Rough Set Rudiments

Zdzisław Pawlak[*1] zpw@ii.pw.edu.pl
Andrzej Skowron[*2] skowron@mimuw.edu.pl

*1 Institute of Theoretical and Applied Informatics
Polish Academy of Sciences
ul. Baltycka 5, 44-000 Gliwice, Poland

*2 Institute of Mathematics
Warsaw University
Banacha 2, 02-097 Warsaw, Poland

Abstract: *The aim of this paper is to give the basic concepts of rough set theory.*

Keywords: Indiscernibility, Set approximations, Rough sets, Rough membership function, Reducts, Decision tables, Decision rules, Dependency of attributes.

1 Introduction

In recent years we witness a rapid grow of interest in rough set theory and its applications, worldwide.

Many international workshops, conferences and seminars included rough sets in their programs. A large number of high quality papers on various aspects of rough sets and their applications have been published in last years.

The aim of this paper is to give the basic concepts of rough set theory.

2 Basic Philosophy

The rough set philosophy is founded on the assumption that with every object of the universe of discourse we associate some information (data, knowledge). For example, if objects are patients suffering from a certain disease, symptoms of the disease form information about patients. Objects characterized by the same information are indiscernible (similar) in view of the available information about them. The indiscernibility relation generated in this way is the mathematical basis of rough set theory.

Any set of all indiscernible (similar) objects is called an elementary set, and forms a basic granule (atom) of knowledge about the universe. Any union of some elementary sets is referred to as crisp (precise) set – otherwise the set is rough (imprecise, vague).

Consequently each rough set has boundary-line cases, i.e., objects which cannot be with certainty classified neither as members of the set nor of its complement. Obviously crisp sets have no boundary-line elements at all. That means that boundary-line cases cannot be properly classified by employing the available knowledge.

Thus, the assumption that objects can be "seen" only through the information available about them leads to the view that knowledge has granular structure. Due to the granularity of knowledge some objects of interest cannot be discerned and appear as the same (or similar). As a consequence vague concepts, in contrast to precise concepts, cannot be characterized in terms of information about their elements. Therefore, in the proposed approach, we assume that any vague concept is replaced by a pair of precise concepts – called the lower and the upper approximation of the vague concept. The lower approximation consists of all objects which surely belong to the concept and the upper approximation contains all objects which possibly belong to the concept. Obviously, the difference between the upper and the lower approximation constitutes the boundary region of the vague concept. Approximations are two basic operations in rough set theory.

3 Approximations and Rough Sets

As mentiones in Section 2, the starting point of rough set theory is the indiscernibility relation, generated by information about objects of interest. The indiscernibility relation is intended to express the fact that due to the lack of knowledge we are unable to discern some objects employing the available information. It means that, in general, we are unable to deal with each particular object but we have to consider clusters of indiscernible objects, as fundamental concepts of our theory.

Now we present above considerations more formally.

Suppose we are given two finite, non-empty sets U and A, where U is the *universe* of *objects*, and A – a set of *attributes*. The pair (U, A) is called an *information table*. With every attribute $a \in A$ we associate a set V_a, of its *values*, called the *domain* of a. Any subset B of A determines a binary relation $I(B)$ on U, called an *indiscernibility relation*, defined as follows:

$$xI(B)y \text{ if and only if } a(x) = a(y) \text{ for every } a \in B,$$

where $a(x)$ denotes the value of attribute a for object x.

Obviously $I(B)$ is an equivalence relation. The family of all equivalence classes of $I(B)$, i.e., the partition determined by B, will be denoted by $U/I(B)$, or simply U/B; an equivalence class of $I(B)$, i.e., the block of the partition U/B, containing x will be denoted by $B(x)$.

If $(x, y) \in I(B)$ we will say that x and y are B-*indiscernible*. Equivalence classes of the relation $I(B)$ (or blocks of the partition U/B) are referred to as B-*elementary sets*. In the rough set approach the elementary sets are the basic building blocks (concepts) of our knowledge about reality. The unions of B-*elementary sets* are called B-*definable sets*.

The indiscernibility relation will be further used to define basic concepts of rough set theory. Let us define now the following two operations on sets

$$\begin{aligned} B_*(X) &= \{x \in U : B(x) \subseteq X\}, \\ B^*(X) &= \{x \in U : B(x) \cap X \neq \emptyset\}, \end{aligned}$$

assigning to every subset X of the universe U two sets $B_*(X)$ and $B^*(X)$ called the B-*lower* and the B-*upper approximation* of X, respectively. The set

$$BN_B(X) = B^*(X) - B_*(X)$$

will be referred to as the B-*boundary region* of X.

If the boundary region of X is the empty set, i.e., $BN_B(X) = \emptyset$, then the set X is *crisp* (*exact*) with respect to B; in the opposite case, i.e., if $BN_B(X) \neq \emptyset$, the set X is referred to as *rough* (*inexact*) with respect to B.

Rough set can be also characterized numerically by the following coefficient

$$\alpha_B(X) = \frac{|B_*(X)|}{|B^*(X)|},$$

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

Several generalizations of the classical rough set approach based on approximation spaces defined by (U, R), where R is an equivalence relation (called indiscernibility relation) in U, have been reported in the literature (for references see the papers and bibliography in [PaS], [PS1], [PS2]). Let us mention two of them.

A generalized approximation space can be defined by $AS = (U, I, \nu)$ where I is the *uncertainty function* defined on U with values in the powerset $P(U)$ of U ($I(x)$ is the *neighboorhood* of x) and ν is the *inclusion function* defined on the Cartesian product $P(U) \times P(U)$ with values in the interval $[0, 1]$ measuring the degree of inclusion of sets. The lower AS_* and upper AS^* approximation operations can be defined in AS by

$$\begin{aligned} AS_*(X) &= \{x \in U : \nu(I(x), X) = 1\}, \\ AS^*(X) &= \{x \in U : \nu(I(x), X) > 0\}. \end{aligned}$$

In the classical case $I(x)$ is equal to the equivalence class $B(x)$ of the indiscernibility relation $I(B)$; in case when a tolerance (similarity) relation $\tau \subseteq U \times U$ is given we take $I(x) = \{y \in U : x\tau y\}$, i.e., $I(x)$ is equal to the tolerance class of τ defined by x. The standard inclusion relation is defined by $\nu(X, Y) = \frac{|X \cap Y|}{|X|}$ if X is non-empty, and otherwise $\nu(X, Y) = 1$. For applications it is important to have some constructive definitions of I and ν.

One can consider another way to define $I(x)$. Usually together with AS we consider some set F of formulae describing sets of objects in the universe U of AS defined by semantics $\|\cdot\|_{AS}$, i.e., $\|\alpha\|_{AS} \subseteq U$ for any $\alpha \in F$. Now, one can take the set

$$N_F(x) = \{\alpha \in F : x \in \|\alpha\|_{AS}\}$$

and $I(x) = \|\alpha\|_{AS}$ where α is selected or constructed from $N_F(x)$. Hence, more general uncertainty functions having values in $P(P(U))$ can be defined. The parametric approximation spaces are examples of such approximation spaces. These spaces have interesting applications. For example, by tuning of their parameters one can search for the optimal, under chosen criteria (e.g. the minimal description length), approximation space for concept description.

The approach based on inclusion functions has been generalized to the *rough mereological approach*. The inclusion relation $x\mu_r y$ with the intended meaning *x is a part of y in 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 in satisfactory degree or for control in such complex environment. Moreover, rough mereology has been recently used for developing foundations of the *information granule calculus*, an attempt towards formalization of the Computing with Words paradigm, recently formulated by Lotfi Zadeh.

Research on rough mereology has shown importance of another notion, namely *closeness* of complex objects (e.g., concepts). This can be defined by $xcl_{r,r'}y$ if and only if $x\mu_r y$ and $y\mu_{r'}x$.

The inclusion and closeness definitions of complex information granules are dependent on applications. However, it is possible to define the granule syntax and semantics as a basis for the inclusion and closeness definitions.

4 Rough Sets and Membership Function

Rough sets can be also introduced using a *rough membership function*, defined by

$$\mu_X^B(x) = \frac{|X \cap B(x)|}{|B(x)|}.$$

Obviously $0 \leq \mu_X^B(x) \leq 1$. The membership function $\mu_X(x)$ is a kind of conditional probability and its value can be interpreted as a degree of *certainty* to which x belongs to X.

The rough membership function, can be used to define approximations and the boundary region of a set, as shown below:

$$\begin{aligned} B_*(X) &= \{x \in U : \mu_X^B(x) = 1\}, \\ B^*(X) &= \{x \in U : \mu_X^B(x) > 0\}, \\ BN_B(X) &= \{x \in U : 0 < \mu_X^B(x) < 1\}. \end{aligned}$$

5 Decision Tables and Decision Rules

Sometimes we distinguish in an information table (U, A) a partition of A into two classes $C, D \subseteq A$ of attributes, called *condition* and *decision* (*action*) attributes, respectively. The tuple $\mathcal{A} = (U, C, D)$ is called a *decision table.*

Let $V = \bigcup\{V_a \mid a \in C\} \cup V_d$. Atomic formulae over $B \subseteq C \cup D$ and V are expressions $a = v$ called *descriptors* (*selectors*) over B and V, where $a \in B$ and $v \in V_a$. The set $\mathcal{F}(B, V)$ of formulae over B and V is the least set containing all atomic formulae over B and V and closed with respect to the propositional connectives $\wedge$ (conjunction), $\vee$ (disjunction) and $\neg$ (negation).

By $\|\varphi\|_{\mathcal{A}}$ we denote the meaning of $\varphi \in \mathcal{F}(B, V)$ in the decision table $\mathcal{A}$ which is the set of all objects in U with the property φ. These sets are defined as follows: $\|a = v\|_{\mathcal{A}} = \{x \in U \mid a(x) = v\}$, $\|\varphi \wedge \varphi'\|_{\mathcal{A}} = \|\varphi\|_{\mathcal{A}} \cap \|\varphi'\|_{\mathcal{A}}$; $\|\varphi \vee \varphi'\|_{\mathcal{A}} = \|\varphi\|_{\mathcal{A}} \cup \|\varphi'\|_{\mathcal{A}}$; $\|\neg\varphi\|_{\mathcal{A}} = U - \|\varphi\|_{\mathcal{A}}$ The formulae from $\mathcal{F}(C, V)$, $\mathcal{F}(D, V)$ are called *condition formulae of* $\mathcal{A}$ and *decision formulae of* $\mathcal{A}$, respectively.

Any object $x \in U$ belongs to a *decision class* $\|\bigwedge_{a \in D} a = a(x)\|_{\mathcal{A}}$ of $\mathcal{A}$. All decision classes of $\mathcal{A}$ create a partition of the universe U.

A *decision rule* for $\mathcal{A}$ is any expression of the form $\varphi \Rightarrow \psi$, where $\varphi \in \mathcal{F}(C, V)$, $\psi \in \mathcal{F}(D, V)$, and $\|\varphi\|_{\mathcal{A}} \neq \emptyset$. Formulae φ and ψ are referred to as the *predecessor* and the *successor* of decision rule $\varphi \Rightarrow \psi$. Decision rules are often called "$IF \ldots THEN \ldots$" rules.

Decision rule $\varphi \Rightarrow \psi$ is *true* in $\mathcal{A}$ if and only if $\|\varphi\|_{\mathcal{A}} \subseteq \|\psi\|_{\mathcal{A}}$. Otherwise one can measure its *truth degree* by introducing some inclusion measure of $\|\varphi\|_{\mathcal{A}}$ in $\|\psi\|_{\mathcal{A}}$.

Each object x of a decision table determines a *decision rule* $\bigwedge_{a \in C} a = a(x) \Rightarrow \bigwedge_{a \in D} a = a(x)$. Decision rules corresponding to some objects can have the same condition parts but different decision parts. Such rules are called *inconsistent* (*nondeterministic, conflicting, possible*); otherwise the rules are referred to as *consistent* (*certain, sure, deterministic, nonconflicting*) rules. Decision tables containing inconsistent decision rules are called *inconsistent* (*nondeterministic, conflicting*); otherwise the table is *consistent* (*deterministic, nonconflicting*).

Numerous methods have been developed for different decision rule generation (see, e.g., [PaS, PS1, PS2, PS3]).

When a set of rules have been induced from a decision table containing a set of training examples, they can be inspected to see if they reveal any novel relationships between attributes that are worth pursuing for further research. Furthermore, the rules can be applied to a set of unseen cases in order to estimate their classificatory power. For a systematic overview of rule application methods the reader is referred to papers in [PaS, PS1, PS2].

6 Dependency of Attributes

Another important issue in data analysis is discovering dependencies between attributes. Intuitively, a set of attributes D depends totally on a set of attributes C, denoted $C \Rightarrow D$, if the values of attributes from C uniquely determine the values of attributes from D. In other words, D depends totally on C, if there exists a functional dependency between values of C and D.

Formally dependency can be defined in the following way. Let D and C be subsets of A.

We will say that D *depends on* C in a *degree* k $(0 \leq k \leq 1)$, denoted $C \Rightarrow_k D$, if

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

where

$$POS_C(D) = \bigcup_{X \in U/D} C_*(X),$$

called a *positive region* of the partition U/D with respect to C, is the set of all elements of U that can be uniquely classified to blocks of the partition U/D, by means of C.

If $k = 1$ we say that D *depends totally* on C, and if $k < 1$, we say that D *depends partially* (in a *degree* k) on 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, employing attributes C and will be called the *degree of the dependency*.

It can be easily seen that if D depends totally on C then $I(C) \subseteq I(D)$. It means that the partition generated by C is finer than the partition generated by D. Notice, that the concept of dependency discussed above corresponds to that considered in relational databases.

Summing up: D is *totally* (*partially*) dependent on C, if *all* (*some*) elements of the universe U can be uniquely classified to blocks of the partition U/D, employing C.

7 Reduction of Attributes

We often face a question whether we can remove some data from a data-table preserving its basic properties, that is – whether a table contains some superfluous data.

Let us express this idea more precisely.

Let $C, D \subseteq A$, be sets of condition and decision attributes respectively. We will say that $C' \subseteq C$ is a *D-reduct* (reduct with *respect* to D) of C, if C' is a minimal subset of C such that $\gamma(C, D) = \gamma(C', D)$.

The intersection of all D-reducts is called a *D-core* (core with *respect* to D).

Because the core is the intersection of all reducts, it is included in every reduct, i.e., each element of the core belongs to some reduct. Thus, in a sense, the core is the most important subset of attributes, since none of its elements can be removed without affecting of the classification power of attributes.

Many other kinds of redutcs and their approximations are discussed in literature. It turns out that they can be efficiently computed using heuristics based on Boolean reasoning approach.

8 Discernibility and Boolean Reasoning

The ability to discern between perceived objects is important for constructing many entities like reducts, decision rules or decision algorithms. In the classical rough set approach the discernibility relation $DIS(B) \subseteq U \times U$ is defined by $xDIS(B)y$ if and only if $non(xI(B)y)$. However, this is in general not the case for the generalized approximation spaces (one can define indiscernibility by $x \in I(y)$ and discernibility by $I(x) \cap I(y) = \emptyset$ for any objects x, y).

The idea of Boolean reasoning is based on construction for a given problem P a corresponding Boolean function f_P with the following property: the solutions for the problem P can be decoded from prime implicants of the Boolean function f_P. Let us mention that to solve real-life problems it is necessary to deal with Boolean functions having large number of variables.

A successful methodology based on the discernibility of objects and Boolean reasoning has been developed for computing of many important for applications entities like 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 (for references see the papers and bibliography in [PaS], [PS1], [PS2])).

Most of the problems related to generation of the above mentioned entities are NP-complete or NP-hard. However, it was possible to develop efficient heuristics returning suboptimal solutions of the problems. The results of experiments on many data sets are very promising. They show very good quality of solutions generated by the heuristics in comparison with other methods reported in literature (e.g. with respect to the classification quality of unseen objects). Moreover, they are very efficient from the point of view of time necessary for computing of the solution.

It is important to note that the methodology allows to construct heuristics having a very important *approximation property* which can be formulated as follows: expressions generated by heuristics (i.e., implicants) *close* to prime implicants define approximate solutions for the problem.

9 Conclusions

In this paper we gave the basic concepts of rough set theory.

It turned out, however that the "basic model" of rough set presented here was not sufficient for many applications and needed some extensions. Besides, theoretical inquiry into the rough set concept also led to its various generalizations. Some of them have been mentioned in the paper.

A variety of methods for decision rules generation, reducts computation and continuous variable discretization are very important issues not discussed here. We have only emphasized the developed powerful methodology based on discernibility and Boolean reasoning for efficient computation of different entities including reducts and decision rules.

Also the relationship of rough set theory to many other theories has been extensively investigated. In particular, its relationship to fuzzy set theory the theory of evidence, Boolean reasoning methods, statistical methods, and decision theory has been clarified and seems to be thoroughly understood.

There are reports on many hybrid methods obtained by combining rough set approach with other ones like fuzzy sets, neural networks, genetic algorithms, principal component analysis, singular value decomposition.

Recently, it has been shown that rough set approach can be used for synthesis of concept approximations in distributed environment of intelligent agents. In particular, the rough set methods are used for construction of interfaces between agents equipped with different sets of concepts [SZ1].

Readers interested in the above issues are advised to consult the enclosed references (e.g. [PaS, PS1, PS2, PS3, SZ1]).

Many important issues, like for example, various logics related to rough sets and many advanced algebraic properties of rough sets are not covered by the paper. These issues have rather advanced structures and are deliberately dropped here. The reader can find details in [Or1, PaS, PS1, PS2, PS3].

Acknowledgments. The research of A. Skowron was supported by the grant ESPRIT-CRIT 2 No. 20288.

References

[LI1] Lin, T.Y. (ed.): Journal of the Intelligent Automation and Soft Computing **2/2** (1996) (special issue)

[LI1] Lin, T.Y. (ed.): International Journal of Approximate Reasoning **15/4** (1996) (special issue)

[LC1] Lin, T. Y. Cercone, N. (eds.): Rough Sets and Data Mining - Analysis of Imperfect Data. Kluwer Academic Publishers, Boston, London, Dordrecht (1997)

[LW1] Lin, T. Y., Wildberger A. M., (eds.): The Third International Workshop on Rough Sets and Soft Computing Proceedings (RSSC'94). San Jose State University, San Jose, California, USA, November 10–12 (1994); see also: Lin, T. Y., Wildberger A. M., (eds.): Soft Computing: Rough Sets, Fuzzy Logic, Neural Networks, Uncertainty Management, Knowledge Discovery. Simulation Councils, Inc., San Diego, CA (1995)

[Or1] Orłowska, E. (ed.): Incomplete Information: Rough Set Analysis. Physica–Verlag, Heidelberg (1997)

[Pa1] Pawlak, Z.: Rough Sets - Theoretical Aspects of Reasoning about Data. Kluwer Academic Publishers, Boston, London, Dordrecht (1991)

[PaS] Pal, S.K., Skowron, A. (eds.): Rough Fuzzy Hybridization: A New Trend in Decision–Making. Springer-Verlag, Singapore (1999)

[PS1] Polkowski, L., Skowron, A. (eds.): Rough Sets in Knowledge Discovery 1: Methodology and Applications. Physica-Verlag, Heidelberg (1998)

[PS2] Polkowski, L., Skowron, A. (eds.): Rough Sets in Knowledge Discovery 2: Applications, Case Studies and Software Systems. Physica-Verlag, Heidelberg (1998)

[PS3] Polkowski, L., Skowron, A. (eds.): Proceedings of the 1st International Conference on Rough Sets and Current Trends in Computing (RSCTC'98). Warsaw, June 2, 22-26, 1998, Lecture Notes in Artificial Intelligence **1424**, Springer-Verlag, Heidelberg (1998)

[SZ1] Skowron, A., Zhong, N. (eds.): Proceedings of the 7-th International Workshop on Rough Sets, Fuzzy Sets, Data Mining, and Granular-Soft Computing (RSFDGrC'99). Yamaguchi, November 9-11, 1999, Lecture Notes in Artificial Intelligence **1711**, Springer-Verlag, Heidelberg (1999)(in print)

[Sł2] Słowiński, R. (ed.): Intelligent Decision Support. Handbook of Applications and Advances of the Rough Set Theory, Kluwer Academic Publishers, Boston, London, Dordrecht (1992)

[SS1] Słowiński, R., Stefanowski, J. (eds.): Proceedings of the First International Workshop on Rough Sets: State of the Art and Perspectives. Kiekrz – Poznań, Poland, September 2–4 (1992); see also Słowiński, R., Stefanowski, J. (eds.): Foundations of Computing and Decision Sciences **18/3–4** (1993) 155–396 (special issue)

[Ts1] Tsumoto, S., et al. (eds.): The Fourth Internal Workshop on Rough Sets, Fuzzy Sets and Machine Discovery, proceedings. November 6-8, The University of Tokyo (1996)

[Zi1] Ziarko, W. (ed.): Proceedings of the Second International Workshop on Rough Sets and Knowledge Discovery (RSKD'93). Banff, Alberta, Canada, October 12–15 (1993), see also: Ziarko, W., (ed.): Rough Sets, Fuzzy Sets and Knowledge Discovery. Proceedings of the International Workshop on Rough Sets and Knowledge Discovery (RSKD'93), Banff, Alberta, Canada, October 12–15, Workshops in Computing, Springer–Verlag & British Computer Society, London, Berlin (1994)

[Zi2] Ziarko, Z. (ed.): Computational Intelligence: An International Journal **11/2** (1995) (special issue)

[Zi3] Ziarko, Z. (ed.): Fundamenta Informaticae **27/2–3** (1996) (special issue)