PRACE IPI PAN • ICS PAS REPORTS

Zdzisław Pawlak

Rough sets

Algebraic and topological approach

482

June 1982

WARSZAWA

INSTYTUT PODSTAW INFORMATYKI POLSKIEJ AKADEMII NAUK
INSTITUTE OF COMPUTER SCIENCE POLISH ACADEMY OF SCIENCES
00-901 WARSAW, P.O. Box 22, POLAND

Zdzisław Pawlak

ROUGH SETS

Algebraic and topological approach

Warsaw, June 1982

Pracę zgłosił Witold Lipski

Mailing address: Zdzisław Pawlak
Institute of Computer Science
Polish Academy of Sciences
00-901 Warsaw, PKiN
P.O. Box 22

ISSN 0138-0648

P r i n t e d   a s   a   m a n u s c r i p t

N a   p r a w a c h   r ę k o p i s u

Nakład 500 egz. Ark. wyd. 1,30; ark. druk. 2,00.
Papier offset. kl. III, 70 g, 70 x 100. Oddano do
druku w czerwcu 1982 r. W.D.N. Zam. nr 285/0/82

Abstract . Содержание . Streszczenie

In this paper the theory of rough sets is developed in an algebraic and topological framework. Applications of the theory of rough sets are described for constructing decision support systems and learning systems.

Приближенные множества. Алгебраическое и топологическое отношение

В работе развивается теория приближенных множеств в алгебраическом и топологическом отношении. Представлены применения конструирования учащихся и решающих систем.

Zbiory przybliżone. Ujęcie algebraiczne i topologiczne

W pracy rozwinięta jest teoria zbiorów przybliżonych w ujęciu algebraicznym i topologicznym. Przedstawione są zastosowania tej teorii w zakresie konstruowania systemów uczących się i systemów decyzyjnych.

Apart from t... unknown, what ...
Harold ...
(in the Homeco...

INTRODUCTION

This paper is an modified extension of [15,16], where ... cept of a rough set is studied.

It seems that the concept of the rough set can be of so... value in some branches of artificial intelligence such as ind... ive reasoning, automatic classification, pattern recognition, learning algorithm etc.

Some results concerning rough sets are given in [4,5,6,7,1... 15,16,19,20].

The idea of a rough set is in a way related to that of a fuzzy set [21], concept of alternative set theory [18] and non-standard analysis [17], however there are essential differences between those concepts, but we shall not discuss that problem here.

Thanks are due to prof. W. Marek and prof. E. Orłowska for their helpful comments and remarks.

1. APPROXIMATION SPACE; APPROXIMATIONS

1.1. Basic notions

Let U be a certain set called the universum, and let R be an equivalence relation on U. The pair $A = (U,R)$ will be called an approximation space. We shall call R an indiscernibility relation. If $x,y \in U$ and $(x,y) \in R$ we say that x and y are indistinguishable in A.

Subsets of U will be denoted by X, Y, Z possibly with indices. The empty set will be denoted by 0, and the universum U

will be also denoted by 1.

Equivalence classes of the relation R will be called <u>elementary sets</u> (<u>atoms</u>) in A or in short elementary sets. The set of all atoms in A will be denoted by U/R.

We assume that the empty set is also elementary in every A.

Every union (finite) of elementary sets in A will be called a <u>composed set in A</u>, or in short a composed set. The family of all composed sets in A will be denoted as $Com(A)$. Certainly $Com(A)$ is a Boolean set algebra, i.e. the family of all composed sets is closed under intersection, union and complement of sets.

Let X be a certain subset of U. The least composed set in A containing X will be called the <u>best upper approximation</u> of X in A, in symbols $\overline{Apr}_A(X)$; the greatest composed set in A contained in X will be called the <u>best lower approximation</u> of X in A, in symbols $\underline{Apr}_A(X)$. If A is known instead of $\overline{Apr}_A(X)$ $(\underline{Apr}_A(X))$ we shall write $\overline{Apr}(X)$ $(\underline{Apr}(X))$.

Since the intersection of arbitrary number of composed sets is again a composed set therefore we find the best upper approximation of a set exists. Similarly best lower approximation exists.

The set $Bnd_A(X) = \overline{Apr}_A(X) - \underline{Apr}_A(X)$ (in short $Bnd(X)$) will be called <u>boundary</u> of X in A.

Sets $\underline{Edg}_A(X) = X - \underline{Apr}_A(X)$ (in short $\underline{Edg}(X)$) and $\overline{Edg}_A(X) = \overline{Apr}_A(X) - X$, (in short $\overline{Edg}(X)$) are referred to as an <u>internal</u> and an <u>external edge</u> of X in A, respectively.

Of course $Bnd_A(X) = \underline{Edg}_A(X) \cup \overline{Edg}_A(X)$.

Fig. 1 depicts the notion of an upper and lower approximation in a two dimension approximation space consisting of a rectangle partitioned into elementary squars.

Let us define the membership functions $\underline{\in}_A$, $\overline{\in}_A$ (called <u>strong</u> and <u>weak</u> membership, respectively), as follows:

$$x \underline{\in}_A X \quad \text{iff} \quad x \in \underline{Apr}_A(X)$$

$$x \overline{\in}_A X \quad \text{iff} \quad x \in \overline{Apr}_A(X)$$

If $x \underline{\in}_A X$ we say "x <u>surely</u> belongs to X in A", and $x \overline{\in}_A$ is to mean "x <u>possibly</u> belongs to X in A". Thus we can interprete approximations as counterparts of necessity and possibility in modal logic.

1.2. Approximation space and topological space

It is easy to check that the approximation space $A = (U,R)$ defines uniquely the topological space $T(A)$ (in short T_A), where $T_A = (U, Com(A))$, and $Com(A)$ is the family of all open sets in T_A, and U/R is a base for T_A. From the definition of approximations (lower and upper) follows that $Com(A)$ is both the set of all open and closed sets in T_A. Thus $\underline{Apr}_A(X)$ and $\overline{Apr}_A(X)$ can be interpreted as an interior and closure of the set X in the topological space T_A, respectively.

If $\underline{Apr}_A(X) = \overline{Apr}_A(X)$ for every $X \subset U$, then $A = (U,R)$ will be called a <u>discrete</u> approximation space.

One can easily check that if A is a discrete approximation space, then all atoms in A are one element sets.

Of course a discrete approximation space A generates discrete topological space T_A.

Let us notice that, unless A is discrete, T_A is not a Hausdorff space.

1.3. Properties of approximations

From the topological interpretation of the approximation operations, follows that for every $X, Y \subset U$, and every approximation space $A = (U, R)$ the following properties are valid:

(A1) $\overline{Apr}_A(X) \supset X \supset \underline{Apr}_A(X)$

(A2) $\underline{Apr}_A(1) = \overline{Apr}_A(1) = 1$

(A3) $\underline{Apr}_A(0) = \overline{Apr}_A(0) = 0$

(A4) $\overline{Apr}_A(\overline{Apr}_A(x)) = \underline{Apr}_A(\overline{Apr}_A(X)) = \overline{Apr}_A(X)$

(A5) $\underline{Apr}_A(\underline{Apr}_A(X)) = \overline{Apr}_A(\underline{Apr}_A(X)) = \underline{Apr}_A(X)$

(A6) $\overline{Apr}_A(X \cup Y) = \overline{Apr}_A(X) \cup \overline{Apr}_A(Y)$

(A7) $\underline{Apr}_A(X \cap Y) = \underline{Apr}_A(X) \cap \underline{Apr}_A(Y)$

(A8) $\overline{Apr}_A(X) = -\underline{Apr}_A(-X)$

(A9) $\underline{Apr}_A(X) = -\overline{Apr}_A(-X)$

where $-X$ is an abbreviation for $U - X$.

Moreover we have

(B1) $\overline{Apr}_A(X \cap Y) \subset \overline{Apr}_A(X) \cap \overline{Apr}_A(Y)$

(B2) $\underline{Apr}_A(X \cup Y) \supseteq \underline{Apr}_A(X) \cup \underline{Apr}_A(Y)$

(B3) $\overline{Apr}_A(X - Y) \supset \overline{Apr}_A(X) - \overline{Apr}_A(Y)$

(B4) $\underline{Apr}_A(X - Y) \subset \underline{Apr}_A(X) - \underline{Apr}_A(Y)$

The following are counterparts of the law $X \cup -X = 1$, for approximations:

(C1) $\overline{Apr}_A(X) \cup \underline{Apr}_A(-X) = 1$

(C2) $\overline{Apr}_A(X) \cup \overline{Apr}_A(-X) = 1$

(C3) $\underline{Apr}_A(X) \cup \overline{Apr}_A(-X) = 1$

(C4) $\underline{Apr}_A(X) \cup \underline{Apr}_A(-X) = -Bnd_A(X)$

The law $X \cap -X = 0$ has the following analogous for approximations:

(D1) $\overline{Apr}_A(X) \cap \underline{Apr}_A(-X) = 0$

(D2) $\overline{Apr}_A(X) \cap \overline{Apr}_A(-X) = Bnd_A(X)$

(D3) $\underline{Apr}_A(X) \cap \underline{Apr}_A(-X) = 0$

(D4) $\underline{Apr}_A(X) \cap \overline{Apr}_A(-X) = 0$

Moreover we have

(E1) If $X \subset Y$, then $\overline{Apr}_A(X) \subset \overline{Apr}_A(Y)$ and $\underline{Apr}_A(X) \subset \underline{Apr}_A(Y)$, for every A.

Let us notice the following equivalence:

$X = \overline{Apr}(X)$ iff $X = \underline{Apr}(X)$ iff X is a composed set in A.

1.4. Refinement of an approximation space

If our discernment ability increases, that is to say we are able to distinguish smaller clusters of elements in the universum, the corresponding approximation space should have smaller atoms. We can formulate this problem precisely in the following way:

Let $A = (U, R)$ and $A' = (U, R')$ be two approximation spaces. If $R' \subset R$, we say that the space A' if finer than the space A or that the space A is coarser than the space A'.

If A' is finer than A, then

$$\underline{Apr}_{A'}(X) \supset \underline{Apr}_A(X)$$

$$\overline{Apr}_{A'}(X) \subset \overline{Apr}_A(X)$$

$$Bnd_{A'}(X) \subset Bnd_A(X)$$

for every $X \subset U$.

We say that an approximation space A is a refinement of another approximation space A', if there exist k $(k > 1)$ such that each atom of A' is union of k atoms of the space A.

Of course if A is refinement of A' then A is finer than A'.

Certainly discrete approximation spaces are the finest possible approximation spaces.

1.5. Sample of a set

Let $A = (U,R)$ be an approximation space and let $X \subset U$.

We say that Y $(Y \subset X)$ is a lower (upper) sample of X in A if $\underline{Apr}_A(Y) = \underline{Apr}_A(X)$ $(\overline{Apr}_A(Y) = \overline{Apr}_A(X))$.

A lower (upper) sample Y of X in A is proper if Y is the minimal sample of X in A. Thus each lower (upper) sample of X in A has some elements in common with every atom of $\underline{Apr}_A(X)$ $(\overline{Apr}_A(X))$; the proper lower (upper) sample of X in A has exactly one element in common with every atom in $\underline{Apr}_A(X)$ $(\overline{Apr}_A(X))$.

If A is a discrete approximation space, then for every $X \subset U$, X is identical with its samples and X is proper sample of X in A.

In fact this is a characteristic property of discrete spaces.

1.6. Accuracy of an approximation

In order to express the "quality" of an approximation we introduce some accuracy measure.

Let $A = (U,R)$ be an approximation space, and let $X \subset U$.

By $\underline{\mu}_A(X)$ $(\overline{\mu}_A(X))$ we denote the number of atoms in $\underline{Apr}_A(X)$ $(\overline{Apr}_A(X))$, and we call $\underline{\mu}_A(X)$ $(\overline{\mu}_A(X))$ the internal (external) measure of X in A.

In what follows we shall consider only sets with finite internal and external measure.

If $\underline{\mu}_A(X) = \overline{\mu}_A(X)$ we say that X is measurable.

Thus the set X is measurable in A if and only if X is a composed set in A.

Let $A = (U,R)$ be an approximation space and let $X \neq 0$.

By the accuracy of approximation of X in A we mean number.

$$\eta_A(X) = \frac{\underline{\mu}_A(X)}{\overline{\mu}_A(X)} ;$$

Obviously $0 \leq \eta_A(X) \leq 1$ for any approximation space $A = (U,R)$ and any $X \subset U$.

For any set X in a discrete approximation space A = $\eta_A(X) = 1$, and this is the greatest possible accuracy.

If $A = (U,R)$ is a refinement of $A' = (U,R')$, then for every $X \subset U$

$$\eta_A(X) \geq \eta_{A'}(X).$$

Certainly if Y', Y'' are lower and upper proper samples of X in A, then

$$\eta_A(X) = \frac{\overline{\mu}_A(Y')}{\overline{\mu}_A(Y'')} = \frac{card(Y')}{card(Y'')}$$

1.7. Examples

In this paragraph we depict introduced previously notions by means of simple examples.

Example 1. Let R^+ be the set of non-negative real numbers, and let S be the indiscernibility relation on R^+ defined by the following partition:

$$\langle 0,1\rangle, \langle 1,2\rangle, \langle 2,3\rangle, \ldots$$

where $(i,i+1\rangle$, $i = 0,1,2,\ldots$ denotes a half-opened interval. The corresponding approximation space will be denoted as $A=(R^+,S)$.

In that approximation space we have for example

$$\underline{Apr}\{1\} = \emptyset$$

$$\overline{Apr}\{1\} = (0,1\rangle$$

$$\underline{Apr}(1\tfrac{1}{2}, 2\tfrac{1}{2}) = \emptyset$$

$$\overline{Apr}(1\tfrac{1}{2}, 2\tfrac{1}{2}) = (1,3\rangle$$

If $N = \{1,2,3,\ldots\}$ is the set of natural numbers, then

$$\underline{Apr}(N) = \emptyset$$

$$\overline{Apr}(N) = R^+$$

Thus the set of natural numbers is a disperse set in A, and N is the proper upper sample of R^+.

Example 2. Let $A = (R^+,S)$ be an approximation space as in the previous example, and let us consider approximations of an open interval $(0,r)$, where $n \leq r \leq n+1$ fof a certain $n \geq 0$. From the definition we have

$$\underline{Apr}(0,r) = \bigcup_{i=0}^{n-1} (i,i+1\rangle = (0,n\rangle, \text{ for } n \geq 1 \text{ and } \emptyset \text{ for } n = 0$$

$$\overline{Apr}(0,r) = \bigcup_{i=0}^{n} (i,i+1\rangle = (0,n+1\rangle .$$

The internal and external measures of $(0,r)$ in A are

$$\underline{\mu}(0,r) = n$$

$$\overline{\mu}(0,r) = n+1,$$

and the accuracy of $(0,r)$ in A is

$$\eta(0,r) = \frac{n}{n+1} .$$

Thus we can interprete the approximation space $A = (R^+,S)$ as a measurement system, where

$$\overline{\mu}_A(i,i+1\rangle = \underline{\mu}_A(i,i+1\rangle = 1, \quad i = 0,1,\ldots$$

is the unit of measurement in A, and $\eta(0,r)$ is the ... of $(0,r)$ in A.

Example 3. Let V be a finite set called a vocabulary and V^* be the set of all finite sequences over V. Any subse... will be called a language over V.

Ler $R \subseteq V^* \times V^*$ be an indiscernibility relation, and $A = (V, R)$ be an approximation space defined by V^* and ...

A language $L \subset V^*$ is recognizable in A if $\underline{Apr}_A(L) = \overline{Apr}_A(L)$.

The family of all recognizable languages in A, denoted Rec(A), is the topology induced by $A = (V^*,R)$ and the ... the topology is V^*/R.

Example 4. Let $S = \langle X,A,V,\rho\rangle$ be an information system [14], where

X - is the set of objects

A - is the set of attributes

$V = \bigcup V_a$, V_a - is the set of values of attribute $a \in A$

$\rho : X \times A \to V$ is an information function, $\rho_x : A \to V$, $x \in X$ is called an information about x in S, where

$$\rho_x(a) = \rho(x,a)$$

for every $x \in X$ and $a \in A$.

We define a binary relation S over X in the following way:

$$x \underset{S}{\sim} y \qquad \text{iff} \qquad \rho_x = \rho_y .$$

Obviously $\tilde{S}$ is an equivalence relation and $A = (X,S)$ will be called the approximation space induced by the information system S.

Any subset $Y \subset X$ is called describable in S iff $\underline{Apr}_A(Y) = \overline{Apr}_A(Y)$. The set of all describable sets in S, denoted as

Des(S), is a zero dimensional topology induced by S on X, and the base of the topology is $X/\tilde{S}$.

2. ROUGH EQUALITY OF SETS

2.1. Basic definitions

Let $A = (U,R)$ be an approximation space and let $X, Y \subset U$.

We say that

(a) The sets X,Y are roughly bottom - equal in A, in symbols $X \underset{A}{\eqsim} Y$, iff

$$\underline{Apr}_A(X) = \underline{Apr}_A(Y).$$

(b) The sets X,Y are roughly top - equal in A, in symbols $X \underset{A}{\simeq} Y$, iff

$$\overline{Apr}_A(X) = \overline{Apr}_A(Y).$$

(c) The sets X,Y are roughly equal in A, in symbols $X \underset{A}{\approx} Y$, iff

$$X \underset{A}{\eqsim} Y \text{ and } X \underset{A}{\simeq} Y.$$

It is easy to check that $\underset{A}{\simeq}$, $\underset{A}{\eqsim}$, $\underset{A}{\approx}$ are equivalence relations on $P(U)$.

In what follows we shall omit the subscript A if the approximation space A is understood - and write $\eqsim$, $\simeq$, $\approx$ instead $\underset{A}{\eqsim}$, $\underset{A}{\simeq}$, $\underset{A}{\approx}$.

2.2. Properties of rough equality

For any approximation space $A = (U,R)$ and any $X, Y \subset U$ the following properties are true:

(A1) If $X \eqsim Y$, then $X \cap Y \eqsim X \eqsim Y$

(A2) If $X \simeq Y$, then $X \cup Y \simeq X \simeq Y$

(A3) If $X \simeq X'$ and $Y \simeq Y'$, then $X \cup Y \simeq X' \cup Y'$

(A4) If $X \eqsim X'$ and $Y \eqsim Y'$, then $X \cap Y \eqsim X' \cap Y'$

(A5) If $X \eqsim Y$, then $X - Y \eqsim 0$

(A6) $X - Y \simeq 0$ iff $X = Y$

(A7) If $X \eqsim Y$, then $-(-X) \eqsim Y$

(A8) If $X \simeq Y$, then $-(-X) \simeq Y$

(A9) If $X \approx Y$, then $-(-X) \approx Y$

(A10) If $X \eqsim Y$, then $X \cup -Y \eqsim 1$

(A11) If $X \simeq Y$, then $X \cup -Y \simeq 1$

(A12) If $X \eqsim Y$, then $X \cap -Y \eqsim 0$

(A13) If $X \simeq Y$, then $X \cap -Y \simeq 0$

Set X will be called roughly dense (r - dense) in $X \underset{A}{\simeq} 1$.

Set X will be called roughly co-dense (r-co-dense) in if $X \underset{A}{\eqsim} 0$.

Set X will be called roughly dispersed (r - dispersed) in A if X is both r-dense and r-co-dense in A.

One can easily show the following properties:

(B1) If $X \subset Y$ and $Y \eqsim 0$, then $X \eqsim 0$

(B2) If $X \subset Y$ and $X \simeq 1$, then $Y \simeq 1$

(B3) If $X \simeq 1$, then $-X \eqsim 0$

(B4) If $X \eqsim 0$, then $-X \simeq 1$

(B5) If X is a r-dispersed set, then so is $-X$.

(B6) $Y \cap X \eqsim 0$ iff $X \eqsim 0$ or $Y \eqsim 0$

(B7) $Y \cup X \simeq 1$ iff $X \simeq 1$ or $Y \simeq 1$

(B8) If X,Y are both r-dense, then $X \simeq Y$

(B9) If X,Y are both r-co-dense then $X \eqsim Y$

(B10) If X, Y are both r-dispersed then $X \approx Y$.

(B11) $\overline{Apr}_A(X)$ is the union of all sets Y such that $X \simeq_A Y$

(B12) $\underline{Apr}_A(X)$ is the intersection of all sets Y, such that $X \eqsim_A Y$.

3. ROUGH INCLUSION OF SETS

3.1. Basic definitions

Let $A = (U,R)$ be an approximation space and let $X, Y \subset U$.

We introduce the following definitions

(a) we say that X is roughly bottom included in Y, in A, in symbols $X \underset{\sim}{\subset}_A Y$, if $\underline{Apr}_A(X) \subset \underline{Apr}_A(Y)$.

(b) We say that X is roughly top included in Y, in A in symbols $X \overset{\sim}{\subset}_A Y$, if $\overline{Apr}_A(X) \subset \overline{Apr}_A(Y)$.

(c) We say that X is roughly included (r-included) in Y, in A in symbols $X \underset{\sim}{\overset{\sim}{\subset}}_A Y$ if $X \underset{\sim}{\subset}_A Y$ and $X \overset{\sim}{\subset}_A Y$.

If A is understood instead of $X \underset{\sim}{\subset}_A Y$, $X \overset{\sim}{\subset}_A Y$, $X \underset{\sim}{\overset{\sim}{\subset}}_Y Y$, we shall write $X \underset{\sim}{\subset} Y$, $X \overset{\sim}{\subset} Y$, $X \underset{\sim}{\overset{\sim}{\subset}} Y$, respectively.

If $X \overset{\sim}{\subset}_A Y$, X is called rough upper-subset of Y in A;

If $X \underset{\sim}{\subset}_A Y$, X is called rough lower-subset of Y in A:

If $X \underset{\sim}{\overset{\sim}{\subset}}_A Y$, X is called rough subset of Y in A.

One can easily check that all rough inclusions $\underset{\sim}{\subset}$, $\overset{\sim}{\subset}$, $\underset{\sim}{\overset{\sim}{\subset}}$, are ordering relations.

3.2. Properties of rough inclusions

It is easy to prove by simple computations that the following properties are true:

(A1) If $X \subset Y$, then $X \underset{\sim}{\subset} Y$, $X \overset{\sim}{\subset} Y$, $X \underset{\sim}{\overset{\sim}{\subset}} Y$

(A2) If $X \underset{\sim}{\subset} Y$ and $Y \underset{\sim}{\subset} X$, then $X \eqsim Y$

(A3) If $X \overset{\sim}{\subset} Y$ and $Y \overset{\sim}{\subset} X$, then $X \simeq Y$

(A4) If $X \underset{\sim}{\overset{\sim}{\subset}} Y$ and $X \underset{\sim}{\overset{\sim}{\supset}} Y$, then $X \approx Y$

(A5) $X \overset{\sim}{\subset} Y$ iff $X \cup Y \simeq Y$

(A6) $X \underset{\sim}{\subset} Y$ iff $X \cap Y \eqsim Y$

(B1) If $X \subset Y$ and $X \eqsim X'$, $Y \eqsim Y'$ then $X' \underset{\sim}{\subset} Y'$

(B2) If $X \subset Y$ and $X \simeq Y'$, $Y \simeq Y'$ then, $X' \overset{\sim}{\subset} Y'$

(B3) If $X \subset Y$, and $X \approx X'$, $Y \approx Y'$ then $X' \underset{\sim}{\overset{\sim}{\subset}} Y'$

(C1) If $X \overset{\sim}{\supset} X'$ and $Y \overset{\sim}{\supset} Y'$, then $X \cup Y \overset{\sim}{\supset} X' \cup Y'$

(C2) If $X \underset{\sim}{\supset} X'$ and $Y \underset{\sim}{\supset} Y'$, then $X \cap Y \underset{\sim}{\supset} X' \cap Y'$

(C3) $X \cap Y \underset{\sim}{\subset} X \overset{\sim}{\subset} X \cup Y$

(C4) If $X \underset{\sim}{\subset} Y$ and $X \eqsim Z$, then $Z \underset{\sim}{\subset} Y$

(C5) If $X \overset{\sim}{\subset} Y$ and $X \simeq Z$, then $Z \overset{\sim}{\subset} Y$

(C6) If $X \underset{\sim}{\overset{\sim}{\subset}} Y$ and $X \approx Z$, then $Z \underset{\sim}{\overset{\sim}{\subset}} Y$.

3.3. Rough powersets

The family of all r-subsets (lower, upper) of X in A will be denoted by $P_A(X)$ $(P_{\underline{A}}(X), P_{\overline{A}}(X))$ and will be called rough (lower, upper) powerset of X in A.

Thus

$$P_{\underline{A}}(X) = \{Y : Y \underset{\sim}{\subset}_A X\}$$

$$P_{\overline{A}}(X) = \{Y : Y \overset{\sim}{\subset}_A X\}$$

$$P_A(X) = \{Y : Y \underset{\sim}{\overset{\sim}{\subset}}_A X\},$$

It is easy to see that

$$P(X) \subset P_{\underline{A}}(X)$$

$$P(X) \subset P_{\overline{A}}(X)$$

$$P(X) \subset P_A(X)$$

for every A.

4. ROUGH SETS

4.1. Basic notions

Let $A = (U,R)$ be an approximation space, and let $\underset{A}{\eqsim}$, $\underset{A}{\simeq}$, $\underset{A}{\approxeq}$, be the induced equivalence relations on $P(U)$.

Every approximation space $A = (U,R)$ defines three following approximation spaces

$$\underline{A}^* = (P(U), \underset{A}{\eqsim})$$

$$\overline{A}^* = (P(U), \underset{A}{\simeq})$$

$$A^* = (P(U), \underset{A}{\approxeq})$$

in which objects are subsets of U and the relations $\underset{A}{\eqsim}$, $\underset{A}{\simeq}$, $\underset{A}{\approxeq}$ are the indiscernibility relations in the corresponding spaces $\underline{A}^*$, $\overline{A}^*$, A^*.

The approximation space $A^*(\underline{A}^*,\overline{A}^*)$ will be called the extension (lower, upper) of A.

Equivalence classes of the relation $\underset{A}{\approxeq}$ ($\underset{A}{\eqsim}$, $\underset{A}{\simeq}$) will be called rough sets (lower, upper).

Thus rough set (lower, upper) is a family of subsets of U, which are equivalent with respect to the indiscernibility relation $\underset{A}{\approxeq}$ ($\underset{A}{\eqsim}$, $\underset{A}{\simeq}$).

Every approximation space $\underline{A}^*$, $\overline{A}^*$, A^* induces a topology $Com(\underline{A}^*)$, $Com(\overline{A}^*)$, $Com(A^*)$ respectively, and consequently - the topological spaces

$$T_{\underline{A}^*} = (P(U), Com(\underline{A}^*))$$

$$T_{\overline{A}^*} = (P(U), Com(\overline{A}^*))$$

$$T_{A^*} = (P(U), Com(A^*)),$$

and $P(U)/\underset{A}{\eqsim}$, $P(U)/\underset{A}{\simeq}$, $P(U)/\underset{A}{\approxeq}$ are the bases for the corresponding topological spaces.

In other words $P(U)/\underset{A}{\eqsim}$, $P(U)/\underset{A}{\simeq}$, $P(U)/\underset{A}{\approxeq}$ are families of equivalence classes of the relations $\underset{A}{\eqsim}$, $\underset{A}{\simeq}$, $\underset{A}{\approxeq}$ respectively,

i.e families of elementary classes in the corresponding [infor]mation spaces $\underline{A}^*$, $\overline{A}^*$, A^*.

If $X,Y \subseteq U$, and $X \underset{A}{\approxeq} Y$ ($X \underset{A}{\eqsim} Y$, $X \underset{A}{\simeq} Y$) we say that Y are close (bottom, top) in A, otherwise sets X,Y are [remo]te in A.

That is to say, that if X,Y are close (bottom, top) in they belong to the same equivalence class in the approximation space A^* ($\underline{A}^*,\overline{A}^*$). Thus sets which are close (bottom, top) are in a sense similar and we are unable to distinguish them our approximation space $A^*(\underline{A}^*,\overline{A}^*)$.

Thus we are able to introduce the upper and lower approximations of a family of subsets of U in $\underline{A}^*$, $\overline{A}^*$, A^*, i.e., if is a certain family of subsets of U we can introduce the following approximations of F in $\underline{A}^*$, $\overline{A}^*$, A^*

$$\underline{Apr}_{\underline{A}^*}(F), \overline{Apr}_{\underline{A}^*}(F)$$

$$\underline{Apr}_{\overline{A}^*}(F), \overline{Apr}_{\overline{A}^*}(F)$$

$$\underline{Apr}_{A^*}(F), \overline{Apr}_{A^*}(F).$$

but we shall not consider this problem here.

There is also another important problem considered in artificial intelligence. Given a family F of subsets of the universum U; the task is to classify members of F, such that sets in the same equivalence class are according to a certain criterium similar.

In our approach we can formulate the problem as follows:

Let $A = (U,R)$ be an approximation space and let $F \subset P(U)$, be a certain family (non-empty) of subsets of the universum U.

By $\underset{A}{\approxeq} \cap F^2$ ($\underset{A}{\eqsim} \cap F^2$, $\underset{A}{\simeq} \cap F^2$) we mean the restriction of the relation $\underset{A}{\approxeq}$ ($\underset{A}{\eqsim}$, $\underset{A}{\simeq}$), to the family F. Then $F/\underset{A}{\approxeq}\cap F^2$ ($F/\underset{A}{\eqsim}\cap F^2$, $F/\underset{A}{\simeq}\cap F^2$) is to mean the family of equivalence classes of the relation $\underset{A}{\approxeq}$ ($\underset{A}{\eqsim}$, $\underset{A}{\simeq}$) restricted to the family F.

That is to mean that each aporoximation space $A = (U,R)$ induces on any family $F \subset P(U)$ three "natural" classifications $F/\underset{\underline{A}}{\approx} \cap F^2$, $F/\underset{\overline{A}}{\approx} \cap F^2$, $F/\underset{A}{\approx} \cap F^2$, denoted as $C_A(F)$, $C_{\underline{A}}(F)$ and $C_{\overline{A}}(F)$, respectively.

Thus in each equivalence class of the classification $C_{\underline{A}}(F)$ all sets have the same lower approximation, in $C_{\overline{A}}(F)$ - the same upper approximation and in $C_A(F)$ - the same both lower and upper approximations.

4.2. Extensions of higher order

In a similar way as before we can extend each of the approximation spaces A^*, $\underline{A}^*$, $\overline{A}^*$ obtaining approximation spaces of higher order.

The k-extension (lower, upper) $k \geq 0$, will be defined inductively:

(i) An approximation space $A = (U,R)$ is of order 0,

(ii) If A^k is the k-extension of A, then k+1-extension (lower, upper) of A are defined as follows:

$$A^{k+1} = (P^{k+1}(U), \approx_{A^k})$$

$$\underline{A}^{k+1} = (P^{k+1}(U), \approx_{\underline{A}^k})$$

$$\overline{A}^{k+1} = (P^{k+1}(U), \approx_{\overline{A}^k}),$$

where $P^k(U)$ is defined as

(a) $P^0(U) = U$,

(b) $P^{k+1}(U) = P(P^k(U))$.

Thus $A^0 = A$, $A^1 = A^*$ etc.

In this way every approximation space $A = (U,R)$ defines uniquely infinite sequence of approximation spaces of higher or-

ders, allowing to cluster sets, families of sets etc., shall not discuss that problem here.

4.3. Ordering of rough sets

Every approximation space $A = (U,R)$ induces three rela tions $\leq_A$, $\lessdot_A$, $\lessapprox_A$ (in short $\leq$, $\lessdot$, $\lessapprox$) on the fa lies $P(U)/\underset{\underline{A}}{\approx}$, $P(U)/\underset{\overline{A}}{\approx}$ $P(U)/\underset{A}{\approx}$, respectively.

Let $A = (U,P)$ be an approximation space, and let $X, Y \subset$ then

(a) $[X]_{\underset{\sim}{=}} \leq [Y]_{\underset{\sim}{=}}$ iff $X \underset{\sim}{\subseteq} Y$

(b) $[X]_{\simeq} \lessdot [Y]_{\simeq}$ iff $X \tilde{\subset} Y$

(c) $[X]_{\approx} \lessapprox [Y]_{\approx}$ iff $X \underset{\sim}{\tilde{\subseteq}} Y$.

One can easily check that $\leq$, $\lessdot$, $\lessapprox$ are ordering relations on $P(U)/\underset{\sim}{=}$, $P(U)/\simeq$, $P(U)/\approx$ respectively. Thus

$$\underline{P}^*(U) = (P(U)/\underset{\sim}{=}, \leq)$$

$$\overline{P}^*(U) = (P(U)/\simeq, \lessdot)$$

$$P^*(U) = (P(U)/\approx, \lessapprox)$$

are partially ordered families of sets.

The smallest element in $\underline{P}^*(U)$ is $[0]_{\underset{\sim}{=}}$, i.e. the family of all co-dense sets in the topological space T_A. The greatest element in $\underline{P}^*(U)$ is $[1]_{\underset{\sim}{=}} = 1$, i.e. the class consisting only of one set, the universum U.

The smallest element in $\overline{P}^*(U)$ is $[0]_{\simeq} = 0$, i.e. the class consisting of only one set, the empty set, the greatest element in $\overline{P}^*(U)$ is $[1]_{\simeq}$, i.e. the family of all dense sets in the topological space T_A.

The smallest element in $P^*(U)$ is $[0]_{\approx} = 0$, and the greatest element in $P^*(U)$ is $[1]_{\approx} = 1$.

Finally let us remark that if an approximation space $A=(U,R)$ has k atoms (elementary sets), then there are 2^k equivalence classes in a family $P(U)/\eqsim_{\underline{A}}$, (and $P(U)/\simeq_{\overline{A}}$), but there are

$$\sum_{i=1}^{k} \binom{k}{i} 2^{k-i} = 3^k$$

equivalence classes in a family $P(U)/\approx_{A}$.

4.4. Algebraic properties

Let $A = (U,R)$ be an approximation space and let $X \subset U$ be a certain subset of the universum U. Rough set in $A^*(\underline{A}^*, \overline{A}^*)$ containing set X will be denoted by $[X]_{\approx_A}$ ($[X]_{\eqsim_{\underline{A}}}$, $[X]_{\simeq_{\overline{A}}}$) or in short $[X]_{\approx}$, ($[X]_{\eqsim}$, $[X]_{\simeq}$) when $A^*(\underline{A}^*, \overline{A}^*)$ is known.

We introduce three operations on rough sets (lower, upper) denoted as $\vee$, $\wedge$, $\sim$, and called union, intersection and complement respectively.

Let $A = (U,R)$ be an approximation space and let $X,Y,Z \subset U$.

For the approximation space $\underline{A}^* = (P(U), \eqsim_{\underline{A}})$ we define

(A1) $[X]_{\eqsim} \vee [Y]_{\eqsim} = \{Z \subset U : \underline{Apr}(Z) = \underline{Apr}(X \cup Y)\}$.

(A2) $[X]_{\eqsim} \wedge [Y]_{\eqsim} = \{Z \subset U : \underline{Apr}(Z) = \underline{Apr}(X \cap Y)\}$.

(A3) $\sim [X]_{\eqsim} = \{Z \subset U : \underline{Apr}(Z) = \overline{Apr}(-X)\}$

Directly from the definition of a rough set one can easily obtain that

(i) $[X]_{\eqsim} \vee [Y]_{\eqsim}$ may not be equal $[X \cup Y]_{\eqsim}$

(ii) $[X]_{\eqsim} \wedge [Y]_{\eqsim} = [X \cap Y]_{\eqsim}$

(iii) $\sim [X]_{\eqsim} = [-X]_{\simeq}$.

For the approximation space $\overline{A}^* = (P(U), \simeq_{\overline{A}})$ we define

(B1) $[X]_{\simeq} \vee [Y]_{\simeq} = \{Z \subset U : \overline{Apr}(Z) = \overline{Apr}(X \cup Y)\}$,

(B2) $[X]_{\simeq} \wedge [Y]_{\simeq} = \{Z \subset U : \overline{Apr}(Z) = \overline{Apr}(X \cap Y)\}$,

(B3) $\sim [X]_{\simeq} = \{Z \subset U : \overline{Apr}(Z) = \underline{Apr}(-X)\}$.

As before, we obtain for this approximation space

(iv) $[X]_{\simeq} \vee [Y]_{\simeq} = [X \cup Y]_{\simeq}$

(v) $[X]_{\simeq} \wedge [Y]_{\simeq}$ may not be equal $[X \cap Y]_{\simeq}$

(vi) $\sim [X]_{\simeq} = [-X]_{\eqsim}$

A family F of subsets of the set U is called an ideal if the following conditions are satisfied:

(a) if $X \in F$ any $Y \subset X$, then $Y \in F$

(b) if $X \in F$ and $Y \in F$, then $X \cup Y \in F$.

A family G of subsets of the set U is called a filter if the following conditions are satisfied:

(c) if $X \in G$ and $X \subset Y$, then $Y \in G$

(d) if $X \in G$ and $Y \in G$, then $X \cap Y \in G$.

Let $A = (U,R)$ be an approximation space, and let $\underline{A}^* = (P(U), \eqsim)$ ($\overline{A}^* = (P(U), \simeq)$) be its lower (upper) extension, and let $P(U)/\eqsim$ ($P(U)/\simeq$) be the family of lower, (upper) rough sets generated by $\underline{A}^*(\overline{A}^*)$.

A family F of rough sets in $\overline{A}^*$ will be called <u>rough ideal</u> in $\overline{A}^*$ if the conditions are satisfied:

(1) if $[X]_{\simeq}$ is in F and $Y \subset X$, then $[Y]_{\simeq}$ is in F

(2) if $[X]_{\simeq}$ and $[Y]_{\simeq}$ are in F, then $[X]_{\simeq} \vee [Y]_{\simeq}$ is in F.

A family G of rough sets in $\underline{A}^*$ will be called <u>rough filter</u> in G if the following conditions are satisfied:

(3) if $[X]_{\approx}$ is in G and $X \subseteq Y$ then $[Y]_{\approx}$ is in G

(4) if $[X]_{\approx}$ and $[Y]_{\approx}$ are in G; then $[X]_{\approx} \wedge [Y]_{\approx}$ is in G.

One can easily show that $P(U)/\approx$ is a rough ideal in $\overline{A}^*$, and $P(U)/\approx$ is a rough filter in $\underline{A}^*$.

Let us remark that

$$[X]_{\approx} \vee [0]_{\approx} = [X]_{\approx}$$
$$[X]_{\approx} \vee [1]_{\approx} = [1]_{\approx}$$
$$[X]_{\approx} \wedge [0]_{\approx} = [0]_{\approx}$$
$$[X]_{\approx} \wedge [1]_{\approx} = [1]_{\approx}$$
$$[X]_{\approx} \vee [X]_{\approx} = [X]_{\approx}$$
$$[X]_{\approx} \wedge [X]_{\approx} = [X]_{\approx}$$
$$\sim (\sim [X]_{\approx}) = [X]_{\approx}$$
$$\sim (\sim [X]_{\approx}) = [X]_{\approx}$$
$$\sim [0]_{\approx} = [1]_{\approx}$$
$$\sim [1]_{\approx} = [0]_{\approx}$$
$$\sim [0]_{\approx} = [1]_{\approx}$$
$$\sim [1]_{\approx} = [0]_{\approx}$$

5. EXAMPLES

5.1. Characteristic symptoms

Let us consider an information system $S = \langle X,A,V,\rho \rangle$ as in example 4, section 1.7, and let us assume that X is a set of patients in a certain hospital, A - is the set of attributes like temperature, blood preasure etc., $V = \bigcup V_a$, $a \in A$ is the set of values of attributes, and the function $\rho_X : A \to V$ describes symptoms of patient X.

Obviously patients belonging to the same equivalence class of the relation S have the same symptoms.

Thus each information system $S = \langle X,A,V,\rho \rangle$ induces an approximation space $A = (X,\tilde{S})$. Suppose we are given the set $Y \subset X$ of patients suffering from a certain disease (the set Y can be given by an expert) and we are interested in finding the characteristic symptoms of that disease.

It follows from the previous considerations that we can give those characteristic symptoms only in that case when Y is a composed set in S, otherwise we can give only symptoms of lower or upper approximation of Y in the approximation space $A = (X,S)$. In other words if Y is not a composed set in S we are not able to give the characteristic symptoms of the set Y, but we can give only the symptoms of patients who surely have the disease Y (symptoms of patients belonging to the lower approximation of Y) or the symptoms of patients who possibly have the disease Y (symptoms of patients belonging to the upper approximation of Y). Let us remark that we identify here the disease with the set of patients having this disease according to the opinion of a certain expert. Another expert can point out of course different set of patients having the considered disease.

Of course, in order to get the characteristic symptoms of the disease Y, we can use not the whole set Y but its lower and upper sample.

5.2. Learning

Suppose we are given an information system as in section 5.1, and suppose that an expert, on the basis of his knowledge, chooses the set $Y \subset X$ of patients having a certain disease. The question arises whether a student can get the knowledge of the expert on the basis of symptoms of the disease Y? In other words - whether the student can define the set Y by means of symptoms of the patients belonging to the set Y.

In general case of course the answer is not, the student can describe the set Y pointed out by an expert in terms of symptoms only if Y is a composed set in S. Otherwise, the student can give only approximate description of the disease Y, i.e. symptoms of lower and upper approximations of Y in S.

We understand that if Y contains patients having a certain disease, then the set $-Y$, does not contain patients having this disease. That is to say that the expert classifies all patients in two classes Y, $-Y$ such that Y contains all patients having a certain disease, and $-Y$ - not having this disease.

Sometimes an expert may be unable to classify patients in two classes, as before, since in some cases he may be unable to classify a patient to the class Y or $-Y$. That is to say sometimes an expert does not know how to classify some objects. In fact in this case he may classify patients in three classes Y^+, Y^-, Y^o, such that Y^+ contains patients who are ill, Y^- - those who are not ill and in Y^o there are patients about which an expert is unable to decide whether they are ill or not.

The question arises how this not complete classification influences the process of learning?

From the previous consideration follows that if the set $Y^o \underset{A}{\approx} 0$, the learning is not affected be the "incomplete" knowledge of an expert, and a student can get exactly the same results as in the case when the expert classification were complete. Otherwise, i.e. if Y^o is not bottom equal to zero, a student is unable to learn (even approximate) classification properly.

That is to say if the "incompleteness" of knowledge of an expert is small enough it does not affect learning, otherwise the learning is affected.

5.3. The case of many experts

Let us consider an information system as in previous sections, and let us suppose, that we employed k experts, to pick up all patients having a certain disease. Thus we obtain a family $F = \{X_1, X_2, \ldots, X_n\}$ of subsets of X such that the set X_i contains all the patients having the considered disease according the opinion of expert i.

The question arises what is the difference between opinions of experts, or in other words how to classify opinions of experts in such a way that "close" opinions are in the same class and "remote" opinions are in different classes.

To do this we can use the three natural classifications $C_{\underline{A}}(F)$, $C_{\overline{A}}(F)$, $C_A(F)$, which have in this case the following meaning:

Each equivalence class of the classification $C_{\underline{A}}(F)$, contains all subsets of the family F having the same lower approximations, i.e. sets which are "close" with respect to symptoms which surely occur among patients in all sets in each equivalence class.

Each equivalence class of the classification $C_{\overline{A}}(F)$ contains all subsets of the family F having the same upper approximations, i.e. sets which are "close" with respect to symptoms which possibly occur among patients in all sets in each equivalence class.

Each equivalence class in the third classification $C_A(F)$ contains sets which have the same sure and possible symptoms.

Thus we can cluster opinions (or experts) into natural similarity classes.

Classification

Again let us consider information system as before and a family $F = \{X_1, X_2, \dots, X_n\}$ of subsets of X. Suppose that F has been given by an export and each X_i represent, according to his knowledge different disease, i.e. all patients having the disease i, belong to the subset X_i.

The question is whether we are able to distinguish all subset of the family F by symptoms? or in other words whether we are able to classify all subset of F in similarity classes, so that in each similarity class we have all subsets of F which are undistinguishable in the approximation space $A = (X,S)$.

To solve this problem we can use the three "natural" classifications $C_{\underline{A}}(F)$, $C_{\overline{A}}(F)$, $C_A(F)$ as in the previous section.

The meaning of the classification $C_{\underline{A}}(F)$ is that in each equivalence class of $C_{\underline{A}}(F)$ we have all subsets of F (or diseases) which we are unable to distinguish by means of symptoms available in our information system, and which surely occur in each disease being in the same equivalence class.

The meaning of the classification $C_{\overline{A}}(F)$ and $C_A(F)$ is obvious.

Thus we can cluster diseases (subsets of the family F) into classes such that in each equivalence class there are diseases, which we are not able to distinguish by means of symptoms available in the information system S.

5.5. Diagnosis

Suppose again that we are given an information system as previously, and the family $F = \{X_1, X_2, \dots, X_k\}$ of subsets of X, determined by an expert such that each X_i contains all patients having a certain disease.

The problem is the following: given a symptom ρ a) what diseases surely have symptom ρ b) what diseases possibly have symptom ρ ?

Let E_ρ denote an equivalence class of the relation $\tilde{S}$, defined by the symptom ρ.

Of course all diseases $X_{i_j} \in F$ such that $\underline{Apr}_A(X_{i_j}) \supset E_\rho$ surely have the symptom ρ, and all diseases $Y_{i_l} \in F$ such that $\overline{Apr}_A(Y_{i_l}) \supset E_\rho$ possibly have the symptom ρ.

If we classify diseases F according to the classifications $C_{\underline{A}}(F)$ and $C_{\overline{A}}(F)$, instead of checking whether lower (upper) approximation of each subset X_i of the family F contains E_ρ, we can simply check whether the corresponding classes contain E_ρ or not - what considerably simplifies the algorithm.

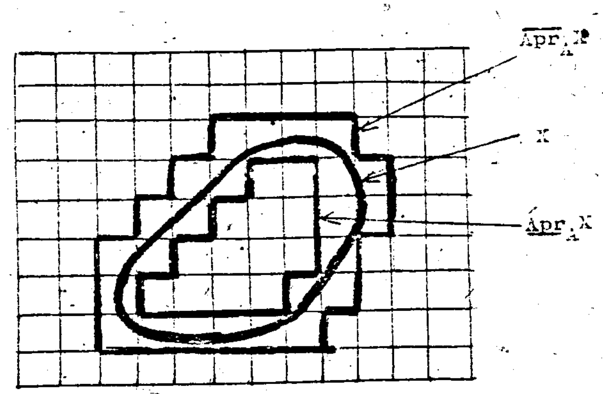

Fig. 1

REFERENCES

[1] Engelking, R., General topology, PWN, 1977

[2] Konrad, E., Orłowska, E., Pawlak, Z., Knowledge representation systems, ICS PAS Reports (1981) No 433

[3] Konrad, E., Orłowska, E., Pawlak, Z., An approximate concept learning, Bericht 81-7 October 81, Berlin

[4] Marek, W., Pawlak, Z., Rough sets and information systems, ICS PAS Reports, (1981) No 441

[5] Orłowska, E., Semantics of vague concepts, Application of rough sets, ICS PAS Reports, (1982) No 469

[6] Orłowska, E., Logic of vague concepts, Application of rough sets, ICS PAS Reports, (1982) No 474

[7] Orłowska, E., Languages of approximate information, ICS PAS Reports (1982), No 479

[8] Orłowska, E., Pawlak, Z., Expressive power of knowledge representation systems, ICS PAS Reports (1981) No 432

[9] Orłowska, E., Pawlak, Z., Measurement and observability. Application of rough sets, ICS PAS Reports (1982) (to appear)

[10] Pawlak, Z., Classification of objects by means of attributes, ICS PAS Reports (1981) No 429

[11] Pawlak, Z., Rough sets, ICS PAS Reports (1981) No 431

[12] Pawlak, Z., Rough relations, ICS PAS Reports (1981) No 435

[13] Pawlak, Z., Rough functions, ICS PAS Reports (1982) No 456

[14] Pawlak, Z., Information systems, theoretical foundations (in Polish) book (to appear)

[15] Pawlak, Z., Some remarks about rough sets, ICS PAS Reports (1982) No 456

[16] Pawlak, Z., Rough sets, powerset hierarchy, ICS PAS Reports (1982) (to appear)

[17] Robinson, A., Non-standard analysis, North-Ho... ing Company, Amsterdam, 1966

[18] Vopenka, P., Mathematics in the Alternative S... Teubner-Texte zur Mathematik, Leipzig, 1979

[19] Żakowski, W., On a concept of rough sets, Demon... Mathematicae (to appear)

[20] Żakowski, W., Approximations in the space (U,... stratio Mathematicae (1982) (to appear)

[21] Zadeh, L.A., Fuzzy sets, Information and Contr... 8, 338-353.

CONTENTS

Introduction 5
1. Approximation space; approximations 5
1.1. Basic notions 5
1.2. Approximation space and topological space . 7
1.3. Properties of approximations 8
1.4. Refinement of an approximation space 9
1.5. Sample of a set 10
1.6. Accuracy of an approximation 10
1.7. Examples 11
2. Rough equality of sets 14
2.1. Basic definitions 14
2.2. Properties of rough equality 14
3. Rough inclusion of sets 16
3.1. Basic definitions 16
3.2. Properties of rough inclusions 16
3.3. Rough powersets 17
4. Rough sets 18
4.1. Basic notions 18
4.2. Extensions of higher order 20
4.3. Ordering of rough sets 21
4.4. Algebraic properties 22
5. Examples 24
5.1. Characteristic symptoms 24
5.2. Learning 25
5.3. The case of many experts 27
5.4. Classification 28
5.5. Diagnosis 28
Bibliography 31