B6. A Rough Set Perspective on Data and Knowledge

Andrzej Skowron[1], *Jan Komorowski*[4,5], *Zdzisław Pawlak*[2], *Lech Polkowski*[3,5]

[1] Institute of Mathematics, Warsaw University
02-097 Warszawa, Poland
skowron@mimuw.edu.pl

[2] Institute of Theoretical and Applied Informatics, Polish Academy of Sciences
ul. Baltycka 5, 44-000 Gliwice, Poland
zpw@ii.pw.edu.pl

[3] Institute of Mathematics, Warsaw University of Technology
Pl. Politechniki 1, 00-665 Warszawa, Poland
polk@mimuw.edu.pl

[4] Department of Computer and Information Science
Norwegian University of Science and Technology (NTNU)
7034 Trondheim, Norway
janko@idi.ntnu.no

[5] Polish-Japanese Institute of Computer Techniques
Koszykowa 86, 02-008 Warszawa, Poland

Abstract

Rough set theory was proposed by Zdzisław Pawlak [43, 44] in the early 1980's. Since then we have witnessed a systematic, world-wide growth of interest in rough set theory and its applications. Rough set approach has been introduced to deal with vague or imprecise concepts, to derive knowledge from data and to reason about knowledge derived from data.

In the first part of this article we outline the basic notions of rough sets, especially those which are related to knowledge extraction from data.

Searching for knowledge is usually guided by some constraints [25]. A wide class of such constraints can be expressed by discernibility of objects. Knowledge derived from data by the rough set approach consists of different constructs. Among them there are reducts, which are the central construct in the rough set approach, different kinds of rules (such as, for example, decision rules or association rules), dependencies, patterns (templates) or classifiers. The reducts are of special importance since all other constructs can be derived from different kinds of reducts using the rough set approach.

Strategies for searching reducts apply Boolean (propositional) reasoning [4] since the constraints (e.g. constraints related to the discernibility of objects) are expressible by propositional formulae. Moreover, using Boolean reasoning, minimal description length data models [56], [30] can be induced since they correspond to constructs of Boolean functions called prime implicants (or their approximations).

The second part of this article includes illustrative examples of applications of this general scheme to inducing from data various forms of knowledge.

Keywords: indiscernibility, Boolean reasoning, lower and upper approximations, rough sets, boundary region, positive region, rough membership function, decision rules, patterns, rough mereology.

B6.1 Basic rough set approach

We start by presenting the basic notions of classical rough set approach [44] that was introduced to deal with imprecise or vague concepts.

Information systems

A data set can be represented by a table where each row represents, for instance, an object, a case, or an event. Every column represents an attribute, or an observation, or a property that can be measured for each object; it can also be supplied by a human expert or user. This table is called an *information system*. More formally, it is a pair $\mathcal{A} = (U, A)$ where U is a non-empty finite set of *objects* called the *universe* and A is a non-empty finite set of *attributes* such that $a : U \to V_a$ for every $a \in A$. Set V_a is called the *value set* of a. By $Inf_B(x) = \{(a, a(x)) : a \in B\}$ we denote the *information signature of x with respect to B*, where $B \subseteq A$ and $x \in U$.

Decision systems

In many cases the target of the classification, that is, the family of concepts to be approximated is represented by an additional attribute called decision. Information systems of this kind are called *decision systems*. A decision system is any system of the form $\mathcal{A} = (U, A, d)$, where $d \notin A$ is the *decision attribute* and A is a set of *conditional attributes* or simply *conditions*.

Let $\mathcal{A} = (U, A, d)$ be given and let $V_d = \{v_1, \ldots, v_{r(d)}\}$. Decision d determines a partition $\{X_1, \ldots, X_{r(d)}\}$ of the universe U, where $X_k = \{x \in U \ : \ d(x) = v_k\}$ for $1 \leq k \leq r(d)$. The set X_i is called the *i-th decision class of $\mathcal{A}$*. By $X_d(u)$ we denote the decision class $\{x \in U \ : \ d(x) = d(u)\}$, for any $u \in U$.

One can generalize the above definition to a case of decision systems of the form $\mathcal{A} = (U, A, D)$ where the set of decision attributes $D = \{d_1, ...d_k\}$ and A are assumed to be disjoint. Formally, this system can be treated as a decision system $\mathcal{A} = (U, A, d_D)$ where $d_D(x) = (d_1(x), ..., d_k(x))$ for $x \in U$.

Decision tables may be identified with training samples known in Machine Learning and used to induce concept approximations in the process known as supervised learning [30].

Rough set approach allows to precisely define the notion of concept approximation. It is based [44] on the indiscernibility relation between objects defining a partition (or covering) of universe U of objects. Since objects are perceived by means of the values of the available attributes, the objects having the same (or similar) values of attributes are indiscernible.

Indiscernibility relation

Let $\mathcal{A} = (U, A)$ be an information system, then with any $B \subseteq A$ there is associated an equivalence relation $IND_{\mathcal{A}}(B)$:

$$IND_{\mathcal{A}}(B) = \{(x, x') \in U^2 \ : \ \forall a \in B \ a(x) = a(x')\}.$$

$IND_{\mathcal{A}}(B)$ (or, $IND(B)$, for short) is called the *B-indiscernibility relation* and its equivalence classes are denoted by $[x]_B$. X/B denotes the partition of U which is defined by the indiscernibility relation $IND(B)$.

We will now discuss what sets of objects may be expressed (defined) by formulae constructed by means of attributes and their values. The simplest formulae, called *descriptors*, are of the form $a = v$ where $a \in A$ and $v \in V_a$. (It is also possible to consider *generalized descriptors* of the form $a \in S$ where $S \subseteq V_a$.) The descriptors can be combined into more complex formulae using propositional connectives. The meaning $|\varphi|_{\mathcal{A}}$ of formula φ in $\mathcal{A}$ is defined inductively as follows:

1. if φ is of the form $a = v$ then $|\varphi|_{\mathcal{A}} = \{x \in U : a(x) = v\}$;
2. $|\varphi \wedge \varphi'|_{\mathcal{A}} = |\varphi|_{\mathcal{A}} \cap |\varphi'|_{\mathcal{A}}$; $|\varphi \vee \varphi'|_{\mathcal{A}} = |\varphi|_{\mathcal{A}} \cup |\varphi'|_{\mathcal{A}}$; $|\neg\varphi|_{\mathcal{A}} = U - |\varphi|_{\mathcal{A}}$.

The above definition may be easily extended to generalized descriptors.

Any set of objects $X \subseteq U$ definable in $\mathcal{A}$ by some formula φ (i.e., $X = |\varphi|_{\mathcal{A}}$) is referred to as a *crisp* (exact) set – otherwise the set is *rough* (*inexact, vague*). Vague concepts may be only approximated by crisp concepts; these approximations are defined now [44].

Lower and upper approximation of sets, boundary regions

Let $\mathcal{A} = (U, A)$ be an information system and let $B \subseteq A$ and $X \subseteq U$. We can approximate X using only the information contained in B by constructing the so-called *B-lower* and *B-upper approximations of* X, denoted $\underline{B}X$ and $\overline{B}X$ respectively, where $\underline{B}X = \{x : [x]_B \subseteq X\}$ and $\overline{B}X = \{x : [x]_B \cap X \neq \emptyset\}$.

The lower approximation corresponds to certain rules while the upper approximation to possible rules (rules with confidence greater than 0) (see Section **??**). The B-lower approximation of X is the set of all objects which can be certainly classified to X using attributes from B. The set $U - \overline{B}X$ is called the *B-outside region of* X and consists of those objects which can be certainly classified as not belonging to X using attributes from B. The set $BN_B(X) = \overline{B}X - \underline{B}X$ is called the *B-boundary region of* X thus consisting of those objects that on the basis of the attributes from B cannot be unambiguously classified into X. A set is said to be *rough* (respectively *crisp*) if the boundary region is non-empty (respectively empty). Consequently, each rough set has boundary-line cases, i.e., objects which can neither be certainly classified as members of the set nor of its complement. Obviously, crisp sets have no boundary-line elements at all. It follows that boundary-line cases cannot be properly classified by employing the available knowledge. The size of the boundary region can be used as a measure of the quality of set approximation (in U).

It can be easily seen that the lower and upper approximations of a set are, respectively, the interior and the closure of this set in the topology generated by the indiscernibility relation.

It is possible to consider weaker indiscernibility relations defined by so called tolerance relations defining coverings of the universe of objects by tolerance (similarity) classes. An extension of rough set approach based on tolerance relations

has been used for pattern extraction and concept approximation (see, e.g., [63], [67], [37], [34]).

Quality measures of concept approximation and measures of inclusion and closeness of concepts

We now present some examples of measures of quality approximation as well as of inclusion and closeness (approximate equivalence). These notions are instrumental in evaluating the strength of rules and closeness of concepts. They are also applicable in determining plausible reasoning schemes [49], [54]. An important role is also played by entropy measures (see e.g., [11]).

Let us consider first an example of a quality measure of approximations.

Accuracy of approximation. A rough set X can be characterized numerically by the following coefficient

$$\alpha_B(X) = \frac{|\underline{B}(X)|}{|\overline{B}(X)|},$$

called the *accuracy of approximation*, where $|X|$ denotes the cardinality of $X \neq \emptyset$ and B is a set of attributes. Obviously, $0 \leq \alpha_B(X) \leq 1$. If $\alpha_B(X) = 1$, X is *crisp* with respect to B (X is *exact* with respect to B); otherwise, i.e., if $\alpha_B(X) < 1$, X is *rough* with respect to B (X is *vague* with respect to B).

Rough membership function. In classical set theory either an element belongs to a set or it does not. The corresponding membership function is the characteristic function of the set, i.e., the function takes values 1 and 0, respectively. In the case of rough sets the notion of membership is different. The *rough membership function* quantifies the degree of relative overlap between the set X and the equivalence class to which x belongs. It is defined by

$$\mu_X^B(x) : U \to [0,1] \text{ and } \mu_X^B(x) = \frac{|[x]_B \cap X|}{|[x]_B|}.$$

The rough membership function can be interpreted as a frequency–based estimate of $\Pr(y \in X \mid u)$, the conditional probability that object y belongs to set X, given the information signature $u = Inf_B(x)$ of object x with respect to attributes B. The value $\mu_X^B(x)$ measures the degree of inclusion of $\{y \in U : Inf_B(x) = Inf_B(y)\}$ in X.

Positive region and its measure. If $X_1, \ldots, X_{r(d)}$ are the decision classes of $\mathcal{A}$, then the set $\underline{B}X_1 \cup \ldots \cup \underline{B}X_{r(d)}$ is called the *B–positive region of* $\mathcal{A}$ and is denoted by $POS_B(d)$. The number $|POS_B(d)|/|U|$ measures the degree of inclusion of the partition defined by attributes from B in the partition defined by the decision.

Dependencies to a degree. Another important issue in data analysis is discovering dependencies among attributes. Intuitively, a set of attributes D depends

totally on a set of attributes C, denoted $C \Rightarrow D$, if all values of the attributes from D are uniquely determined by the values of the attributes from C. In other words, D depends totally on C, if there exists a functional dependency between values of D and C. Dependency can be formally defined as follows.

Let D and C be subsets of A. We will say that D *depends on* C to a *degree* k $(0 \leq k \leq 1)$, denoted $C \Rightarrow_k D$, if

$$k = \gamma(C, D) = \frac{|POS_C(D)|}{|U|},$$

where $POS_C(D) = POS_C(d_D)$.
Obviously,

$$\gamma(C, D) = \sum_{X \in U/D} \frac{|\underline{C}(X)|}{|U|}.$$

If $k = 1$, then D *depends totally* on C, and if $k < 1$, then D *depends partially* (to a *degree* k) on C. $\gamma(C, D)$ describes the closeness of the partition U/D and its approximation with respect to the conditions from C.

The coefficient k expresses the ratio of all elements of the universe which can be properly classified to blocks of the partition U/D by employing attributes C. It will be called the *degree of the dependency.*

Inclusion and closeness to a degree. Instead of the classical exact set inclusion, inclusion to a degree is often used in the process of deriving knowledge from data. A well-known measure of inclusion of two non-empty sets $X, Y \subseteq U$ is described by $|X \cap Y|/|X|$ (see [2] and [49]); their closeness may be defined by

$$min\left(|X \cap Y|/|X|, |X \cap Y|/|Y|\right).$$

B6.2 Searching for knowledge

We have pointed out that rough set approach has been introduced by Z. Pawlak [44] in order to deal with vague or imprecise concepts. More generally, it is an approach for deriving knowledge from data and for reasoning about knowledge derived from data. Searching for knowledge is usually guided by some constraints [25]. A wide class of such constraints can be expressed using rough set framework or its generalizations (e.g., rough mereology [49] or granular computing [54]). Knowledge derived from data by rough set approach may consist of different constructs. Among these constructs are reducts, which are fundamental to rough set approach, different kinds of rules (e.g., decision rules or association rules), dependencies, patterns (also called templates) or classifiers. The reducts are of special importance since all other constructs may be derived from different kinds of reducts.

Searching strategies for reducts are based on Boolean (propositional) reasoning [4] since constraints (e.g. related to discernibility of objects) are suitably expressible by propositional formulae. Moreover, using Boolean reasoning, it is

possible to induce data models with a minimum description length ([56], [30]) since they correspond to the constructs of Boolean functions called prime implicants (or their approximations).

Searching for knowledge can be performed in the language close to data or in a language with more abstract concepts; this is closely related to the issues of feature selection and feature extraction in Machine Learning or Pattern Recognition [30]. Let us also mention that data models derived from data by using rough set approach are controlled using statistical test procedures (for more details see, e.g., [11], [10]). A thorugh analysis of the quality of rough set classifiers, including discrimination and callibration as well as the so-called ROC analysis [69] has been originally introduced to rough sets in [38] and is available in the ROSETTA system (see Section D2.1.3 in this Handbook).

In this section we present illustrative examples showing how the outlined general scheme is used for deriving knowledge.

Finally, we would like to mention that extensions of rough sets (e.g., rough mereology [49] or granular computing [54]), have been developed for extracting knowledge and reasoning about knowledge related to more complex data models such as, for instance, those in distributed environment or related to qualitative reasoning (e.g., spatial reasoning [57]).

Now, it will be important to make some remarks on Boolean reasoning since most the methods discussed later are based on generation of reducts using Boolean reasoning.

Boolean reasoning

The combination of rough set approach with Boolean reasoning [4] has created a powerful methodology that allows to formulate and efficiently solve searching problems for different kinds of reducts and their approximations.

The idea of Boolean reasoning is as follows. Given problem P construct a corresponding Boolean function f_P. This function has a property that solutions of problem P may be recovered from prime implicants of f_P. We recall that an implicant of a Boolean function f is any conjunction of literals (variables or their negations) such that if the values of these literals are true under an arbitrary valuation v of variables then the value of the function f under v is also true. A prime implicant is a minimal implicant.

Using rough set approach, searching strategies for data models under a given partition of objects are based on discernibility and Boolean reasoning (see e.g., [37], [34],[61], [67], [68], [51], [52]). This process covers also tuning of parameterssuch as thresholds used to extract relevant partitions (or coverings), to measure the degree of inclusion (or closeness) of sets, or to extract parameters measuring the quality of approximation.

It is necessary to deal with Boolean functions of large size in order to solve real-life problems. Consequently, a successful methodology for computing many of the constructs important for applications such as reducts and their approximations, decision rules, association rules, discretization of real value attributes,

symbolic value grouping, searching for new features defined by oblique hyperplanes or higher order surfaces, pattern extraction from data as well as conflict resolution or negotiation, has been developed. The methodology is based on discernibility of objects and Boolean reasoning.

Reducts are also basic tools in extracting from data functional dependencies or functional dependencies to a degree (for references see the papers and bibliography in [61], [40], [51], [52]).

Most of the problems related to generation of the above mentioned constructs are of high computational complexity (i.e., they are NP-complete or NP-hard). This is also showing that most of the problems related to, e.g., feature selection or pattern extraction from data, have intrinsic high computational complexity. However, using the above methodology reasoning it was possible to discover efficient heuristics returning suboptimal solutions of the problems.

The reported results of experiments on many data sets are very promising. In comparison with other methods reported in literature, they show very good quality of solutions (expressed by the classification quality of unseen objects and time necessary for the construction of solutions) as generated by the heuristics. Moreover, a method to deal with large relational databases (see e.g., [33]) and a decomposition method based on patterns called templates have been developed for processing large data sets (see e.g., [37], [34]). The former method, (see e.g., [33]) has shown that Boolean reasoning methodology can be extended to large relational data bases. Its main idea is based on the observation that Boolean variables relevant to a very large formula (corresponding to an analyzed relational data base) may be discovered by analyzing some statistical information. This statistical information can be efficiently extracted from large data bases. The latter method is based on a decomposition of large data into regular subdomains which are of a size feasible for processing with previously developed methods. We will discuss this approach later.

Another interesting statistical approach is based on different sampling strategies. Samples are analyzed using the developed strategies and stable constructs for sufficiently large number of samples are considered as relevant for the whole table. This approach has been successfully used for generating different kinds of the so-called dynamic reducts (see e.g., [3]). It has been used for example for generation of so called dynamic decision rules. Experiments on different data sets have proven that these methods are applicable to large data sets.

Our approach is strongly related to propositional reasoning [58] and progress in propositional reasoning will bring further progress in the developing of our methods. It is important to note that our methodology allows to construct heuristics that have a very important *approximation property*. It may be formulated as follows: expressions (i.e., implicants) generated by heuristics *close* to prime implicants define approximate solutions for the problem [58]. This property is important since the time complexity of heuristics generating implicants close to prime implicants may be much lower than for generating prime implicants.

In the sequel we will discuss different kinds of reducts and their applications in deriving different forms of knowledge from data.

B6.2.1 Reducts in information systems and decision systems

We start from reducts of information systems. Given an $\mathcal{A} = (U, A)$, a *reduct* is a minimal set of attributes $B \subseteq A$ such that $IND_{\mathcal{A}}(B) = IND_{\mathcal{A}}(A)$. In other words, a reduct is a minimal set of attributes from A that preserves the original classification defined by the set A of attributes. Finding a minimal reduct is NP-hard [62]; one can also show that for any m there exists an information system with m attributes having an exponential number of reducts. There fortunately exist good heuristics that compute sufficiently many reducts in an acceptable time.

Let $\mathcal{A}$ be an information system with n objects. The *discernibility matrix* of $\mathcal{A}$ is a symmetric $n \times n$ matrix with entries c_{ij} as given below. Each entry consists of the set of attributes upon which objects x_i and x_j differ.

$$c_{ij} = \{a \in A \mid a(x_i) \neq a(x_j)\} \text{ for } i, j = 1, ..., n.$$

A *discernibility function* $f_{\mathcal{A}}$ for an information system $\mathcal{A}$ is a Boolean function of m Boolean variables $a_1^*, ..., a_m^*$ (corresponding to the attributes $a_1, ..., a_m$) defined by

$$f_{\mathcal{A}}(a_1^*, ..., a_m^*) = \bigwedge \left\{ \bigvee c_{ij}^* \mid 1 \leq j \leq i \leq n,\, c_{ij} \neq \emptyset \right\}$$

where $c_{ij}^* = \{a^* \mid a \in c_{ij}\}$. In the sequel we will write a_i instead of a_i^*.

The discernibility function $f_{\mathcal{A}}$ describes constraints which should be preserved in order to maintain discernibility between all pairs of discernible objects from $\mathcal{A}$. It requires to keep at least one attribute from each non-empty entry of the discernibility matrix, i.e., corresponding to any pair of discernible objects. It has been shown [62] that the sets of all minimal sets of attributes preserving discernibility between objects, i.e., reducts, correspond to prime implicants of the discernibility function $f_{\mathcal{A}}$.

The intersection of all reducts is called *core*.

In general, the decision is not constant for the indiscernibility classes. Let $\mathcal{A} = (U, A, d)$ be a decision system. The *generalized decision in* $\mathcal{A}$ is the function $\partial_A : U \longrightarrow \mathcal{P}(V_d)$ defined by $\partial_A(x) = \{i \mid \exists x' \in U\ x'\, IND(A)\, x \text{ and } d(x') = i\}$. A decision system $\mathcal{A}$ is called *consistent (deterministic)*, if $|\partial_A(x)| = 1$ for any $x \in U$, otherwise $\mathcal{A}$ is *inconsistent (non-deterministic)*. Any set consisting of all objects with the same generalized decision value is called a *generalized decision class*.

It is easy to see that a decision system $\mathcal{A}$ is consistent if, and only if, $POS_A(d) = U$. Moreover, if $\partial_B = \partial_{B'}$, then $POS_B(d) = POS_{B'}(d)$ for any pair of non-empty sets $B, B' \subseteq A$. Hence the definition of a decision-relative reduct: a subset $B \subseteq A$ is a *relative reduct* if it is a minimal set such that $POS_A(d) = POS_B(d)$. Decision-relative reducts may be found from a discernibility matrix: $M^d(\mathcal{A}) = (c_{ij}^d)$ assuming $c_{ij}^d = c_{ij} - \{d\}$ if $(|\partial_A(x_i)| = 1$ or $|\partial_A(x_j)| = 1)$ and $\partial_A(x_i) \neq \partial_A(x_j)$, $c_{ij}^d = \emptyset$, otherwise. Matrix $M^d(\mathcal{A})$ is called *the decision-relative discernibility matrix of* $\mathcal{A}$. Construction of *the decision-relative discernibility function* from this matrix follows the construction of the discernibility

function from the discernibility matrix. It has been shown [62] that the set of *prime implicants* of $f_M^d(\mathcal{A})$ defines the set of all *decision-relative reducts* of $\mathcal{A}$.

In some applications, instead of reducts we prefer to use their approximations called α-reducts, where $\alpha \in [0,1]$ is a real parameter. For a given information system $\mathcal{A} = (U, A)$, the set of attributes $B \subseteq A$ is called α-reduct if B has a non-empty intersection with at least $\alpha \cdot 100\%$ of the non-empty sets $c_{i,j}$ of the discernibility matrix of $\mathcal{A}$.

B6.2.2 Reducts and Boolean reasoning: Examples of applications

We will present examples showing how a combination of rough set methods with Boolean reasoning may be successfully used to solve several KDD problems. Reducts are the crucial constructs. They are (prime) implicants of suitably chosen Boolean functions expressing discernibility conditions which should be preserved during reduction.

Feature selection

Selection of relevant features is an important problem and has been extensively studied in Machine Learning and Pattern Recognition (see e.g., [30]). It is also a very active research area in the rough set community.

One of the first ideas [44] was to consider the *core* of the reduct set of the information system $\mathcal{A}$ as the source of relevant features. One can observe that relevant feature sets, in a sense used by the machine learning community, can be interpreted in most cases as the decision-relative reducts of decision systems obtained by adding appropriately constructed decisions to a given information system.

Another approach is related to dynamic reducts (for references see e.g., [51]). The attributes are considered relevant if they belong to dynamic reducts with a sufficiently high stability coefficient, i.e., they appear with sufficiently high frequency in random samples of a given information system. Several experiments (see [51]) show that the set of decision rules based on such attributes is much smaller than the set of all decision rules. At the same time the quality of classification of new objects increases or does not change if one only considers rules constructed over such relevant features.

The idea of attribute reduction can be generalized through an introduction of a concept of *significance of attributes* which enables to evaluate attributes not only in the two-valued scale *dispensable – indispensable* but also in the multi-value case by assigning to an attribute a real number from the interval [0,1] that expresses the importance of an attribute in the information table.

Significance of an attribute can be evaluated by measuring the effect of removing the attribute from an information table.

Let C and D be sets of condition and decision attributes, respectively, and let $a \in C$ be a condition attribute. It was shown previously that the number $\gamma(C, D)$ expresses the degree of dependency between attributes C and D, or the accuracy of the approximation of U/D by C. It may be now checked how the

coefficient $\gamma(C,D)$ changes when attribute a is removed. In other words, what is the difference between $\gamma(C,D)$ and $\gamma((C-\{a\},D)$. The difference is normalized and the significance of attribute a is defined by

$$\sigma_{(C,D)}(a) = \frac{(\gamma(C,D) - \gamma(C-\{a\},D))}{\gamma(C,D)} = 1 - \frac{\gamma(C-\{a\},D)}{\gamma(C,D)},$$

Coefficient $\sigma_{C,D}(a)$ can be understood as a classification error which occurs when attribute a is dropped. The significance coefficient can be extended to sets of attributes as follows:

$$\sigma_{(C,D)}(B) = \frac{(\gamma(C,D) - \gamma(C-B,D))}{\gamma(C,D)} = 1 - \frac{\gamma(C-B,D)}{\gamma(C,D)}.$$

Another possibility is to consider as relevant the features that come from approximate reducts of sufficiently high quality.

Any subset B of C is called an *approximate reduct* of C and the number

$$\varepsilon_{(C,D)}(B) = \frac{(\gamma(C,D) - \gamma(B,D))}{\gamma(C,D)} = 1 - \frac{\gamma(B,D)}{\gamma(C,D)},$$

is called an *error of reduct approximation*. It expresses how exactly the set of attributes B approximates the set of condition attributes C with respect to determining D. Using a similar approach, [20] showed how feature selection can be applied to identify population subgroups.

Several other methods of reduct approximation based on measures different from positive region have been developed. All experiments confirm the hypothesis that by tuning the level of approximation the quality of the classification of new objects may be increased in most cases. It is important to note that it is once again possible to use Boolean reasoning to compute the different types of reducts and to extract from them relevant approximations.

Feature extraction

The rough set community has been committed to constructing efficient algorithms for (new) feature extraction [52]. Rough set methods combined with Boolean reasoning [4] lead to several successful approaches to feature extraction. The most successful methods are: (i) discretization techniques (see, e.g., [32], [61]); (ii) methods of partitioning of nominal attribute value sets (see e.g. [34], [61]) and (iii) combinations of the above methods (see e.g. [61]). The discretization problems and symbolic value partition problems are NP-complete or NP-hard which clearly justifies the importance of designing efficient heuristics.

Our illustrative example concerns symbolic (nominal, qualitative) attribute value grouping. We also present some experimental results of heuristics based on our methods that are applied to the case of mixed nominal and numeric attributes.

In case of symbolic value attribute (i.e., without pre-assumed order on values of given attributes), the problem of searching for new features of the form $a \in V$

is, in a sense, from practical point of view more complicated than for the real value attributes. However, it is possible to develop efficient heuristics for this case using Boolean reasoning.

Let $\mathcal{A} = (U, A \cup \{d\})$ be a decision table. Any function $P_a : V_a \to \{1, \ldots, m_a\}$ (where $m_a \leq |V_a|$) is called *a partition of* V_{a_i}. The *rank of* P_{a_i} is the value $rank\,(P_i) = |P_{a_i}\,(V_{a_i})\,|$. The family of partitions $\{P_a\}_{a \in B}$ is *consistent with* B ($B-consistent$) iff the condition $[(u, u') \notin IND(B/\{d\})$ implies $\exists_{a \in B}[P_a(a(u)) \neq P_a(a(u'))]]$ holds for any $(u, u') \in U$. It means that if two objects u, u' are discerned by B and d, then they must be discerned by partition attributes defined by $\{P_a\}_{a \in B}$. We consider the following optimization problem

PARTITION PROBLEM: SYMBOLIC VALUE PARTITION PROBLEM:
Given a decision table $\mathcal{A} = (U, A \cup \{d\})$ and a set of attributes $B \subseteq A$, search for the minimal $B-consistent$ family of partitions (i.e., such $B-consistent$ family $\{P_a\}_{a \in B}$ that $\sum_{a \in B} rank\,(P_a)$ is minimal).

In order to discern between pairs of objects will use new binary features $a_v^{v'}$ (for $v \neq v'$) defined by $a_v^{v'}(x, y) = 1$ iff $a(x) = v \neq v' = a(y)$. One can apply Johnson's heuristics [18] for the new matrix with these attributes to search for minimal set of new attributes that discerns all pairs of objects from different decision classes. After extracting these sets, for each attribute a_i we construct graph $\Gamma_a = \langle V_a, E_a \rangle$ where E_a is defined as the set of all new attributes (propositional variables) found for the attribute a. Any vertex coloring of Γ_a defines a partition of V_a. The colorability problem is solvable in polynomial time for $k = 2$, but remains NP-complete for all $k \geq 3$. However, similarly to discretization, it is possible to apply some efficient heuristics searching for optimal partition.

Let us consider an example Fig. 1 of a decision table presented in Table 1 and (a reduced form) of its discernibility matrix in Table 2.

Fig. 1. A decision table and its discernibility matrix

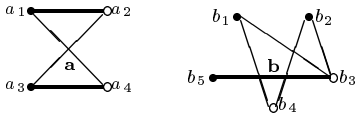
$\mathcal{A}$	**a**	**b**	**d**
u_1	a_1	b_1	0
u_2	a_1	b_2	0
u_3	a_2	b_3	0
u_4	a_3	b_1	0
u_5	a_1	b_4	1
u_6	a_2	b_2	1
u_7	a_2	b_1	1
u_8	a_4	b_2	1
u_9	a_3	b_4	1
u_{10}	a_2	b_5	1

$\Longrightarrow$

$\mathcal{M}(\mathcal{A})$	u_1	u_2	u_3	u_4
u_5	$\mathbf{b}_{b_4}^{b_1}$	$\mathbf{b}_{b_4}^{b_2}$	$\mathbf{a}_{a_2}^{a_1},\ \mathbf{b}_{b_4}^{b_3}$	$\mathbf{a}_{a_3}^{a_1},\ \mathbf{b}_{b_4}^{b_1}$
u_6	$\mathbf{a}_{a_2}^{a_1},\ \mathbf{b}_{b_2}^{b_1}$	$\mathbf{a}_{a_2}^{a_1}$	$\mathbf{b}_{b_3}^{b_2}$	$\mathbf{a}_{a_3}^{a_2},\ \mathbf{b}_{b_2}^{b_1}$
u_7	$\mathbf{a}_{a_2}^{a_1}$	$\mathbf{a}_{a_2}^{a_1},\ \mathbf{b}_{b_2}^{b_1}$	$\mathbf{b}_{b_3}^{b_1}$	$\mathbf{a}_{a_3}^{a_2}$
u_8	$\mathbf{a}_{a_4}^{a_1},\ \mathbf{b}_{b_2}^{b_1}$	$\mathbf{a}_{a_4}^{a_1}$	$\mathbf{a}_{a_4}^{a_2},\ \mathbf{b}_{b_3}^{b_2}$	$\mathbf{a}_{a_4}^{a_3},\ \mathbf{b}_{b_2}^{b_1}$
u_9	$\mathbf{a}_{a_3}^{a_1},\ \mathbf{b}_{b_4}^{b_1}$	$\mathbf{a}_{a_3}^{a_1},\ \mathbf{b}_{b_4}^{b_2}$	$\mathbf{a}_{a_3}^{a_2},\ \mathbf{b}_{b_4}^{b_3}$	$\mathbf{b}_{b_4}^{b_1}$
u_{10}	$\mathbf{a}_{a_2}^{a_1},\ \mathbf{b}_{b_5}^{b_1}$	$\mathbf{a}_{a_2}^{a_1},\ \mathbf{b}_{b_5}^{b_2}$	$\mathbf{b}_{b_5}^{b_3}$	$\mathbf{a}_{a_3}^{a_2},\ \mathbf{b}_{b_5}^{b_1}$

From the Boolean function $f_{\mathcal{A}}$ with Boolean variables of the form $\mathbf{a}_{v_1}^{v_2}$, one can find the shortest prime implicant: $\mathbf{a}_{a_2}^{a_1} \wedge \mathbf{a}_{a_3}^{a_2} \wedge \mathbf{a}_{a_4}^{a_1} \wedge \mathbf{a}_{a_4}^{a_3} \wedge \mathbf{b}_{b_4}^{b_1} \wedge \mathbf{b}_{b_4}^{b_2} \wedge \mathbf{b}_{b_3}^{b_2} \wedge \mathbf{b}_{b_3}^{b_1} \wedge \mathbf{b}_{b_5}^{b_3}$ which can be represented by graphs (see Fig. 2).

Fig. 2. Coloring of the attribute value graphs and the reduced table.



$\Longrightarrow$

$\mathbf{a}^{P_a}$	$\mathbf{b}^{P_b}$	**d**
1	**1**	0
2	**2**	0
1	**2**	1
2	**1**	1

We can color vertices of those graphs as it is shown in Fig. 2. The colors are corresponding to the partitions:

$$P_{\mathbf{a}}(a_1) = P_{\mathbf{a}}(a_3) = 1; \quad P_{\mathbf{a}}(a_2) = P_{\mathbf{a}}(a_4) = 2$$
$$P_{\mathbf{b}}(b_1) = P_{\mathbf{b}}(b_2) = P_{\mathbf{b}}(b_5) = 1; \quad P_{\mathbf{b}}(b_3) = P_{\mathbf{b}}(b_4) = 2.$$

At the same time one can construct a new decision table (Fig. 2).

One can extend this approach to the case when a given decision system contains nominal and numeric attributes (see e.g., [35]). The obtained heuristics are of a very good quality. Experiments with classification methods (see [35]) have been carried over decision systems using two techniques called *"train-and-test"* and *"n-fold-cross-validation"*. Table 1 shows some experimental results obtained by applying the proposed methods MD (using only discretization based on MD-heuristics using Johnson approximation strategy [32], [61]) and MD-G (using discretization and symbolic value grouping [34], [61]) to the classification tasks for some data tables from the "UC Irvine repository". The results reported in [12] are summarized in columns labeled by S-ID3 and C4.5 in Table 1). Let us note that the heuristics MD and MD-G are also very efficient with respect to time complexity.

In the case of real value attributes one can search for features in the feature set that contains the characteristic functions of half-spaces determined by hyper-planes or parts of spaces defined by more complex surfaces in the multi-dimensional spaces. Genetic algorithms have been applied in searching for semi-optimal hyper-planes [32]. The reported results show a substantial increase in the quality of classification of unseen objects but at the price of increased time for searching for a semi-optimal hyper-plane.

Decision rules

Reducts serve the purpose of inducing *minimal* decision rules. Any such rule contains a minimal number of descriptors in the conditional part so that their conjunction defines the largest subset of a generalized decision class (decision class, if the decision table is deterministic). Hence, information included in conditional part of any minimal rule is sufficient for predicting the generalized decision value for all objects satisfying this part. The conditional parts of minimal rules define largest object sets relevant for generalized decision classes approximation. It turns out that the conditional parts of minimal rules can be computed (by using

Names of Tables	Classification accuracy			
	S-ID3	C4.5	MD	MD-G
Australian	78.26	85.36	83.69	84.49
Breast (L)	62.07	71.00	69.95	69.95
Diabetes	66.23	70.84	71.09	76.17
Glass	62.79	65.89	66.41	69.79
Heart	77.78	77.04	77.04	81.11
Iris	96.67	94.67	95.33	96.67
Lympho	73.33	77.01	71.93	82.02
Monk-1	81.25	75.70	100	93.05
Monk-2	69.91	65.00	99.07	99.07
Monk-3	90.28	97.20	93.51	94.00
Soybean	100	95.56	100	100
TicTacToe	84.38	84.02	97.7	97.70
Average	78.58	79.94	85.48	87.00

Table 1. A quality comparison of various decision tree methods and our heuristics. Abbreviations: MD: MD-heuristic; MD-G: MD-heuristic with symbolic value partition

Boolean reasoning) as the so-called reducts relative to objects or local reducts (see e.g., [60], [3]). Once the reducts have been computed, the conditional parts of rules are easily constructed by laying the reducts over the original decision system and reading off the values. In the discussed case the generalized decision value is preserved during the reduction. One can consider stronger constraints which should be preserved. For example, in [65] the constraints are described by probability distributions corresponding to information signatures of objects. Once more, the same methodology can be used to compute the reducts corresponding to these constraints.

The main challenge in inducing rules from decision systems lies in determining which attributes should be included in the conditional part of the rule. Using the outlined above strategy first minimal rules are computed. Their conditional parts describe largest object sets (definable by conjunctions of descriptors) with the same generalized decision value in a given decision system. Hence, they create the largest sets still relevant for defining the decision classes (or sets of decision classes when the decision system is inconsistent). Although such minimal decision rules can be computed, this approach can result in set of rules with an unsatisfactory quality of classification. These rules will be too detailed and will over-fit so that unseen cases will be poorly classified. Shorter rules should rather be synthesized. Although they will not be perfect on the known cases there is a good chance that they will be of high quality when classifying new cases. They can be constructed by computing approximations of the above mentioned reducts. Approximations of reducts received by dropping some descriptors from the conditional parts of minimal rules define larger sets, not entirely included in the decision classes but included to a satisfactory degree. It means that these shorter descriptions may be more relevant for decision class (concept)

approximation than the exact reducts. This leads to the following observation: if dropping a descriptor from the conditional part provides a description of the object set which is almost included in the approximated decision class, then this descriptor is a good candidate for removal. [1] use systematic approaches to pruning sets of classification rules.

For estimation of the quality of decision classes approximation global measures based on the positive region [60] or entropy [11] are used. Methods of boundary region thinning [76] can be based, e.g., on the idea that neighborhoods included in decision classes in satisfactory degree can be treated as parts of the lower approximations of decision classes. Hence the lower approximations of decision classes are enlarged and decision rules generated for them are usually stronger (e.g., they are supported by more examples). The degree of inclusion is tuned experimentally to achieve, e.g., high classification quality of new cases.

When a set of rules has been induced from a decision system containing a set of training examples, they can be used to classify new objects. However, to resolve conflict between different decision rules recognizing new objects one should develop strategies for resolving conflicts between them when they are voting for different decisions (see the bibliography in [51] and [52]). Recently [70], it has been shown that rough set methods can be used to learn from data the strategy for conflict resolving between decision rules when they are classifying new objects contrary to existing methods using some fixed strategies.

α-reducts and association rules

In this section we discuss a relationship between association rules [2] and approximations of reducts.

We consider formulae called *templates* that are conjunctions of descriptors. Templates will be denoted by $\mathbf{T}$, $\mathbf{P}$, $\mathbf{Q}$ and descriptors by D with or without subscripts. $support_{\mathcal{A}}(\mathbf{T})$ denotes the cardinality of $|\mathbf{T}|_{\mathcal{A}}$ and $confidence_{\mathcal{A}}(\mathbf{P} \rightarrow \mathbf{Q})$ denotes the ratio $support_{\mathcal{A}}(\mathbf{P} \wedge \mathbf{Q})/support_{\mathcal{A}}(\mathbf{P})$.

The reduct approximations mentioned above are descriptions of the object sets matched by templates. They describe these sets in an approximate sense expressed by coefficients called support and confidence.

There are two main steps rule generation methods for a given information system $\mathcal{A}$ and two parameters (support s and confidence c):

1. Extract from the data as many as possible templates $\mathbf{T} = D_1 \wedge D_2 ... \wedge D_k$ such that $support_{\mathcal{A}}(\mathbf{T}) \geq s$ and $support_{\mathcal{A}}(\mathbf{T} \wedge D) < s$ for any descriptor D different from descriptors of $\mathbf{T}$ (i.e., generation of maximal templates among those supported by more than s objects);
2. Search for a partition $\mathbf{T} = \mathbf{P} \wedge \mathbf{Q}$ for any of generated template $\mathbf{T}$ satisfying the following conditions:
   (a) $support_{\mathcal{A}}(\mathbf{P}) < \frac{support_{\mathcal{A}}(\mathbf{T})}{c}$
   (b) $\mathbf{P}$ has the shortest length among templates satisfying the previous condition.

The second step can be solved using rough set methods and Boolean reasoning approach.

Let $\mathbf{T} = D_1 \wedge D_2 \wedge \ldots \wedge D_m$ be a template with $support_{\mathcal{A}}(\mathbf{T}) \geq s$. For a given confidence threshold $c \in (0;1)$, the decomposition $\mathbf{T} = \mathbf{P} \wedge \mathbf{Q}$ is called c-irreducible if $confidence_{\mathcal{A}}(\mathbf{P} \rightarrow \mathbf{Q}) \geq c$ and for any decomposition $\mathbf{T} = \mathbf{P}' \wedge \mathbf{Q}'$ such that $\mathbf{P}'$ is a sub-template of $\mathbf{P}$, we have $confidence_{\mathcal{A}}(\mathbf{P}' \rightarrow \mathbf{Q}') < c$.

Now we explain that the problem of searching for c-irreducible association rules from the given template is equivalent to the problem of searching for local α-reducts (for some α) from a decision table. The last problem is a well known problem in rough set theory.

Let us define a new decision table $\mathcal{A}|_{\mathbf{T}} = (U, A|_{\mathbf{T}}, d)$ from the original information system $\mathcal{A}$ and the template $\mathbf{T}$ by

1. $A|_{\mathbf{T}} = \{a_{D_1}, a_{D_2}, ..., a_{D_m}\}$ is a set of attributes corresponding to the descriptors of $\mathbf{T}$ such that $a_{D_i}(u) = \begin{cases} 1 \text{ if the object } u \text{ satisfies } D_i, \\ 0 \text{ otherwise.} \end{cases}$
2. the decision attribute d determines if the object satisfies template $\mathbf{T}$, i.e., $d(u) = \begin{cases} 1 \text{ if the object } u \text{ satisfies } \mathbf{T}, \\ 0 \text{ otherwise.} \end{cases}$

The following facts [61], [36] describe the relationship between association rules and approximations of reducts.

For the given information table $\mathcal{A} = (U, A)$, the template $\mathbf{T}$, the set of descriptors $\mathbf{P}$. The implication $\left(\bigwedge_{D_i \in \mathbf{P}} D_i \longrightarrow \bigwedge_{D_j \notin \mathbf{P}} D_j\right)$ is

1. 100%-irreducible association rule from $\mathbf{T}$ if and only if $\mathbf{P}$ is a reduct in $\mathcal{A}|_{\mathbf{T}}$.
2. c-irreducible association rule from $\mathbf{T}$ if and only if $\mathbf{P}$ is an α-reduct of $\mathcal{A}|_{\mathbf{T}}$, where $\alpha = 1 - (\frac{1}{c} - 1)/(\frac{n}{s} - 1)$, n is the total number of objects from U and $s = support_{\mathcal{A}}(\mathbf{T})$.

It can be shown that the problem of searching for the shortest α-reducts is NP-hard [36]. From the above facts it follows that extracting association rules from data is strongly related to extraction of reduct approximations [36].

Decomposition of large data tables

Several methods based on rough sets have been developed in order to deal with large data tables, i.e., in order to generate strong decision rules for such tables. We will discuss one of the methods based on a decomposition of tables using patterns, called templates, that describe regular sub-domains of the universe (e.g., they describe a large number of customers having a large number of common features).

Long templates with large support are preferred in many Data Mining tasks. Several quality functions can be used to compare templates. For example they can be defined by $quality^1_{\mathcal{A}}(\mathbf{T}) = support_{\mathcal{A}}(\mathbf{T}) + length(\mathbf{T})$ and $quality^2_{\mathcal{A}}(\mathbf{T}) = support_{\mathcal{A}}(\mathbf{T}) \times length(\mathbf{T})$. Problems of high quality template generation (using

different optimization criteria) are of high computational complexity. However, efficient heuristics have been developed for solving them (see e.g., [2, 75]), [34]).

Templates extracted from data are used to decompose large data tables. Consequently, a decision tree is built. Its internal nodes are labeled by the templates extracted from the data, and the edges outgoing from them are labeled by 0 (false) and 1 (true). Any leaf is labeled by a subtable (sub-domain) consisting of all objects from the original table matching all templates or their complements appearing on the path from the root of the tree to the leaf. The process of the decomposition is continued until the size of subtables attached to leaves is feasible for the rough set algorithms at hand (e.g., decision rules for them can be generated efficiently). The reported experiments show that such decomposition returns interesting patterns of regular sub-domains of large data tables (for references see [34], [37], [51] and [52]).

It is also possible to search for patterns that are almost included in the decision classes, i.e., default rules [31]. For a presentation of generating default rules see the bibliography in [51] and [52].

Conclusions

We have shown that rough set theory constitutes a sound basis for KDD: minimal concept descriptions, classifiers, dependencies, etc, are systematically synthesized and their quality can be evaluated using statistical methods. Features can be extracted and selected. The strict concepts are extended with the approximate ones that usually improve the quality of classification. Methods for processing very large data sets are developed. Successful software tools are implemented and used by thousands of researchers world-wide.

There has been done a substantial progress in developing rough set methods for KDD (like methods for extraction from data rules, partial or total dependencies, methods for elimination of redundant data, methods dealing with missing data, dynamic data and others reported e.g., in [6], [7], [8], [16], [19], [26], [31], [32], [40], [51], [52], [53], [78]). New methods for extracting patterns from data (see e.g., [23], [37], [31]), [22], [47]), decomposition of decision systems (see e.g., [37]) as well as a new methodology for data mining in distributed and multi-agent systems (see e.g., [50]) have been reported. Recently, rough set based methods have been proposed for data mining in very large relational data bases.

There are numerous areas of successful applications of rough set software systems (see [52] and `http://www.idi.ntnu.no/~aleks/rosetta/` for the ROSETTA system). Many interesting case studies are reported (for references see e.g., [51, 52], [40] and the bibliography in these books, in particular [7], [16], [22], [71], [78]).

We would like to mention some generalizations of rough set approach like rough mereological approach (see e.g., [54], [49]). The inclusion relation $x\mu_r y$ with the intended meaning *x is a part of y to a degree r* has been taken as the basic notion of the rough mereology being a generalization of the Leśniewski mereology. Rough mereology offers a methodology for synthesis and analysis of

objects in distributed environment of intelligent agents, in particular, for synthesis of objects satisfying a given specification to a satisfactory degree, i.e., objects sufficiently close to standard objects (prototypes) satisfying the specification. Moreover, rough mereology has been recently used [50] for developing foundations of the *information granule calculus*, an attempt towards a formalization of the Computing–with–Words paradigm, recently formulated by Lotfi Zadeh [72], [73]. Let us also note that one of the prospects for rough mereological applications is to look for algorithmic methods of extracting logical structures from data such as, for instance, finding relational structures corresponding to relevant feature extraction, synthesizing default rules (approximate decision rules), constructing connectives for uncertainty coefficients propagation and synthesizing schemes of approximate reasoning creating a higher level knowledge extracted from data (e.g. qualitative schemes of reasoning). The development of such methods is crucial to further progress in many applications. It is also one of the central issues of KDD as pointed out in [13].

Several other generalizations of rough sets have been investigated and some of them have been used for real life data analysis (see e.g., [76], [5], [42], [14], [24], [41], [27], [59], [50]).

Finally, we would like to point out that the algebraic and logical aspects of rough sets have been intensively studied since the beginning of rough set theory. The reader interested in that topic is referred to the bibliography in [51].

Acknowledgment. The number of people that are involved in this research is very large and we can only express our general thanks to all colleagues, collaborators, reviewers, students, wives and funding agencies who have contributed to this article in very many direct and indirect ways and who supported us over the years. Authors would like thank to Willi Klösgen, Jan Żytkow and to the anonymous reviewer for their criticism and comments on the previous version of the paper.

Jan Komorowski has been supported in part by the European Union 4$^{\text{th}}$ Framework Telematics project CardiAssist, and by the Norwegian Research Council (NFR) grants #74467/410, #110177/730. Lech Polkowski has been supported by the grant No 8T11C02417 from the State Committee for Scientific Research (KBN) of the Republic of Poland. Andrzej Skowron has been supported by the grant of the Wallenberg Foundation, by the ESPRIT-CRIT 2 project #20288, and by grant from the State Committee for Scientific Research (KBN) of the Republic of Poland.

References

1. T. Ågotnes, J. Komorowski, T. Løken (1999), *Taming large rule models in rough set approaches,* Proceedings of the 3rd European Conference of Principles and Practice of Knowledge Discovery in Databases, September 15-18, 1999, Prague, Czech Republic, Lecture Notes in Artificial Intelligence **1704**, Springer-Verlag, Berlin, pp. 193-203.
2. R. Agrawal, H. Mannila, R. Srikant, H. Toivonen, A. Verkano(1996), *Fast discovery of association rules*, Fayyad U.M., Piatetsky-Shapiro G., Smyth P., Uthurusamy R.

(Eds.): Advances in Knowledge Discovery and Data Mining, The AAAI Press/The MIT Press 1996, pp. 307–328.

3. J. G. Bazan (1998). *A comparison of dynamic and non-dynamic rough set methods for extracting laws from decision system,* in: [51], pp. 321–365.
4. F. M. Brown (1990), *Boolean Reasoning*, Kluwer Academic Publishers, Dordrecht 1990.
5. G. Cattaneo (1998), *Abstract approximation spaces for rough theories*, In: Polkowski and Skowron [51], pp. 59–98.
6. J. Cios, W. Pedrycz, R.W. Swiniarski (1998), *Data Mining in Knowledge Discovery*, Academic Publishers (in press).
7. A. Czyżewski (1998), *Soft processing of audio signals.* In: Polkowski and Skowron [52], pp. 147–165.
8. J. Deogun, V. Raghavan, A. Sarkar, H. Sever (1997), *Data mining: Trends in research and development*, In: Lin and Cercone [26], pp. 9–45.
9. Dougherty J., Kohavi R., Sahami M.(1995), *Supervised and unsupervised discretization of continuous features,* In: Proceedings of the Twelfth International Conference on Machine Learning, Morgan Kaufmann, San Francisco, CA.
10. I. Duentsch, G. Gediga (1997): *Statistical evaluation of rough set dependency analysis,* International Journal of Human-Computer Studies **46**, pp. 589-604.
11. I. Duentsch, G. Gediga (2000): *Rough set data analysis,* in: Encyclopedia of Computer Science and Technology, Marcel Dekker (to appear).
12. J. Friedman, R. Kohavi, Y. Yun (1996), *Lazy Decision Trees*, Proc. of AAAI-96, pp. 717–724.
13. U. Fayyad, G. Piatetsky-Shapiro, G. (Eds.) (1996), *Advances in knowledge discovery and data mining*, MIT/AAAI Press.
14. S. Greco, B. Matarazzo, R. Słowiński (1998), *Rough Approximation of a Preference Relation in a Pairwise Comparison Table.* In: Polkowski and Skowron [52] pp. 13–36.
15. M. R. Chmielewski, J. W. Grzymała-Busse(1994), *Global discretization of attributes as preprocessing for machine learning,* Proceedings of the Third International Workshop on Rough Sets and Soft Computing (RSSC'94),San Jose State University, San Jose, California, USA, November 10–12, pp. 294–301.
16. J.W. Grzymała–Busse (1998), *Applications of the rule induction system LERS.* In: Polkowski and Skowron [51], pp. 366–375.
17. P. J. Huber (1981), *Robust statistics*, Wiley, New York.
18. D.S. Johnson (1974), *Approximation algorithms for combinatorial problems*, J. of Computer and System Sciences, 9:256–278.
19. J. Komorowski, J. Żytkow (Eds.) (1997), *The First European Symposium on Principles of Data Mining and Knowledge Discovery (PKDD'97).* June 25–27, Trondheim, Norway, Lecture Notes in Artificial Intelligence **1263**, Springer-Verlag, Berlin pp. 1–396.
20. J. Komorowski, A. Øhrn (1999), *Modelling Prognostic Power of cardiac tests using rough sets*, Artificial Intelligence in Medicine, 15(2):167–191, January.
21. J. Komorowski, Z. Pawlak, L. Polkowski, and A. Skowron (1999), *Rough sets: A tutorial*, in: S.K. Pal and A. Skowron (Eds.), Rough fuzzy hybridization: A new trend in decision–making, Springer-Verlag, Singapore, pp. 3–98.
22. W. Kowalczyk (1998), *Rough data modelling, A new technique for analyzing data.* In: Polkowski and Skowron [51], pp. 400–421.

23. K. Krawiec, R. Słowiński, and D. Vanderpooten (1998), *Learning decision rules from similarity based rough approximations.* In: Polkowski and Skowron [52], pp. 37–54.
24. M. Kryszkiewicz (1997), *Generation of rules from incomplete information systems.* In: Komorowski and Żytkow [19] pp. 156–166.
25. Langley, P., Simon, H.A., Bradshaw, G.L., Żytkow, J.M. (1987): Scientific Discovery, Computational Explorations of the Creative Processes, The MIT Press, Cambridge, Massachusetts.
26. T. Y. Lin, N. Cercone (Eds.) (1997), *Rough sets and data mining. Analysis of imprecise data*, Kluwer Academic Publishers, Boston.
27. T.Y. Lin (1989), *Granular computing on binary relations I, II.* In: Polkowski and Skowron [51], pp. 107–140.
28. V.M. Marek, M. Truszczyński (1999), *Contributions to the theory of rough sets*, Fundamenta Informaticae **39(4)**, 1999, pp. 389-409.
29. D. Michie, D.J. Spiegelhalter, C.C. Taylor (Eds.). *Machine learning, Neural and Statistical Classification.* Ellis Horwood, New York, 1994.
30. T.M. Mitchell (1997), *Machine Learning*, Mc Graw-Hill, Portland.
31. T. Mollestad and J. Komorowski (1998), *A Rough Set Framework for Propositional Default Rules Data Mining.* In: S.K. Pal and A. Skowron (Eds.) (1998), Rough – fuzzy hybridization: New trend in decision making, Springer–Verlag, Singapore.
32. H. S. Nguyen (1997), *Discretization of Real Value Attributes, Boolean Reasoning Approach*, Ph.D. Dissertation, Warsaw University (1997), pp. 1–90.
33. H. S. Nguyen (1999), *Efficient SQL-learning Method for Data Mining in Large Data Bases*, Proceedings of the Sixteenth International Joint Conference on Artificial Intelligence (IJCAI'99), 1999, pp. 806–811.
34. S. H. Nguyen (2000), *Data regularity analysis and applications in data mining*, Ph.D. Dissertation, Warsaw University, 2000.
35. H.S. Nguyen and S.H. Nguyen (1998), *Pattern extraction from data*, Fundamenta Informaticae **34**, pp. 129–144.
36. H.S. Nguyen and S.H. Nguyen (1999), *Rough sets and association rule generation,* Fundamenta Informaticae **40/4** pp. 383–405.
37. S. H. Nguyen, A. Skowron, P. Synak (1998), *Discovery of data patterns with applications to decomposition and classification problems.* In: Polkowski and Skowron [52], pp. 55–97.
38. A. Øhrn (1999), *Discernibility and Rough Sets in Medicine: Tools and Applications*, PhD thesis, Norwegian University of Science and Technology, Trondheim, IDI-rapport 1999:14.
39. E. Orłowska (Ed.) (1998), *Incomplete Information, Rough Set Analysis*, Physica–Verlag, pp. 1–613.
40. S. K. Pal, A. Skowron (1999), *Rough–fuzzy hybridization: New trend in decision making*, Springer–Verlag, Singapore (in print).
41. G. Paun, L. Polkowski, A. Skowron (1996), *Parallel communicating grammar systems with negotiations.* Fundamenta Informaticae **28/3-4**, pp. 315–330.
42. Z. Pawlak (1981), *Information systems – theoretical foundations.* Information Systems **6**, pp. 205–218.
43. Z. Pawlak (1982), *Rough sets.* International Journal of Computer and Information Sciences **11**, pp. 341–356.
44. Z. Pawlak (1991), *Rough Sets – Theoretical Aspects of Reasoning about Data*, Kluwer Academic Publishers, Dordrecht.

45. Z. Pawlak and Z. Ras, (Eds.), (1996), *Proc. Ninth International Symposium on Methodologies for Intelligent Systems (ISMIS'96)*, June, Springer Verlag, LNAI.
46. Z. Pawlak, A. Skowron (1999), *Rough set rudiments*, Bulletin of the International Rough Set Society **3/4**, pp. 181–185.
47. Z. Piasta, A. Lenarcik (1998), *Rule induction with probabilistic rough classifiers.* Machine Learning (to appear).
48. L. Polkowski (2000), *On synthesis of constructs for spatial reasoning via rough mereology*, Fundamenta Informaticae (to appear).
49. L. Polkowski, A. Skowron (1996), *Rough mereology: A new paradigm for approximate reasoning*, International Journal of Approximate Reasoning **15/4**, pp. 333–365.
50. L. Polkowski, A. Skowron (1998), *Rough sets: A perspective*, In: Polkowski and Skowron [51], pp. 31–58.
51. L. Polkowski, A. Skowron (Eds.) (1998), *Rough Sets in Knowledge Discovery 1: Methodology and Applications*, Physica-Verlag, Heidelberg.
52. L. Polkowski, A. Skowron (Eds.) (1998), *Rough Sets in Knowledge Discovery 2: Applications, Case Studies and Software Systems*, Physica-Verlag, Heidelberg.
53. L. Polkowski, A. Skowron (Eds.) (1998), *Proc. First International Conference on Rough Sets and Soft Computing* – RSCTC'98, Warszawa, Poland, June 22–27, Springer-Verlag, LNAI **1424**.
54. L. Polkowski, A. Skowron (1999), *Towards adaptive calculus of granules*, In: [74], **1**, pp. 201–227.
55. J.R. Quinlan (1993), *C4.5. Programs for machine learning.* Morgan Kaufmann, San Mateo, CA.
56. J. J. Rissanen (1978), Modeling by Shortest Data Description, Automatica **14**, pp. 465-471.
57. J.F. Roddick J.F., M. Spiliopoulou (1999), *A bibliography of temporal, spatial, and temporal data mining research*, Newsletter of the Special Interest Group (SIG) on Knowledge Discovery & Data Mining **1/1**, pp. 34-38.
58. B. Selman, H. Kautz and D. McAllester. *Ten challenges in propositional reasoning and search.* Proc. IJCAI'97, Japan.
59. Z.W. Ras (1996), *Cooperative knowledge–based systems.* Journal of the Intelligent Automation Soft Computing **2/2** (special issue edited by T.Y. Lin), pp. 193–202
60. Skowron A. (1995), *Synthesis of adaptive decision systems from experimental data.* In: A. Aamodt, J. Komorowski (Eds), Proc. of the Fifth Scandinavian Conference on Artificial Intelligence (SCAI'95), May 1995, Trondheim, Norway, IOS Press, Amsterdam, pp. 220–238.
61. A. Skowron, H.S. Nguyen (1999), *Boolean reasoning scheme with some applications in data mining,* Proceedings of the 3-rd European Conference on Principles and Practice of Knowledge Discovery in Databases, September 1999, Prague Czech Republic, Lecture Notes in Computer Science **1704**, pp. 107–115.
62. A. Skowron, C. Rauszer (1992), *The Discernibility Matrices and Functions in Information Systems.* In: Słowiński [66], pp. 331–362.
63. A. Skowron, J. Stepaniuk (1996), *Tolerance Approximation Spaces*, Fundamenta Informaticae **27**, pp. 245–253.
64. A. Skowron, J. Stepaniuk, S. Tsumoto (1999), *Information Granules for Spatial Reasoning*, Bulletin of the International Rough Set Society **3/4**, pp. 147–154.
65. D. Ślęzak (1998), *Approximate reducts in decision tables,* In: Proceedings of the Sixth International Conference, Information Processing and Management of Un-

certainty in Knowledge-Based Systems (IPMU'96) Vol. 3, July 1–5, Granada, Spain, pp. 1159–1164.
66. R. Słowiński, (Ed.) (1992), *Intelligent Decision Support – Handbook of Applications and Advances of the Rough Sets Theory*, Dordrecht, Kluwer Academic Publishers.
67. R. Słowiński, D. Vanderpooten (1995), *Similarity relation as a basis for rough approximations.* In: P. Wang (Ed.): Advances in Machine Intelligence & Soft Computing, Bookwrights, Raleigh NC (1997) pp. 17–33.
68. R. Słowiński, D. Vanderpooten (1999), *A generalized definition of rough approximations based on similarity.* IEEE Trans. on Data and Knowledge Engineering (to appear).
69. J.A. Swets, R.M. Picket (1982), *Evaluation of Diagnostic Systems: Methods from Signal Detection Theory*, Academic Press, New York, NY, 1982.
70. M. Szczuka (1999), *Symbolic and neural network methods for classifiers construction*, Ph.D. Dissertation, Warsaw University.
71. S. Tsumoto (1998), *Modelling diagnostic rules based on rough sets.* In: Polkowski and Skowron [53], pp. 475–482.
72. L.A. Zadeh (1996), *Fuzzy logic = computing with words*, IEEE Trans. on Fuzzy Systems **4**, pp. 103–111.
73. L.A. Zadeh (1997), *Toward a theory of fuzzy information granulation and its certainty in human reasoning and fuzzy logic*, Fuzzy Sets and Systems **90**, pp. 111–127.
74. Zadeh, L.A., Kacprzyk, J. (Eds.): Computing with Words in Information/Intelligent Systems Vol. 1-2, Physica-Verlag, Heidelberg, 1999.
75. M.J. Zaki, S. Parthasarathy, M. Ogihara, W. Li (1997), *New parallel algorithms for fast discovery of association rules.* In: Data Mining and Knowledge Discovery : An International Journal, special issue on Scalable High-Performance Computing for KDD **1/4**, pp. 343–373.
76. W. Ziarko (1993), *Variable Precision Rough Set Model* , J. of Computer and System Sciences, 46, pp. 39–59.
77. W. Ziarko (ed.) (1994), *Rough Sets, Fuzzy Sets and Knowledge Discovery (RSKD'93).* Workshops in Computing, Springer–Verlag & British Computer Society, London, Berlin.
78. W. Ziarko (1998), *Rough sets as a methodology for data mining.* In: L. Polkowski A, Skowron (Eds.), Rough Sets in Knowledge Discovery 1: Methods Applications, Physica-Verlag, Heidelberg, pp. 554–576.

This article was processed using the LaTeX macro package with LMAMULT style