

Rough Sets - Basic Notions

by
Zdzislaw Pawlak

ICS Research Report 52/93

Warsaw University of Technology

Warsaw, December 1993

ROUGH SETS - BASIC NOTIONS

Zdzisław Pawlak

1. Introduction

Objectives of this note are to give basic ideas of the rough set theory. More elaborated consideration on rough sets and their applications can be found in Pawlak (1991), Skowron and Rauszer (1991) and Slowinski (1992).

2. Information System

The starting point of our considerations will be the notion of an information system. By an *information system* we will mean an ordered pair $S = (U, A)$, where U is non empty finite set called the *universe*, element of which will be referred to as *objects* - and A - is a nonempty, finite set of elements called *attributes*.

Every attribute $a \in A$ is a total function $a:U \rightarrow V_a$, where V_a - is the set of *values* of a , called the *domain* of a , and $a(x) \in V_a$.

With every subset of attributes $B \subseteq A$, we associate an equivalence relation $IND(B)$, called an indiscernibility relation and defined thus:

$$IND(B) = \{(x,y) \in U^2: \text{for every } a \in B, a(x) = a(y)\}.$$

Obviously we have

$$[x]_{IND(B)} = \bigcap_{a \in B} [x]_a,$$

where $[x]_R$, denotes an equivalence class of the relation R containing an element x .

In fact an information system is a finite table in which rows are labelled by objects, columns - by attributes and entries are values of the corresponding functions. Such tables are very convenient to represent various algorithms related to the rough set theory.

Example of an information system is given in Table 1. We will use this example to illustrate notions considered in this paper.

U	a	b	c	d	e
1	1	0	2	1	0
2	0	0	1	2	1
3	2	0	2	1	0
4	0	0	2	2	2
5	1	2	2	1	0

Table 1

If an object $x \in POS_B(X)$, then x will be called a *B-positive example* of X , and similarly for $NEG_B(X)$ and $BN_B(X)$.

The positive region $POS_B(X)$ or the lower approximation of X is the collection of those objects which can be classified with full certainty as members of the set X , using attributes B .

Similarly, the negative region $NEG_B(X)$ is the collection of objects for which it can be determined without any ambiguity, employing attributes B , that they do not belong to the set X , and, they belong to the complement of X .

The boundary region is in a sense undecidable area of the universe, i.e. none of the objects belonging to the boundary can be classified with certainty into X or \bar{X} employing the set of attributes B .

We will say that set X is *B-definable* iff $\bar{B}X = \underline{B}X$, otherwise the set is *B-undefinable* or *B-rough*.

For example set $X = \{1,3,4\}$ is $\{b,c\}$ -definable but it is $\{a,b\}$ -undefinable.

Properties of approximations can be found in Pawlak (1991).

In order to express numerically how a set can be defined using set of attributes B we will use the *accuracy coefficient* as given below:

$$\alpha_B(X) = \frac{\text{card } \underline{B}X}{\text{card } \bar{B}X},$$

where $X \neq \emptyset$.

Obviously $0 \leq \alpha_B(X) \leq 1$, for every $B \subseteq A$ and $X \subseteq U$; if $\alpha_B(X) = 1$ the B -boundary region of X is empty and the set X is B -definable; if $\alpha_B(X) < 1$ the set X has some non-empty B -boundary region and consequently is B -undefinable.

For example for $B = \{a,b\}$ and $X = \{2,3\}$ we have $\alpha_B(X) = 1/3$.

Besides characterization of rough sets by means of numerical values (accuracy coefficient), one can also define another characterization of rough sets employing the notion of the lower and the upper approximation. It turns out then that there are four important and different kinds of rough sets defined as shown below:

- a) If $\underline{B}X \neq \emptyset$ and $\bar{B}X \neq U$, then we say that X is *roughly B-definable*

- b) If $\underline{BX} = \emptyset$ and $\overline{BX} \neq U$, then we say that X is *internally B-undefinable*
- c) If $\underline{BX} \neq \emptyset$ and $\overline{BX} = U$, then we say that X is *externally B-undefinable*
- d) If $\underline{BX} = \emptyset$ and $\overline{BX} = U$, then we say that X is *totally B-undefinable*

For example for $B = \{a\}$ we have

- $X = \{2,4\}$ is B -definable
 $X = \{1,2,4\}$ is roughly B -definable
 $X = \{2,5\}$ is internally B -undefinable
 $X = \{1,2,3,5\}$ is externally B -undefinable

There is no totally $\{a\}$ -undefinable set in the system.

It is also important to have the notion of approximation of classifications. Let $F = \{X_1, X_2, \dots, X_n\}$, $X_i \in U$, be a classification of U and let $B \subseteq A$. By $\underline{BF} = \{\underline{BX}_1, \underline{BX}_2, \dots, \underline{BX}_n\}$ and $\overline{BF} = \{\overline{BX}_1, \overline{BX}_2, \dots, \overline{BX}_n\}$, we denote the B -lower and the B -upper approximation of the family F .

We will define two measures to describe inexactness of approximate classifications.

The first one is the extension of the coefficient defined to describe accuracy of approximation of sets, defined as follows:

$$\alpha_B(F) = \frac{\sum \text{card } \underline{BX}_i}{\sum \text{card } \overline{BX}_i} .$$

The second coefficient called the quality of approximation of F by R is the following:

$$\gamma_B(F) = \frac{\sum \text{card } \underline{BX}_i}{\text{card } U} .$$

Besides numerical characterisation of roughness (vagueness) of sets we can introduce another numerical coefficient, defined as follows

$$\mu_B^X(X) = \frac{\text{card}(X \cap [x]_B)}{\text{card } U}.$$

Obviously $0 \leq \mu_B^X(X) \leq 1$ and can be consider as a rough membership function (cf. Pawlak and Skowron (1993)) expressing how "strongly" an element x belong to the rough set X in view of information about the element expressed be the set of attributes B .

4. Reduct and Core of Attributes

We will say that attribute $a \in B$ is *superfluous* in B , if $IND(B) = IND(B - \{a\})$; otherwise the attribute a is *indispensable* in B .

If all attributes $a \in B$ are indispensable in B , then B will be called *orthogonal*.

Subset $B' \subseteq B$ is a *reduct* of B , iff B' is orthogonal and $IND(B) = IND(B')$.

The set of all indispensable attributes in B will be called the *core* of B , and will be denoted $CORE(B)$.

Proposition 1

$$CORE(B) = \bigcap_{R \in RED(B)} R,$$

where $RED(B)$ is the family of all reducts of B . ■

To compute easily reducts and the core we will use discernibility matrix (cf. Skowron et al. (1991)), which is defined next.

Let $S = (U, A)$ be an information system with $U = \{x_1, x_2, \dots, x_n\}$, and let $B \subseteq A$. By an *discernibility matrix* of B in S , denoted $M_S(B)$, or $M(B)$ if S is understood - we will mean $n \times n$ matrix defined thus:

$$(c_{ij}) = \{a \in B: a(x_i) \neq a(x_j)\} \text{ for } i, j = 1, 2, \dots, n.$$

Thus entry c_{ij} is the set of all attributes which discern objects x_i and x_j .

The discernibility matrix $M(B)$ assigns to each pair of objects x and y a subset of attributes $\delta(x, y) \subseteq B$, with the following properties:

- i) $\delta(x,x) = \emptyset$
- ii) $\delta(x,y) = \delta(y,x)$
- iii) $\delta(x,z) \subseteq \delta(x,y) \cup \delta(y,z)$

These properties resemble properties of semi-distance, and therefore the function δ may be regarded as *qualitative semi-metric* and $\delta(x,y)$ - *qualitative semi-distance*. Thus the discernibility matrix can be seen as a *semi-distance (qualitative) matrix*.

Let us also note that for every $x,y,z \in U$ we have

- iv) $\text{card } \delta(x,x) = 0$
- v) $\text{card } \delta(x,y) = \text{card } \delta(y,x)$
- vi) $\text{card } \delta(x,z) \leq \text{card } \delta(x,y) + \text{card } \delta(y,z)$

It is easily seen that the core is the set of all single element entries of the discernibility matrix $M(B)$, i.e.

$$\text{CORE}(B) = \{a \in B: c_{ij} = (a), \text{ for some } i,j\}.$$

It can be easily seen that $B' \subseteq B$ is the reduct of B , if B' is the minimal (with respect to inclusion) subset of B such that

$$B' \cap c \neq \emptyset \text{ for any nonempty entry } c (c \neq \emptyset) \text{ in } M(B).$$

In other words reduct is the minimal subset of attributes that discerns all objects discernible by the whole set of attributes.

Every discernibility matrix $M(B)$ defines uniquely a *discernibility (boolean) function* $f(B)$ defined as follows.

Let us assign to each attribute $a \in B$ a binary boolean variable \bar{a} , and let $\sum \delta(x,y)$ denotes boolean sum of all boolean variables assigned to the set of attributes $\delta(x,y)$. Then the discernibility function can be defined by the formula

$$f(B) = \prod_{(x,y) \in U^2} \sum \delta(x,y) .$$

The following theorem establishes the relationship between disjunctive normal form of the function $f(B)$ and the set of all reducts of B .

Proposition 2 (Skowron et al. (1991))

All constituents (prime implicants) in the disjunctive normal form of the function $f(B)$ are all reducts of B . ■

For example the indiscernibility matrix for the system given in Table 1 is as follows:

	1	2	3	4	5
1					
2	a, c, d, e				
3	a	a, c, d, e			
4	a, d, e	c, e	a, d, e		
5	b	a, b, c, d, e	a, b	a, b, d, e	

Table 2

The core of the set of attributes $\{a, b, c, e, d\}$ is the set $\{a, b\}$. The discernibility function for this set is

$$(a+c+d+e)a(a+c+d+e)(a+d+e)(c+e)(a+d+e)b(a+b+c+d+e)(a+b) (a+b+d+e).$$

By employing the absorption law $((x+y)x = x)$ and by "multiplying" all the constituents we get the following disjunctive normal formula

$$ab(c+e) = abc+abe.$$

Thus the set of attributes has two reducts $\{a, b, c\}$ and $\{a, b, e\}$. That means that instead Table 1 we can use either Table 3 or Table 4

U	a	b	c
1	1	0	2
2	0	0	1
3	2	0	2
4	0	0	2
5	1	2	2

Table 3

U	a	b	e
1	1	0	0
2	0	0	1
3	2	0	0
4	0	0	2
5	1	2	0

Table 4

Next we can eliminate attribute values, which are unnecessary to discern objects in the system. To this end we can apply similar procedure as to eliminate superfluous attributes, which is defined next.

We will say that the value of attribute $a \in B$, is *superfluous* for x , if $[x]_{IND(B)} = [x]_{IND(B-\{a\})}$; otherwise the value of attribute a is *indispensable* for x .

If for every attribute $a \in B$ the value of a is indispensable for x , then B will be called *orthogonal* for x .

Subset $B' \subseteq B$ is a *reduct* of B for x , iff B' is orthogonal for x and $[x]_{IND(B)} = [x]_{IND(B')}$.

The set of all indispensable values of attributes in B for x will be called the *core* of B for x , and will be denoted $CORE^X(B)$.

The counterpart of Theorem 1 holds also in this case.

Proposition 1'

$$CORE^X(B) = \bigcap_{R \in RED^X(B)} R$$

where $RED^X(B)$ is the family of all reducts of B for x . ■

In order to compute the core and reducts for x we can also use the discernibility matrix as defined before and the discernibility function, which must be slightly modified now, as shown below:

$$f^X(B) = \prod_{y \in U} \sum \delta(x, y)$$

For example let us consider the information system given in Table 3. For this system we have the following discernibility matrix:

	1	2	3	4	5
1					
2	a, c				
3	a	a, c			
4	a	c	a		
5	b	a, b, c	a, b	a, b	

Table 5

For this system we have the following discernibility functions and their normal forms:

$$f^1(A) = (a+c)ab = ab$$

$$f^2(A) = (a+c)c(a+b+c) = c$$

$$f^3(A) = a(a+c)(a+b) = a$$

$$f^4(A) = ac(a+b) = ac$$

$$f^5(A) = b(a+b+c)(a+b) = b.$$

This means that Table 3 can be simplified as shown below in Table 6.

<u>U</u>	<u>a</u>	<u>b</u>	<u>c</u>
1	1	0	x
2	x	x	1
3	2	x	x
4	0	x	2
5	x	2	x

Table 6

where x denotes "do not care" values of attributes. It means in our approach that concepts represented by x are superfluous in our knowledge since they are included in some others, more general concepts.

For Table 4 there corresponds simplified form shown in Table 7 below

<u>U</u>	<u>a</u>	<u>b</u>	<u>e</u>
1	1	0	x
2	x	x	1
3	2	x	x
4	x	x	2
5	x	2	x

Table 7

Often we might be interested whether properties of objects expressed in terms of attributes C can be expressed in terms of attributes B . In turn this is the same as decision tables analysis. In our approach decision table is an information system in which to subsets of attributes $B, C \subseteq A$, called condition and decision attributes respectively, are distinguished. The problem in terms of decision tables boils down to decision tables simplification. To investigate this problem in details we need the notion of a relative reduct and the relative core of attributes.

Let $B, C \subseteq A$, and let

$$POS_B(C) = \bigcup_{x \in U/IND(C)} \underline{B}x$$

We will say that attribute $a \in B$ is C -superfluous in B , if $POS_B(C) = POS_{(B-\{a\})}(C)$; otherwise the attribute a is C -indispensable in B .

If all attributes $a \in B$ are C -indispensable in B , then B will be called C -orthogonal.

Subset $B' \subseteq B$ is a C -reduct of B , iff B' is C -orthogonal and $POS_B(C) = POS_{B'}(C)$.

The set of all C -indispensable attributes in B will be called the C -core of B , and will be denoted $CORE_C(B)$. The counterpart of the Proposition 1 has the form.

Proposition 1''

$$CORE_C(B) = \bigcap_{R \in RED_C(B)} R,$$

where $RED_C(B)$ is the family of all C -reducts of B . ■

If $B = C$ we will get previous definitions. Relative reducts can be computed also using discernibility matrix which needs slight modification.

Let $S = (U, A)$ be an information system with $U = \{x_1, x_2, \dots, x_n\}$, and let $B, C \subseteq A$. By an C -discernibility matrix of B in S , denoted $M_C(B)$, we mean $n \times n$ matrix defined thus:

$$(c_{ij}) = \{a \in B: a(x_i) \neq a(x_j) \text{ and } (x_i, x_j) \notin IND(C)\}$$

for $i, j = 1, 2, \dots, n$ and x_i or x_j belong to $POS_B(C)$.

Thus entry c_{ij} is the set of all attributes which discern objects x_i and x_j which do not belong to the same equivalence class of the relation $IND(C)$.

The remaining definitions need a little changes as shown below.

The C -core is the set of all single element entries of the discernibility matrix $M_C(B)$, i.e.

$$CORE_C(B) = \{a \in B: c_{ij} = \{a\}, \text{ for some } i, j\}$$

Set $B' \subseteq B$ is the C -reduct of B , if B' is the minimal (with respect to inclusion) subset of B such that

$$B' \cap c \neq \emptyset \text{ for any nonempty entry } c (c \neq \emptyset) \text{ in } M_C(B).$$

Thus C -reduct is the minimal subset of attributes that discerns all equivalence classes of the relation $IND(C)$ discernible by the whole set of attributes.

Every discernibility matrix $M_C(B)$ defines uniquely a *discernibility (boolean) function* $f_C(B)$ which is defined as before and the Proposition 2 has now the form

Proposition 2'

All constituents in the disjunctive normal form of the function $f_C(B)$ are all C -reducts of B . ■

For example consider a decision table with $B = \{a,b,c\}$ and $C = \{d,e\}$ as condition and decision attributes respectively. Discernibility matrix for this table is given below:

	1	2	3	4	5
1					
2	a,c				
3	-	a,c			
4	a	c	a		
5	-	a,b,c	-	a,b	

Table 8

The discernibility function and its disjunctive normal form is the following:

$$ac(a+c)(a+b+c)(a+b) = ac.$$

Thus set $B = \{a,b,c\}$ has only one C -reduct, which is $\{a,c\}$. This means the attribute b is superfluous, i.e. objects of the universe can be equally well classified to classes of the equivalence relation $IND(\{d,e\})$ without attribute b , or what is the same the decision table shown in Table 1 can be simplified as Table 9 below:

U	a	c	d	e
1	1	2	1	0
2	0	1	2	1
3	2	2	1	0
4	0	2	2	2
5	1	2	1	0

Table 9

We can be also interested to drop unnecessary values of condition attributes in this decision table. To this end we must also have a notion of a relative reduct and the relative core of values of attributes, which requires slight modification of previous definition.

Suppose we are given $B, C \subseteq A$, and $x \in U$. We say that value of attribute $a \in B$, is C -superfluous for x , if

$$[x]_{IND(B)} \subseteq [x]_{IND(C)} \text{ implies } [x]_{IND(B-\{a\})} \subseteq [x]_{IND(C)} ;$$

otherwise the value of attribute a is C -indispensable for x .

If for every attribute $a \in B$ value of a is C -indispensable for x , then B will be called C -orthogonal for x .

Subset $B' \subseteq B$ is a C -reduct of B for x , iff B' is C -orthogonal for x and

$$[x]_{IND(B)} \subseteq [x]_{IND(C)} \text{ implies } [x]_{IND(B')} \subseteq [x]_{IND(C)}.$$

The set of all C -indispensable for x values of attributes in B will be called the C -core of B for x , and will be denoted $CORE_C^x(B)$.

The counterpart of Theorem 1 now has the form

Proposition 1''''

$$CORE_C^x(B) = \bigcap_{R \in RED_C^x(B)} R,$$

where $RED_C^x(B)$ is the family of all C -reducts of B for x .

For computing reducts and the core for this case we use as a starting point the discernibility matrix $M_C(B)$ and the discernibility function, will have the form:

$$f_C^x(B) = \prod_{y \in U} \sum \delta(x, y).$$

For Table 9 we get the following discernibility functions and their disjunctive normal forms:

$$f_C^1(B) = (a+c)a = a$$

$$f_C^2(B) = (a+c)c(a+c) = c$$

$$f_C^3(B) = (a+c)a = a$$

$$f_C^4(B) = ac$$

$$f_C^5(B) = a .$$

That means that the decision table shown in Table 9 can be presented in equivalent form as shown in Table 10 below:

<u>U</u>	<u>a</u>	<u>c</u>	<u>d</u>	<u>e</u>
1	1	x	1	0
2	x	1	2	1
3	2	x	1	0
4	0	2	2	2
5	1	x	1	0

Table 10

The above decision table can be also regarded as a set of decision rules of the form

$$a_1 \rightarrow d_1 e_0$$

$$a_2 \rightarrow d_1 e_0$$

$$c_1 \rightarrow d_2 e_1$$

$$a_0 c_2 \rightarrow d_2 e_2$$

or

$$a_1 \vee a_2 \rightarrow d_1 e_0$$

$$c_1 \rightarrow d_2 e_1$$

$$a_0 c_2 \rightarrow d_2 e_2$$

where α_i means "attribute α has value i " and symbols " \vee " and " \rightarrow " denote propositional alternative and implication respectively. In the decision rule $\phi \rightarrow \Psi$ formulas ϕ , Ψ are called *condition* and *decision* respectively. Minimization of set of attributes and values of attributes with respect to some other set of attributes means simply reduction of unnecessary conditions in decision rules, which is also known as decision rule generation from data.

5. Dependency of Attributes

Next important definition concerns dependency of attributes.

Intuitively speaking set of attributes $Q \subseteq A$ depends on set of attributes $P \subseteq A$ ($P \Rightarrow Q$), if values of attributes in Q are uniquely determined by values of attributes in P , i.e. if there exists a function which assigns to each set of values of P set values of Q . Formally

$$P \Rightarrow Q \text{ iff } IND(P) \subseteq IND(Q).$$

Below a property which establishes relation between reducts and dependency is given.

Proposition 3

Let $S = (U, A)$ be an information system and let $B \subseteq A$. If B' is a reduct of B , then $B' \Rightarrow B - B'$. ■

The following property is direct consequence of the definition of dependency.

Proposition 4

$$P \Rightarrow Q, \text{ implies } P \Rightarrow Q', \text{ for every } Q' \subseteq Q. \quad \blacksquare$$

Propositions 3 and 4 enables us to find all dependencies among attributes.

Proposition 5

If B' is a reduct of B , then neither $\{a\} \Rightarrow \{b\}$ nor $\{b\} \Rightarrow \{a\}$ holds, for every $a, b \in B'$, i.e. all attributes in the reduct are pairwise independent. ■

Assume for example that for the system presented in Table 1, $B = \{a, b, c\}$ and $C = \{d, e\}$. It is easily seen that $B \Rightarrow C$, which yields the dependencies $\{a, b, c\} \Rightarrow \{d\}$ and $\{a, b, c\} \Rightarrow \{e\}$. Now it can be seen clearly the role of the relative reduct. It means that instead the above dependencies we can use dependencies $\{a, c\} \Rightarrow \{d, e\}$, $\{a, c\} \Rightarrow \{d\}$ and $\{a, c\} \Rightarrow \{e\}$.

The above definition of the dependency of attributes can be generalized as follows.

Let $B, C \subseteq A$. We say that B depends in a degree k ($0 \leq k \leq 1$) on C , symbolically $B \Rightarrow_k C$, if

$$k = \gamma_B(C) = \frac{\text{card } POS_B(C)}{\text{card } (U)} .$$

If $k = 1$, then we say that C totally depends on B , and the definition coincides with the previous one; if $0 < k < 1$, we say that C partially depends on B , and if $k = 0$ we say that C is totally independent on B .

For example in the system presented in Table 1 we have the following partial dependencies of attributes

$$\{a, b\} \Rightarrow_{0.6} \{c\}$$

$\{b,c\} \Rightarrow_{0.4} \{d\}$

$\{a,c\} \Rightarrow_{0.6} \{b\}$.

Referencies

Pawlak, Z. (1991). *Rough Sets, Theoretical Aspects of Reasoning about Data*, KLUWER ACADEMIC PUBLISHERS

Pawlak, Z and Skowron, A. (1993). *Rough Membership Functions: A Tool for Reasoning with Uncertainty. Algebraic Methodes in Logic and Computer Science, Banach Center Publications, Vol. 28, Polish Academy of Sciences, pp.135 -150.*

Skowron, A. and Rauszer, C. (1991). *The Discernibility Matrices and Functions in Information Systems, in Intelligen Decision Support - Handbook of Applications and Advances of the Rough Set Theory, Slowinski, R. (Ed.).Kluwer Academic Publishers, Dordrecht, Boston, London, pp.331-362.*

Slowinski, R. (1992). (Ed.) *Intelligen Decision Support - Handbook of Applications and Advances of the Rough Set Theory, Slowinski, R. (Ed.).Kluwer Academic Publishers, Dordrecht, Boston, London.*

**Recently published Research Reports
of the Institute of Computer Science, W.U.T.**

- 37/93 Zdzisław Pawlak, On some Issues Connected with Conflict Analysis, September 1993.
- 38/93 Andrzej Skowron, Zbigniew Suraj, Synthesis of Concurrent Systems Specified by Information Systems, September 1993.
- 39/93 Michał Gawryś, Jacek Sienkiewicz, RSL - The Rough Set Library, September 1993.
- 40/93 Mieczysław Muraszkiewicz, Henryk Rybiński, Towards a Parallel Rough Sets Computer, September 1993.
- 41/93 Lech Polkowski, Andrzej Skowron, Analytical Morphology: Mathematical Morphology of Decision Tables, October 1993.
- 42/93 Marzena Kryszkiewicz, Fast Algorithm Finding Minima of Monotonic Boolean Functions, November 1993.
- 43/93 Marcin S. Szczuka, Applications of Rough Sets Methods in Neural Network Processing, November 1993.
- 44/93 Nguyen Hung Son, Nguyen Tuan Trung, An Approach to the Handwriting Digit Recognition Problem Based on Modal Logic, November 1993
- 45/93 Andrzej Skowron, Extracting Laws from Decision Tables: A Rough Set Approach, December 1993.
- 46/93 Andrzej Skowron, Management of Uncertainty in AI: A Rough Set Approach, December 1993.
- 47/93 Mohua Banerjee, M. K. Chakraborty, Rough Algebra, December 1993.
- 48/93 M. K. Chakraborty, Mohua Banerjee, Rough Consequence, December 1993.
- 49/93 M. K. Chakraborty, Mohua Banerjee, Rough Logic with Rough Quantifiers, December 1993.
- 50/93 Solomon Marcus, Imprecision, between Variety and Uniformity: The Conjugate Pairs, December 1993.
- 51/93 Solomon Marcus, Tolerance Rough Sets Čech Topologies, Learning Processes, December 1993.