International Journal of Computer and Information Sciences, Vol. 11, No. 5, 1982

Rough Sets

Zdzisław Pawlak[1]

Received June 1981; revised Semptember 1982

We investigate in this paper approximate operations on sets, approximate equality of sets, and approximate inclusion of sets. The presented approach may be considered as an alternative to fuzzy sets theory and tolerance theory. Some applications are outlined.

KEY WORDS: Artificial intelligence; automatic classification; cluster analysis; fuzzy sets; inductive reasoning; learning algorithms; measurement theory; pattern recognition; tolerance theory.

Apart from the known and the unknown, what else is there?
Harold Pinter (The Homecoming)

1. INTRODUCTION

The aim of this paper is to describe some properties of rough sets, introduced in Ref. 7 and investigated in Refs. 1, 2, 4, 5, 6, 8, 9, and 11.

The rough set concept can be of some importance, primarily in some branches of artificial intelligence, such as inductive reasoning, automatic classification, pattern recognition, learning algorithms, etc.

The idea of a rough set could be placed in a more general setting, leading to a fruitful further research and applications in classification theory, cluster analysis, measurement theory, taxonomy, etc.

The key to the presented approach is provided by the exact mathematical formulation of the concept of approximative (rough) equality of sets in a given approximation space; an approximation space is understood as a pair (U, R), where U is a certain set called universe, and $R \subset U \times U$ is an indiscernibility relation. We assume throughout this paper that R is an equivalence relation.

[1] Institute of Computer Sciences, Polish Academy of Sciences, P.O. Box 22, 00-901 Warsaw, PKiN.

0091-7036/82/1000-0341$03.00/0

Some ideas underlying the theory outlined here are common to fuzzy set theory,[13] tolerance theory,[14] nonstandard analysis.[12] However, we are primarily aiming at laying mathematical foundations for artificial intelligence, and not a new set theory or analysis.

Some applications of the presented ideas are given in Refs. 1, 4, 5, 6.

The ideas given in this paper have been inspired by the results of Michalski (see Ref. 3) concerning automatic classification.

We use throughout this paper standard mathematical notations, and we assume that the reader is familiar with basic set theoretical and topological notions.

Thanks are due to Prof. E. Orłowska and Prof. W. Marek for fruitful discussions, and to the reviewer for valuable comments and remarks.

2. APPROXIMATION SPACE; APPROXIMATIONS

2.1. Basic Notions

Let U be a certain set called the *universe*, 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 universe U will also be denoted by 1.

Equivalence classes of the relation R will be called *elementary sets* (*atoms*) in A or, briefly, 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 finite union of elementary sets in A will be called a *composed set in* A, or in short, a composed set. The family of all composed sets in A will be denoted as $\mathrm{Com}(A)$. Obviously $\mathrm{Com}(A)$ is a Boolean algebra, i.e., the family of all composed set 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 *best upper approximation* of X in A, in symbols $\overline{\mathrm{Apr}}_A(X)$; the greatest composed set in A contained in X will be called the *best lower approximation* of X in A, in symbols $\underline{\mathrm{Apr}}_A(x)$.

If A is known, instead of $\overline{\mathrm{Apr}}_A(X)$ ($\underline{\mathrm{Apr}}_A(X)$) we shall write $\overline{\mathrm{Apr}}(X)$ ($\underline{\mathrm{Apr}}(X)$).

The set $\mathrm{Bnd}_A(X) = \overline{\mathrm{Apr}}_A(X) - \underline{\mathrm{Apr}}_A(X)$ (in short $\mathrm{Bnd}(X)$) will be called the *boundary* of X in A.

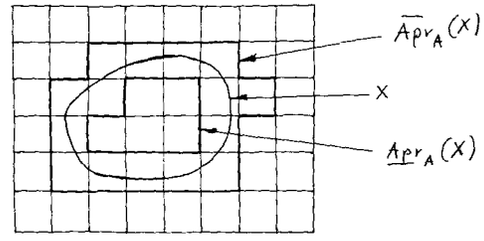

Fig. 1

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

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

Fig. 1 shows the notion of an upper and lower approximation in a two-dimensional approximation space consisting of a rectangle partitioned into elementary squares.

Let us define two membership functions $\underline{\in}_A$, $\overline{\in}_A$ (called *strong* and *weak* membership, respectively), as follows:

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

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

Of course,

$$\underline{\mathrm{Apr}}_A(X) = \{x : x \underline{\in}_A X\}$$
$$\overline{\mathrm{Apr}}_A(X) = \{x : x \overline{\in}_A X\}$$

Thus we can develop our theory in terms of strong and weak membership functions or in terms of approximations. For the sake of simplicity we shall use here the approximational approach.

2.2. 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 $T_A = (U, \mathrm{Com}(A))$, and $\mathrm{Com}(A)$ are the family of all open sets in T_A, and U/R is a base for T_A.

It follows from the definition of (lower and upper) approximations that $\mathrm{Com}(A)$ is both the set of all open and closed sets in T_A. Thus, $\underline{\mathrm{Apr}}_A(X)$ and

$\overline{\text{Apr}}_A(X)$ can be interpreted as an interior and closure of the set X in the topological space T_A, respectively.

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

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

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

2.3. Properties of Approximations

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

$$\overline{\text{Apr}}(X) \supset X \supset \underline{\text{Apr}}(X) \tag{1}$$

$$\underline{\text{Apr}}(1) = \overline{\text{Apr}}(1) = 1 \tag{2}$$

$$\underline{\text{Apr}}(0) = \overline{\text{Apr}}(0) = 0 \tag{3}$$

$$\overline{\text{Apr}}(\overline{\text{Apr}}(X)) = \underline{\text{Apr}}(\overline{\text{Apr}}(X)) = \overline{\text{Apr}}(X) \tag{4}$$

$$\underline{\text{Apr}}(\underline{\text{Apr}}(X)) = \overline{\text{Apr}}(\underline{\text{Apr}}(X)) = \underline{\text{Apr}}(X) \tag{5}$$

$$\overline{\text{Apr}}(X \cup Y) = \overline{\text{Apr}}(X) \cup \overline{\text{Apr}}(Y) \tag{6}$$

$$\underline{\text{Apr}}(X \cap Y) = \underline{\text{Apr}}(X) \cap \underline{\text{Apr}}(Y) \tag{7}$$

$$\overline{\text{Apr}}(X) = -\underline{\text{Apr}}(-X) \tag{8}$$

$$\underline{\text{Apr}}(X) = -\overline{\text{Apr}}(-X) \tag{9}$$

where $-X$ is an abbreviation for $U - X$. Moreover we have

$$\overline{\text{Apr}}(X \cap Y) \subset \overline{\text{Apr}}(X) \cap \overline{\text{Apr}}(Y) \tag{10}$$

$$\underline{\text{Apr}}(X \cup Y) \supset \underline{\text{Apr}}(X) \cup \underline{\text{Apr}}(Y) \tag{11}$$

$$\overline{\text{Apr}}(X - Y) \supset \overline{\text{Apr}}(X) - \overline{\text{Apr}}(Y) \tag{12}$$

$$\underline{\text{Apr}}(X - Y) \subset \underline{\text{Apr}}(X) - \underline{\text{Apr}}(Y) \tag{13}$$

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

$$\overline{\text{Apr}}(X) \cup \underline{\text{Apr}}(-X) = 1 \tag{14}$$

$$\overline{\text{Apr}}(X) \cup \overline{\text{Apr}}(-X) = 1 \tag{15}$$

$$\underline{\text{Apr}}(X) \cup \overline{\text{Apr}}(-X) = 1 \tag{16}$$

$$\underline{\text{Apr}}(X) \cup \underline{\text{Apr}}(-X) = -\text{Bnd}(X) \tag{17}$$

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

$$\overline{\text{Apr}}(X) \cap \underline{\text{Apr}}(-X) = 0 \tag{18}$$

$$\overline{\text{Apr}}(X) \cap \overline{\text{Apr}}(-X) = \text{Bnd}(X) \tag{19}$$

$$\underline{\text{Apr}}(X) \cap \underline{\text{Apr}}(-X) = 0 \tag{20}$$

$$\underline{\text{Apr}}(X) \cap \overline{\text{Apr}}(-X) = 0 \tag{21}$$

De Morgan's laws have the following counterparts:

$$-(\underline{\text{Apr}}(X) \cup \underline{\text{Apr}}(Y)) = \overline{\text{Apr}}(-X) \cap \overline{\text{Apr}}(-Y) \tag{22}$$

$$-(\underline{\text{Apr}}(X) \cup \overline{\text{Apr}}(Y)) = \overline{\text{Apr}}(X) \cap \underline{\text{Apr}}(Y) \tag{23}$$

$$-(\overline{\text{Apr}}(X) \cup \underline{\text{Apr}}(Y)) = \underline{\text{Apr}}(-X) \cap \overline{\text{Apr}}(-Y) \tag{24}$$

$$-(\overline{\text{Apr}}(X) \cup \overline{\text{Apr}}(Y)) = \underline{\text{Apr}}(-X) \cap \underline{\text{Apr}}(-Y) \tag{25}$$

$$-(\underline{\text{Apr}}(X) \cap \underline{\text{Apr}}(Y)) = \overline{\text{Apr}}(-X) \cup \overline{\text{Apr}}(-Y) \tag{26}$$

$$-(\underline{\text{Apr}}(X) \cap \overline{\text{Apr}}(Y)) = \overline{\text{Apr}}(-X) \cup \underline{\text{Apr}}(-Y) \tag{27}$$

$$-(\overline{\text{Apr}}(X) \cap \underline{\text{Apr}}(Y)) = \underline{\text{Apr}}(-X) \cup \overline{\text{Apr}}(-Y) \tag{28}$$

$$-(\overline{\text{Apr}}(X) \cap \overline{\text{Apr}}(X)) = \underline{\text{Apr}}(-X) \cup \underline{\text{Apr}}(-Y) \tag{29}$$

Moreover we have

$$\text{If } X \subset Y, \text{ then } \overline{\text{Apr}}(X) \subset \overline{\text{Apr}}(Y) \text{ and } \underline{\text{Apr}}(X) \subset \underline{\text{Apr}}(Y) \tag{30}$$

Note that $X = \overline{\text{Apr}}_A(X)$ and $X = \underline{\text{Apr}}_A(X)$ iff X is a composed set in A.

2.4. 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 $\underline{\text{Apr}}_A(X)$ ($\overline{\text{Apr}}_A(X)$), and we call $\underline{\mu}_A(X)$ ($\overline{\mu}_A(X)$) the *internal* (*external*) *measure* of X in A.

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

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 $X \subset U$.

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

$$\eta_A(X) = \frac{\underline{\mu}_A(X)}{\bar{\mu}_A(X)}, \qquad \text{where} \quad \bar{\mu}_A(X) \neq 0$$

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

For any measurable set X in A, $\eta_A(X) = 1$. If X is not measurable in A, then $0 \leqslant \eta_A(X) < 1$. In particular $\eta_A(X) = 0$, iff $\underline{\text{Apr}}_A(X) = 0$.

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

2.5. Examples

In this paragraph we illustrate the notions introduced previously with simple examples.

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

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

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

Let us consider approximations of an open interval $(0, r)$, where $n \leqslant r \leqslant n+1$ for a certain $n \geqslant 0$.

By definition we have

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

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

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

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

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

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

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

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

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

is the *unit of measurement* in A, and $\eta(0, r)$ is the accuracy of $(0, r)$ in A. For more detail see Ref. 6.

Example 2. Let V be a finite set called a *vocabulary* and let V^* be the set of all finite sequences over V. Any subset of V^* will be called a *language* over V.

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

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

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

That is to say that if the language L is not recognizable in A we are able to recognize only the lower and upper approximations in A.

This property can be used in speach recognition, pattern recognition, fault tolerant computers, etc.

Example 3. Let $S = \langle X, A, V, \rho\rangle$ be an information system (see Ref. 10), 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 $a \in A$

ρ: $X \times A \to V$ is an *information function*, ρ_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 the binary relation $\tilde{S}$ over X in the following way:

$$x \sim_S y \quad \text{iff} \quad \rho_x = \rho_y$$

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

Any subset $Y \subset X$ is called *describable* in S iff $\underline{\text{Apr}}_A(Y) = \overline{\text{Apr}}_A(Y)$. The set of all describable sets in S, denoted as $\text{Des}(S)$, is a topology induced by S on X, and the base of the topology is $X/\tilde{S}$.

That is to mean that if we classify some objects according to some attributes, in a general case we are unable to define an arbitrary subset of objects by these attributes; only those subsets which are describable in S, can be defined by means of the attributes of the system S.

This property must be taken into consideration, in any classification system in which objects are classified by means of attributes.

3. ROUGH EQUALITY OF SETS

3.1. 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 $X \underset{A}{\approx} Y$, iff $\underline{\mathrm{Apr}}_A(X) = \underline{\mathrm{Apr}}_A(Y)$.

(b) The sets X, Y are *roughly top-equal* in A, in symbols $X \overset{A}{\approx} Y$, iff $\overline{\mathrm{Apr}}_A(X) = \overline{\mathrm{Apr}}_A(Y)$.

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

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

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

3.2. Properties of Rough Equality

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

$$\text{If } X \approx Y, \text{ then } X \cap Y \approx X \approx Y \tag{31}$$

$$\text{If } X \simeq Y, \text{ then } X \cup Y \simeq X \simeq Y \tag{32}$$

$$\text{If } X \simeq X' \text{ and } Y \simeq Y', \text{ then } X \cup Y \simeq X' \cup Y' \tag{33}$$

$$\text{If } X \approx X' \text{ and } Y \approx Y', \text{ then } X \cap Y \approx X' \cap Y' \tag{34}$$

$$\text{If } X \approx Y, \text{ then } X - Y \approx 0 \tag{35}$$

$$X - Y \simeq 0 \text{ iff } X = Y \tag{36}$$

$$\text{If } X \approx Y, \text{ then } -(-X) \approx Y \tag{37}$$

$$\text{If } X \simeq Y, \text{ then } -(-X) \simeq Y \tag{38}$$

$$\text{If } X \approx Y, \text{ then } -(-X) \approx Y \tag{39}$$

$$\text{If } X \simeq Y, \text{ then } X \cup -Y \simeq 1 \tag{40}$$

$$\text{If } X \simeq Y, \text{ then } X \cap -Y \simeq 0 \tag{41}$$

Set X will be called *dense* in A if $X \underset{A}{\simeq} 1$. Set X will be called *codense* in A if $X \underset{A}{\approx} 0$. Set X will be called *dispersed* in A if X is both dense and codense in A.

One can easily show the following properties:

$$\text{If } X \subset Y \text{ and } Y \approx 0, \text{ then } X \approx 0 \tag{42}$$

$$\text{If } X \subset Y \text{ and } X \simeq 1, \text{ then } Y \simeq 1 \tag{43}$$

$$\text{If } X \simeq 1, \text{ then } -X \approx 0 \tag{44}$$

$$\text{If } X \approx 0, \text{ then } -X \simeq 1 \tag{45}$$

$$\text{If } X \text{ is a dispersed set, then so is } -X, \text{ i.e., } X \approx -X \tag{46}$$

and $X \simeq -X$, and hence $X \approx -X$.

$$\text{If } X, Y \text{ are both dense, then } X \simeq Y \tag{47}$$

$$\text{If } X, Y \text{ are both codense then } X \approx Y \tag{48}$$

$$\text{If } X, Y \text{ are both dispersed then } X \approx Y \tag{49}$$

$$X \approx 0 \text{ iff } \underline{\mathrm{Apr}}(X) = 0 \tag{50}$$

$$X \simeq 0 \text{ iff } X = 0 \tag{51}$$

$$X \approx 1 \text{ iff } X = 1 \tag{52}$$

$$X \simeq 1 \text{ iff } \overline{\mathrm{Apr}}(X) = 1$$

$$\overline{\mathrm{Apr}}_A(X) \text{ is the union of all sets } Y \text{ such that } X \simeq_A Y \tag{54}$$

$$\underline{\mathrm{Apr}}_A(X) \text{ is the intersection of all sets } Y, \text{ such that } X \underset{A}{\approx} Y \tag{55}$$

4. ROUGH INCLUSION OF SETS

4.1. 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 $C \subseteq_A Y$, if $\underline{\mathrm{Apr}}_A(X) \subset \underline{\mathrm{Apr}}_A(Y)$.

(b) We say that X is *roughly top-included* in Y, in A in symbols $X \tilde{\subseteq}_A Y$, if $\overline{\mathrm{Apr}}_A(X) \subset \overline{\mathrm{Apr}}_A(Y)$.

(c) We say that X is *roughly included* in Y, in A, in symbols $X \tilde{\tilde{\subseteq}}_A Y$, if $X \tilde{\subseteq}_A Y$ and $X \tilde{\subseteq}_A Y$.

If A is understood then instead of $X \tilde{\subseteq}_A Y$, $X \tilde{\subseteq}_A Y$, and $X \tilde{\tilde{\subseteq}}_A Y$, we shall write $X \tilde{\subseteq} Y$, $X \tilde{\subseteq} Y$, $X \tilde{\tilde{\subseteq}} Y$, respectively. If $X \tilde{\subseteq}_A Y$, X is called a *rough upper-subset* of Y in A; If $X \tilde{\subseteq}_A Y$, X is called a *rough lower-subset* of Y in A; If $X \tilde{\tilde{\subseteq}}_A Y$, X is called a *rough subset* of Y in A.

One can easily check that all rough inclusions $\tilde{\subseteq}$, $\tilde{\subseteq}$, and $\tilde{\tilde{\subseteq}}$ are ordering relations.

The family of all rough (lower, upper) subsets 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,

$$P_{\underline{A}}(X) = \{Y : U \tilde{\subseteq}_A X\}$$
$$P_{\overline{A}}(X) = \{Y : Y \tilde{\subseteq}_A X\}$$
$$P_A(X) = \{Y : Y \tilde{\tilde{\subseteq}}_A X\}$$

It is easy to see that

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

and

$$\text{If } X \approx Y, \text{ then } P_{\overline{A}}(X) = P_{\underline{A}}(Y)$$
$$\text{If } X \simeq Y, \text{ then } P_{\overline{A}}(X) = P_{\overline{A}}(Y)$$
$$\text{If } X \approx Y, \text{ then } P_A(X) = P_A(Y)$$
$$\text{If } X \tilde{\subseteq} Y \text{ then } P_{\underline{A}}(X) \subset P_{\underline{A}}(Y)$$
$$\text{If } X \tilde{\subseteq} Y, \text{ then } P_{\overline{A}}(X) \subset P_{\overline{A}}(Y)$$
$$\text{If } X \tilde{\tilde{\subseteq}} Y, \text{ then } P_A(X) \subset P_A(Y)$$

4.2. Properties of Rough Inclusions

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

$$\text{If } X \subset Y, \text{ then } X \tilde{\subseteq} Y, X \tilde{\subseteq} Y, X \tilde{\tilde{\subseteq}} Y \tag{56}$$

$$\text{If } X \tilde{\subseteq} Y, \text{ and } Y \tilde{\subseteq} X, \text{ then } X \approx Y \tag{57}$$
$$\text{If } X \tilde{\subseteq} Y \text{ and } Y \tilde{\subseteq} X, \text{ then } X \simeq Y \tag{58}$$
$$\text{If } X \tilde{\tilde{\subseteq}} Y \text{ and } X \tilde{\tilde{\subseteq}} Y, \text{ then } X \approx Y \tag{59}$$
$$X \tilde{\subseteq} Y \text{ iff } X \cup Y \simeq Y \tag{60}$$
$$X \tilde{\subseteq} Y \text{ iff } X \cap Y \approx X \tag{61}$$
$$\text{If } X \subset Y \text{ and } X \approx X', Y \approx Y' \text{ then } X' \tilde{\subseteq} Y' \tag{62}$$
$$\text{If } X \subset Y \text{ and } X \simeq X', Y \simeq Y', \text{ then } X' \tilde{\subseteq} Y' \tag{63}$$
$$\text{If } X \subset Y, \text{ and } X \approx X', Y \approx Y', \text{ then } X' \tilde{\tilde{\subseteq}} Y' \tag{64}$$
$$\text{If } X \tilde{\supseteq} X' \text{ and } Y \tilde{\supseteq} Y', \text{ then } X \cup Y \tilde{\supseteq} X' \cup Y' \tag{65}$$
$$\text{If } X \tilde{\supseteq} X' \text{ and } Y \tilde{\supseteq} Y', \text{ then } X \cap Y \tilde{\supseteq} X' \cap Y' \tag{66}$$
$$X \cap Y \tilde{\subseteq} X \tilde{\subseteq} X \cup Y \tag{67}$$
$$\text{If } X \tilde{\subseteq} Y \text{ and } X \approx Z (Y \approx Z) \text{ then } Z \tilde{\subseteq} Y (X \tilde{\subseteq} Z) \tag{68}$$
$$\text{If } X \tilde{\subseteq} Y \text{ and } X \simeq Z (Y \simeq Z), \text{ then } Z \tilde{\subseteq} Y (X \tilde{\subseteq} Z) \tag{69}$$
$$\text{If } X \tilde{\tilde{\subseteq}} Y \text{ and } X \approx Z (Y \approx Z), \text{ then } Z \tilde{\tilde{\subseteq}} Y (X \tilde{\tilde{\subseteq}} Z) \tag{70}$$

5. ROUGH SETS

5.1. Basic Notions

Let $A = (U, R)$ be an approximation space, and let $\approx_A$, $\simeq_A$, $\approx_A$, be equivalence relations on $P(U)$.

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

$$\underline{A}^* = (P(U), \approx_A)$$
$$\overline{A}^* = (P(U), \simeq_A)$$
$$A^* = (P(U), \approx_A)$$

in which objects are subsets of U and the relations $\approx_A$, $\simeq_A$, and $\approx_A$ 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 (*lower*, *upper*) extension of A.

Equivalence classes of the relation $\approx_A (\approx_A, \simeq_A)$ will be called *rough* (*lower*, *upper*) sets.

Thus, a rough (lower, upper) set is a family of subsets of U, which are equivalence with respect to the indiscernibility relation $\approx_A(\underline{\approx}_A, \overline{\approx}_A)$.

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

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

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

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

and $P(U)/\approx_A$, $P(U)/\underline{\approx}_A$, and $P(U)/\overline{\approx}_A$ are the bases for the corresponding topological spaces.

In other words, $P(U)/\approx_A$, $P(U)/\underline{\approx}_A$, $P(U)/\overline{\approx}_A$, are families of equivalence classes of the relations $\approx_A$, $\underline{\approx}_A$, $\overline{\approx}_A$, respectively, i.e., families of elementary classes in the corresponding approximation spaces $\underline{A}^*$, $\overline{A}^*$, A^*. Thus, sets which are in the same equivalence class of an approximation space $A^*(\underline{A}^*, \overline{A}^*)$ are in sense similar and we are unable to distinguish them in the approximation space $A^*(\underline{A}^*, \overline{A}^*)$.[2]

5.2. Rough Sets and Classifications

In artificial intelligence the following problem is of great importance: given a family F of subsets of a certain universe U, the task is to classify members of F, so that the sets in the same equivalence class are similar according to a certain criterion.

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 (nonempty) family of subsets of the universe U.

By $\approx_A \cap F^2(\underline{\approx}_A \cap F^2, \overline{\approx}_A \cap F^2)$ we mean the restriction of the relation $\approx_A(\underline{\approx}_A, \overline{\approx}_A)$ to the family F. Then, $F/\approx_A \cap F^2(F/\underline{\approx}_A \cap F^2, F/\overline{\approx}_A \cap F^2)$ is to mean the family of equivalence classes of $\approx_A(\underline{\approx}_A, \overline{\approx}_A)$ restricted to F. That is to say that each approximation space $A = (U, R)$ induces on the family $F \subset P(U)$ three "natural" classifications $F/\approx_A \cap F^2$, $F/\underline{\approx}_A \cap F^2$, $F/\overline{\approx}_A \cap F^2$, denoted by $C_A(F)$, $C_{\underline{A}}(F)$, and $C_{\overline{A}}(F)$, respectively.

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

We can consider the suggested approach to clustering as an alternative, to cluster analysis based on distance, or similarity functions—in which the indiscernibility relation plays the role of the distance (or similarity) function.

[2] We can also introduce approximation spaces of higher orders but we shall not consider that problem here.

6. EXAMPLES

6.1. Characteristic Symptoms

Let us consider an information system $S = \langle X, A, V, \rho \rangle$ as in example 3, section 2.5, 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 the symptoms of patient x.

Obviously, patients belonging to the same equivalence class of 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 indicated 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 if Y is a composed set in S, otherwise we can give only symptoms of lower or upper approximation of Y in the approximation space $A = (X, \tilde{S})$. In other words, if Y is not a composed set in S we are not able to give the characteristic symptoms of Y, but we can give only the symptoms of patients who surely suffer from the Y (symptoms of patients belonging to the lower approximation of Y) or the symptoms of patients who possibly suffer from the Y (symptoms of patients belonging to the upper approximation of Y). Note that we identify here the disease with the set of patients suffering from this disease according to the opinion of a certain expert. Another expert can, of course indicate a different set of patients having the disease in question.

6.2. Learning

Suppose we are given an information system as in section 6.1, and suppose that an expert, on the basis of his knowledge, chooses the set $Y \subset X$ of patients suffering from a certain disease. The question arises whether a student can obtain the knowledge of that 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 the general case the answer is, of course, in the negative; 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 only give an 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 suffering from a certain

disease, then the set $-Y$, does not contain patients suffering from this disease. This is to say that the expert classifies all patients into two classes, Y and $-Y$, such that Y contains all patients suffering from a certain disease, and $-Y$, those not suffering from that disease.

Sometimes an expert may be unable to classify patients in two classes, as before, since in some cases he may be unable to include a patient in Y or $-Y$. That is to say that sometimes an expert does not know how to classify some objects. In fact in this case he may classify patients into three classes, Y^+, Y^-, Y^0, such that Y^+ contains patients who are ill, Y^- are those who are not ill and in Y^0 there are patients about which an expert is unable to decide whether they are ill or not.

The question arises how this incomplete classification influences the process of learning?

It follows from the previous consideration that if $Y^0 \approx_A 0$, the process of learning is not affected be the incomplete knowledge of an expert, and a student can obtain exactly the same results as when the expert classification is complete. Otherwise, i.e., if Y^0 is not bottom equal to zero, a sutdent is unable to properly learn (even approximately) the classification.

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

6.3. The Case of Many Experts

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

The question arises what is the difference between opinions of experts, or, in other words, how to classify opinions of experts so that similar opinions are in the same class and widely different opinions are in different classes.

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

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

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

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

We can thus cluster opinions (or experts) into natural similarity classes.

6.4. Classification

Let us again consider an information system as before and a family $F = \{X_1, X_2, ..., X_n\}$ of subsets of X. Suppose that F has been given by an expert and each X_i represents, according to his knowledge, a different disease, so that all the patients suffering from the disease i, belong to the subset X_i.

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

To solve this problem we can use the three natural classifications $C_{\underline{A}}(F)$, $C_{\overline{A}}(F)$, and $C_A(F)$ 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 certainly occur in each disease 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 each equivalence class induces diseases which we are not able to distinguish by means of symptoms available in the information system S.

6.5. Diagnosis

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

The problem is the following: given a symptom ρ, (a) what diseases certainly have the symptom ρ, (b) what diseases possibly have the 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 $\underline{\mathrm{Apr}}_A(X_{i_j}) \supset E_\rho$ certainly have the symptom ρ, and all diseases $Y_{i_1} \in F$ such that $\overline{\mathrm{Apr}}_A(Y_{i_1}) \supset E_\rho$ 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 the lower (upper) approximation of each subset X_i of F contains E_p, we can simply check whether the corresponding classes contain E_p or not, which considerably simplifies the algorithm.

REFERENCES

1. E. Konrad, E. Orłowska, and Z. Pawlak, *An approximate concept learning* (Berlin, Bericht, 1981), pp. 81–87.
2. W. Marek and Z. Pawlak, "Rough sets and information systems," *ICS PAS Reports* (441) (1981).
3. R. Michalski, "S., Pattern Recognition as Role-Guided Inductive Interference," *IEEE Transaction on Pattern Analysis and Machine Intelligence* **2**:179–187 (1971).
4. E. Orłowska, "Semantics of vague concepts, Application of rough sets," *ICS PAS Reports* (469) (1982).
5. E. Orłowska, "Logic of vague concepts, Application of rough sets," *ICS PAS Reports* (474) (1982).
6. E. Orłowska and Z. Pawlak, "Measurement and observability, Application of rough sets," (to appear).
7. Z. Pawlak, "Rough sets," *ICS PAS Reports* (431) (1981).
8. Z. Pawlak, "Rough relations," *ICS PAS Reports* (435) (1981).
9. Z. Pawlak, "Rough functions," *ICS PAS Reports* (167) (1981).
10. Z. Pawlak, "Information systems, theoretical foundations," *Information systems* **6** (3):205–218 (1981).
11. Z. Pawlak, "Rough sets, Algebraic and topological approach," *ICS PAS Reports* (482) (1982).
12. A. Robinson, *Non-standard analysis* (North-Holland Publishing Company, Amsterdam, 1966).
13. L. A. Zadah, "Fuzzy sets," *Information and Control* **8**:338–353 (1965).
14. E. O. Zeeman, "The Topology of the Brain and Visual Perception," in *Topology of 3-Manifolds and related topics*, M. K. Fort, ed. (Englewood Cliffs, N.Y., 1962).