(Complications, alcoholism)] = {6}.

The indiscernibility relation $IND(B)$ is known when all elementary sets of $IND(B)$ are known. Such elementary sets of B are intersections of the corresponding attribute-value pairs, i.e., for any case $x \in U$,

$$[x]_B = \cap\{[(a, v)] | a \in B, \rho(x, a) = v\}.$$

We will illustrate the idea how to compute elementary sets of B for Table 1 where $B = A$:

$[1]_A = [4]_A = [(Age, 20..29)] \cap [(Hypertension, no)] \cap [(Complications, none)] = \{1, 4\}$,
$[2]_A = [(Age, 20..29)] \cap [(Hypertension, yes)] \cap [(Complications, obesity)] = \{2\}$,
$[3]_A = [(Age, 20..29)] \cap [(Hypertension, yes)] \cap [(Complications, none)] = \{3\}$,
$[5]_A = [(Age, 30..39)] \cap [(Hypertension, yes)] \cap [(Complications, none)] = \{5\}$,
$[6]_A = [(Age, 30..39)] \cap [(Hypertension, yes)] \cap [(Complications, alcohol)] = \{6\}$,
and
$[7]_A = [(Age, 40..50)] \cap [(Hypertension, no)] \cap [(Complications, none)] = \{7\}$.

A decision table with an incompletely specified (partial) function ρ will be called *incompletely specified*, or *incomplete*. For the rest of the paper we will assume that all decision values are specified, i.e., they 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. Incomplete decision tables are described by characteristic relations instead of indiscernibility relations. Also, elementary sets are replaced by characteristic sets. An example of an incomplete table is presented in Table 2.

Table 2. An incomplete decision table

	Attributes			Decision
Case	Age	Hypertension	Complications	Delivery
1	?	*	none	fullterm
2	20..29	yes	obesity	preterm
3	20..29	yes	none	preterm
4	20..29	no	none	fullterm
5	30..39	yes	?	fullterm
6	*	yes	alcoholism	preterm
7	40..50	no	?	fullterm

For incomplete decision tables the definition of a block of an attribute-value pair must be modified. If for an attribute a there exists a case x such that

$\rho(x, a) = ?$, i.e., the corresponding value is lost, then the case x should not be included in any block $[(a, v)]$ for all values 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 all blocks $[(a, v)]$ for every possible value v of attribute a. This modification of the definition of the block of attribute-value pair is consistent with the interpretation of missing attribute values, lost and "do not care" condition. Thus, for Table 2

[(Age, 20..29)] = {2, 3, 4, 6},
[(Age, 30..39)] = {5, 6},
[(Age, 40..50)] = {6, 7},
[(Hypertension, no)] = {1, 4, 7},
[(Hypertension, yes)] = {1, 2, 3, 5, 6},
[(Complications, none)] = {1, 3, 4},
[(Complications, obesity)] = {2},
[(Complications, alcoholism)] = {6}.

We define a *characteristic set* $KB(x)$ as 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 Table 2 and $B = A$,

$K_A(1) = \{1, 3, 4\}$,
$K_A(2) = \{2, 3, 4, 6\} \cap \{1, 2, 3, 5, 6\} \cap \{2\} = \{2\}$,
$K_A(3) = \{2, 3, 4, 6\} \cap \{1, 2, 3, 5, 6\} \cap \{1, 3, 4\} = \{3\}$,
$K_A(4) = \{2, 3, 4, 6\} \cap \{1, 4, 7\} \cap \{1, 3, 4\} = \{4\}$,
$K_A(5) = \{5, 6\} \cap \{1, 2, 3, 5, 6\} = \{5, 6\}$,
$K_A(6) = \{1, 2, 3, 5, 6\} \cap \{6\} = \{6\}$, and
$K_A(7) = \{6, 7\} \cap \{1, 4, 7\} = \{7\}$.

The characteristic set $K_B(x)$ may be interpreted as the smallest set of cases that are indistinguishable from x using all attributes from B and using a given interpretation of missing attribute values. Thus, $K_A(x)$ is the set of all cases that cannot be distinguished from x using all attributes. The *characteristic relation* $R(B)$ is a relation on U defined for $x, y \in U$ as follows:

$$(x, y) \in R(B) \ if \ and \ only \ if \ y \in K_B(x).$$

We say that $R(B)$ is *implied* by its characteristic sets $K_B(x)$, $x \in U$. The characteristic relation $R(B)$ is reflexive but—in general—does not need to be symmetric or transitive. Also, the characteristic relation $R(B)$ is known if we know characteristic sets $K(x)$ for all $x \in U$. In our example, $R(A) = \{(1, 1),$ $(1, 3), (1, 4), (2, 2), (3, 3), (4, 4), (5, 5), (5, 6), (6, 6), (7, 7)\}$. The most convenient way to define the characteristic relation is through the characteristic sets. Nevertheless, the characteristic relation $R(B)$ may be defined independently of characteristic sets in the following way:

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

3 Lower and Upper Approximations

For completely specified decision tables lower and upper approximations are defined on the basis of the indiscernibility relation. Any finite union of elementary sets, associated with B, will be called a *B-definable set.* Let X be any subset of the set U of all cases. The set X is called a *concept* and is usually defined as the set of all cases defined by a 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

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

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

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

The above shown way of computing lower and upper approximations, by constructing these approximations from singletons x, will be called the *first method.* The B-lower approximation of X is the greatest B-definable set, contained in X. The B-upper approximation of X is the smallest B-definable set containing X.

As it was observed in [12], for complete decision tables we may use a *second method* to define the B-lower approximation of X, by the following formula

$$\cup\{[x]_B | x \in U, [x]_B \subseteq X\},$$

and the B-upper approximation of x may de defined, using the second method, by

$$\cup\{[x]_B | x \in U, [x]_B \cap X \neq \emptyset).$$

For incompletely specified decision tables lower and upper approximations may be defined in a few different ways. First, the definition of definability should be modified. Any finite union of characteristic sets of B is called a *B-definable* set. In this paper we suggest three different definitions of lower and upper approximations. 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 incomplete 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 [9], [10], [14], [15] and [16], i.e., lower and upper approximations are sets of singletons from the universe U satisfying some properties. Thus, lower and upper approximations are defined by analogy with the above first method, by constructing both sets from singletons. We will call these approximations *singleton.* A singleton B-lower approximation of X is defined as follows:

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

A singleton B-upper approximation of X is

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

In our example of the decision table presented in Table 2 let us say that $B = A$. Then the singleton A-lower and A-upper approximations of the two concepts: {1, 4, 5, 7} and {2, 3, 6} are:

$$\underline{A}\{1, 4, 5, 7\} = \{4, 7\},$$

$$\underline{A}\{2, 3, 6\} = \{2, 3, 6\},$$

$$\overline{A}\{1, 4, 5, 7\} = \{1, 4, 5, 7\},$$

$$\overline{A}\{2, 3, 6\} = \{1, 2, 3, 5, 6\}.$$

Note that the set $\overline{A}\{1, 4, 5, 7\} = \{1, 4, 5, 7\}$ is not A-definable (this set cannot be presented as a union of intersections of attribute-value pair blocks). The problem is caused by case 5. This case appears twice in the list of all blocks of attribute-value pairs, namely, in [(Age, 30..39)] and [(Hypertension, yes)]. However, both of these blocks contain also case 6. Hence any intersection of blocks of attribute value pairs, containing case 5, must also contain case 6. Thus, using intersection and union of blocks of attribute-value pairs we may construct the set {1, 4, 5, 6, 7} but not the set {1, 4, 5, 7}. Therefore, singleton approximations are, in general, not A-definable, and, as such, are not useful for rule induction.

The second method of defining lower and upper approximations for complete decision tables uses another idea: lower and upper approximations are unions of elementary sets, subsets of U. Therefore we may define lower and upper approximations for incomplete decision tables by analogy with the second method, using characteristic sets instead of elementary sets. There are two ways to do this. Using the first way, a *subset* B-lower approximation of X is defined as follows:

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

A *subset* B-upper approximation of X is

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

Since any characteristic relation $R(B)$ is reflexive, for any concept X, singleton B-lower and B-upper approximations of X are subsets of the subset B-lower and B-upper approximations of X, respectively. For the same decision table, presented in Table 2, the subset A-lower and A-upper approximations are

$$\underline{A}\{1, 4, 5, 7\} = \{4, 7\},$$

$$\underline{A}\{2, 3, 6\} = \{2, 3, 6\},$$

$$\overline{A}\{1, 4, 5, 7\} = \{1, 3, 4, 5, 6, 7\},$$

$$\overline{A}\{2, 3, 6\} = \{1, 2, 3, 4, 5, 6\}.$$

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

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

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:

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

The concept B-upper approximation of X is a subset of the subset B-upper approximation of X. Thus, concept upper approximations are more useful for rule induction than subset upper approximations. For the decision presented in Table 2, the concept A-lower and A-upper approximations are

$$\underline{A}\{1, 4, 5, 7\} = \{4, 7\},$$

$$\underline{A}\{2, 3, 6\} = \{2, 3, 6\},$$

$$\overline{A}\{1, 4, 5, 7\} = \{1, 3, 4, 5, 6, 7\},$$

$$\overline{A}\{2, 3, 6\} = \{2, 3, 6\}.$$

Note that for complete decision tables, all three definitions of lower approximations, singleton, subset and concept, coalesce to the same definition. Also, for complete decision tables, all three definitions of upper approximations coalesce to the same definition. This is not true for incomplete decision tables, as our example shows.

4 Congruent Decision Tables

In this section, for simplicity, all characteristic relations will be defined for the entire set A of attributes instead of its subset B. In addition, and the characteristic relation will be denoted by R instead of $R(A)$. Finally, 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 ?. Obviously, there is 2^n congruent decision tables, where n is the total number of all missing attribute values in a decision table.

To every incomplete decision table we will assign a *signature* of missing attribute values, a vector $(p_1, p_2, ..., p_n)$, where p_i is equal to either ? or *, the value taken from the incomplete decision table; $i = 1, 2, ..., n$, by scanning the decision table, row after row, starting from the top row, from left to right. Thus every consecutive missing attribute value should be placed as a component of the signature, where p_1 is the first missing attribute value, identified during scanning, and p_n is the last one. For Table 2, the signature is (?, *, ?, *, ?). In the set of all congruent decision tables, a signature uniquely identifies the table and vice versa. On the other hand, congruent decision tables with different signatures may have the same characteristic relations. For example, tables congruent with Table 2, with signatures (?, *, *, *, *) and (*, ?, *, *, *), have the same characteristic relations. Two congruent decision tables that have the same characteristic relations will be called *indistinguishable*.

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 $x \in U$. For two congruent decision tables D_1 and D_2 we define a characteristic relation $R = R_1 \cdot R_2$ as implied by characteristic sets $K_1(x) \cap K_2(x)$. For two signatures p and q, $p \cdot q$ is defined as a signature r with $r_i(x) = *$ if and only if $p_i(x) = *$ and $q_i(x) = *$, otherwise $r_i(x) = ?$, $i = 1, 2, ..., n$.

Let $A = \{a_1, a_2, ..., a_k\}$. Additionally, let us define, for $x \in U$ and $a \in A$, the set $[(a, \rho(x, a))]^+$ in the following way: $[(a, \rho(x, a))]^+ = [(a, \rho(x, a))]$ if $\rho(x, a) \neq *$ and $\rho(x, a) \neq ?$ and $[(a, \rho(x, a))]^+ = U$ otherwise.

Lemma. For $x \in U$, the characteristic set $K(x) = \cap_{i=1}^{k}[(a_i, \rho(x, a_i))]^+$.

Proof. In the definition of $K(x)$, if $\rho(x, a) = *$ or $\rho(x, a) = ?$, the corresponding block $[(a, \rho(x, a))]$ is ignored. Additionally, by our assumption, for every $x \in U$ there exists an attribute $a \in A$ such that $\rho(x, a) \neq *$ and $\rho(x, a) \neq ?$.

Let D be an incomplete decision table and let p be the signature of D. Let ψ be a function that maps a signature p into a characteristic relation R of D.

Homomorphism Theorem. Let p and q be two signatures of congruent decision tables. Then $\psi(p \cdot q) = \psi(p) \cdot \psi(q)$, i.e., ψ is a homomorphism.

Proof. Let D_1, D_2 be two congruent decision tables with functions ρ_1 and ρ_2, signatures p and q, and characteristic relations R_1, R_2, respectively, where $\psi(p) = R_1$ and $\psi(q) = R_2$. Let D be a congruent decision table with function ρ and signature $p \cdot q$ and let $\psi(p \cdot q) = R$. Due to Lemma, for every $x \in U$

$$K_1(x) \cdot K_2(x) = (\cap_{i=1}^{k}[(a_i, \rho_1(x, a_i))]^+) \cap (\cap_{i=1}^{k}[(a_i, \rho_2(x, a_i))]^+) =$$
$$\cap_{i=1}^{k}[(a_i, \rho_1(x, a_i))]^+ \cap [(a_i, \rho_2(x, a_i))]^+$$

If $\rho_j(x, a_i) \neq *$ and $\rho_j(x, a_i) \neq ?$ then $[(a_i, \rho_j(x, a_i))]$ contains $y \in U$ if and only if $\rho_j(y, a_i) = *$, $j = 1, 2$. Moreover, $[(a_i, \rho(x, a_i))]^+$ contains y if and only if $\rho_1(y, a_i) = *$ and $\rho_2(y, a_i) = *$. Thus, $K_1(x) \cdot K_2(x) = \cap_{i=1}^{k}[(a_i, \rho_1(x, a_i))]^+ = K(x)$.

Thus, $\psi(p) \cdot \psi(q)$ is the characteristic relation of a congruent decision table with the signature $p \cdot q$. For the set L of all characteristic relations for the set of all congruent decision tables, the operation $\cdot$ on relations is idempotent, commutative, and associative, therefore, L is a semilattice [1], p. 9. Moreover, L has a universal upper bound $\psi(*, *, ..., *)$ and its length is finite, so L is a lattice, see [1], p. 23. The second lattice operation, resembling addition, is defined directly from the diagram of a semilattice.

Let us define subset E of the set of all congruent decision tables as the set of tables with exactly one missing attribute value "?" and all remaining attribute values equal to "*". Let G be the set of all characteristic relations associated with the set E. The lattice L can be generated by G, i.e., every element of L can be expressed as $\psi(*, *, ...*)$ or as a product of some elements from G.

5 Conclusions

An attribute-value pair block is a very useful tool not only for dealing with completely specified decision tables but, much more importantly, also for incompletely specified decision tables. For completely specified decision tables attribute-value pair blocks provide for easy computation of equivalence classes of the indiscernibility relation. Similarly, for incompletely specified decision tables, attribute-value pair blocks make possible, by equally simple computations, determining characteristic sets and then, if necessary, characteristic relations.

For a given concept of the incompletely specified decision table, lower and upper approximations can be easily computed from characteristic sets—knowledge of characteristic relations is not required. Note that for incomplete decision tables there are three different approximations possible: singleton, subset and concept. The concept approximations are the best fit for the intuitive expectations for lower and upper approximations. Our last observation is that for a given incomplete decision table, the set of all characteristic relations for the set of all congruent decision tables is a lattice.

References

1. Birkhoff, G.: Lattice Theory. American Mathematical Society, Providence, RI (1940).
2. Greco, S., Matarazzo, S.B. and Slowinski, R.: Dealing with missing data in rough set analysis of multi-attribute and multi-criteria decision problems. In Decision Making: Recent developments and Worldwide Applications, ed. by S. H. Zanakis, G. Doukidis and Z. Zopounidis, Kluwer Academic Publishers, Dordrecht, Boston, London (2000) 295–316.
3. Grzymala-Busse, J.W.: 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. Lecture Notes in Artificial Intelligence, vol. 542, Springer-Verlag, Berlin, Heidelberg, New York (1991) 368–377.
4. Grzymala-Busse, J.W.: 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.
5. Grzymala-Busse, J.W.: Rough set strategies to data with missing attribute values. Workshop Notes, Foundations and New Directions of Data Mining, the 3-rd International Conference on Data Mining, Melbourne, FL, USA, November 19–22, 2003, 56–63.

6. Grzymala-Busse, J.W.: Characteristic relations for incomplete data: A generalization of the indiscernibility relation. Proceedings of the RSCTC'2004, the Fourth International Conference on Rough Sets and Current Trends in Computing, Uppsala, Sweden, June 1–5, 2004. Lecture Notes in Artificial Intelligence 3066, Springer-Verlag 2004, 244–253.
7. Grzymala-Busse, J.W. and Hu, M.: 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, Banff, Canada, October 16–19, 2000, 340–347.
8. Grzymala-Busse, J.W. and Wang, A.Y.: 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.
9. Kryszkiewicz, M.: Rough set approach to incomplete information systems. Proceedings of the Second Annual Joint Conference on Information Sciences, Wrightsville Beach, NC, September 28–October 1, 1995, 194–197.
10. Kryszkiewicz, M.: Rules in incomplete information systems. *Information Sciences* **113** (1999) 271–292.
11. Pawlak, Z.: Rough Sets. *International Journal of Computer and Information Sciences* **11** (1982) 341–356.
12. Pawlak, Z.: Rough Sets. Theoretical Aspects of Reasoning about Data. Kluwer Academic Publishers, Dordrecht, Boston, London (1991).
13. Slowinski, R. and Vanderpooten, D.: A generalized definition of rough approximations based on similarity. *IEEE Transactions on Knowledge and Data Engineering* **12** (2000) 331–336.
14. Stefanowski, J.: Algorithms of Decision Rule Induction in Data Mining. Poznan University of Technology Press, Poznan, Poland (2001).
15. Stefanowski, J. and Tsoukias, A.: 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, Ube, Yamaguchi, Japan, November 8–10, 1999, 73–81.
16. Stefanowski, J. and Tsoukias, A.: Incomplete information tables and rough classification. *Computational Intelligence* **17** (2001) 545–566.
17. Yao, Y.Y.: Two views of the theory of rough sets in finite universes. *International J. of Approximate Reasoning* **15** (1996) 291–317.
18. Yao, Y.Y.: Relational interpretations of neighborhood operators and rough set approximation operators. *Information Sciences* **111** (1998) 239–259.
19. Yao, Y.Y.: On the generalizing rough set theory. Proc. of the 9th Int. Conference on Rough Sets, Fuzzy Sets, Data Mining and Granular Computing (RSFDGrC'2003), Chongqing, China, October 19–22, 2003, 44–51.