BULLETIN OF THE POLISH
ACADEMY OF SCIENCES
TECHNICAL SCIENCES
Vol. 33, No. 11–12, 1985

VARIA

COMPUTER SCIENCE

Some Remarks on Rough Sets

by

Zdzisław PAWLAK

Presented by Z. PAWLAK on May 13, 1985

Summary. In this paper we give somewhat more general formulation of the rough set concept, then that presented in previous publications concerning this subject.

1. Introduction. In [1] a new approach to vague and imprecise data analysis has been proposed, different to that offered by fuzzy set theory [2]. The approach is based on the fact that in many applications we are given a set of objects, states, processes, phases etc., but we are unable to distinguish them by available means of measurements, observations or description. For example, if objects are patients suffering from a certain disease and we characterize the health status of each patient in terms of some symptomps it may happen that some patients display the same symptoms, thus we are unable to discern them employing these symptoms.

To deal with this kind of problems the concept of rough set has been introduced and investigated.

In this note we define the concept of the rough set in purely algebraic way, which enables to use standard mathematical tools to deal with problems involving vagueness.

2. An approximation space. An *approximation space* is an ordered pair $\mathbf{A} = (U, \mathbf{R})$, where $U = \emptyset$ is a set called *universe* and $\mathbf{R} = \{R_1, R_2, ..., R_n\}$, $R_i \subseteq U \times U$ is a family of *primitive indiscernibility relations*. In this paper we assume that every R_i is an equivalence relation. Finite intersection of indiscernibility relations is also an indiscernibility relation. Equivalence classes of an indiscernibility relation R in $\mathbf{A}$ are called *R-elementary sets* in $\mathbf{A}$. If R is an indiscernibility relation in $\mathbf{A}$ and $(x, y) \in R$ — we say that x and y are *indiscernible* with respect to R in $\mathbf{A}$.

If $\mathbf{A} = (U, \mathbf{R})$ is an approximation space and $\mathbf{P} \subseteq \mathbf{R}$, then the approximation space $\mathbf{A} = (U, \mathbf{P})$ is called **P**-*subapproximation space* of $\mathbf{A}$ and is denoted by $\mathbf{A}/\mathbf{P}$.

If $\mathbf{A} = (U, \mathbf{R})$ is n approximation space, $X \subseteq U$ and $\mathbf{P} = \{P_1, P_2, \ldots, P_n\}$, where $P_i = R_i/X \times X$, then the approximation space $\mathbf{A} = (X, \mathbf{P})$ is called X-*subapproximation* space of $\mathbf{A}$ and is denoted by $\mathbf{A}/X$.

Let $\mathbf{A} = (U, \mathbf{R})$ be an approximation space, $\mathbf{P} \subseteq \mathbf{R}$ and let $\tilde{\mathbf{P}} = \bigcap_{R \in \mathbf{P}} \tilde{R}$.

A family $\mathbf{P}$ is *independent* in $\mathbf{A}$, if for every $\mathbf{Q} \subset \mathbf{P}$, $\tilde{\mathbf{Q}} \supset \tilde{\mathbf{P}}$; otherwise, i.e. if there exists $\mathbf{Q} \subset \mathbf{P}$ such that $\tilde{\mathbf{Q}} = \tilde{\mathbf{P}}$, then $\mathbf{P}$ is *dependent* in $\mathbf{A}$.

A family $\mathbf{Q} \subseteq \mathbf{P}$ is a *reduct* of $\mathbf{P}$ in $\mathbf{A}$ if $\mathbf{Q}$ is the maximal independent family in $\mathbf{P}$.

A approximation space $\mathbf{A} = (U, \mathbf{R})$ is *independent* (*dependent*, *reduced*) if $\mathbf{R}$ is independent (dependent, reduced). If $\mathbf{P}$ is a reduct of $\mathbf{R}$ in $\mathbf{A}$ then $\mathbf{A}/\mathbf{P}$ is called the reduct of $\mathbf{A}$.

3. Approximation of sets. Let $\mathbf{A} = (U, \mathbf{R})$ be an approximation space, $X \leqslant U$ and $\mathbf{R}$ — an indiscernibility relation in $\mathbf{A}$. For any X and $\mathbf{R}$ we define two sets

$$\underline{R}X = \{x \in U : [x]_R \subseteq X\}$$

$$\bar{R}X = \{x \in U : [x]_R \cap X \neq \emptyset\}$$

called *R-lower* and *R-upper approximation* of X in $\mathbf{A}$, respectively.

The set $Bn_R(X) = \bar{R}X - \underline{R}X$ will be called *R-boundary* of X in $\mathbf{A}$.

If $\bar{R}X = \underline{R}X$, then X is called *R-definable* in $\mathbf{A}$; otherwise, i.e. if $\bar{R}X \neq \underline{R}X$ set X is called *R-nondefinable* in $\mathbf{A}$. R-nondefinable sets in $\mathbf{A}$ will be called *rough sets* with respect to R in $\mathbf{A}$.

The number

$$\mu_R(X) = \frac{\text{card}(\underline{R}X)}{\text{card}(\bar{R}X)}$$

is called the *accuracy of* X with respect to R in $\mathbf{A}$, and the number

$$\eta_R(X) = 1 - \mu_R(X)$$

is called the *roughness* of X with respect to R in $\mathbf{A}$.

We shall employ the following definitions:

a) the set $\underline{R}X$ is called the *positive region* of the set X with respect to R in $\mathbf{A}$

b) the set $Bn_R(X)$ is called the *doubtful region* of the set X with respect to R in $\mathbf{A}$

c) the set $U - \bar{R}X$ is called the *negative region* of the set X with respect to R in $\mathbf{A}$

4. Properties of approximations. Each indiscernibility relation R in $\mathbf{A} = (U, \mathbf{R})$ defines the topological space $T_R = (U, \mathrm{Def}_{\mathbf{A}}(R))$, where $\mathrm{Def}_{\mathbf{A}}(R)$ — the family of all R-definable sets in $\mathbf{A}$ — is the topology for U and it is the family of open and closed sets in T_R. The family of all R-elementary sets in $\mathbf{A}$ is a base of T_R. The R-lower and R-upper approximation of X in $\mathbf{A}$ are interior and closure operations in the topological space T_R, respectively, and hence the following properties

A1) $\underline{R}X \subseteq X \subseteq \bar{R}X$

A2) $\underline{R}U = \bar{R}U = U$

A3) $\underline{R}\emptyset = \bar{R}\emptyset = \emptyset$

A4) $\bar{R}(X \cup Y) = \bar{R}X \cup \bar{R}Y$

A5) $\underline{R}(X \cup Y) \supseteq \underline{R}X \cup \underline{R}Y$

A6) $\bar{R}(X \cap Y) \subseteq \bar{R}X \cap \bar{R}Y$

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

A8) $\bar{R}(-X) = -\underline{R}(X)$

A9) $\underline{R}(-X) = -\bar{R}(X)$

moreover, we have

A10) $\underline{R}\underline{R}X = \bar{R}\underline{R}X = \underline{R}X$

A11) $\bar{R}\bar{R}X = \underline{R}\bar{R}X = \bar{R}X$.

5. Classification of rough sets. Let $\mathbf{A} = (U, \mathbf{R})$, R — be an indiscernibility relation in $\mathbf{A}$ and $X \subseteq U$ be R-nondefinable set in $\mathbf{A}$ (rough with respect to R in $\mathbf{A}$). We can classify rough sets as follows

B1) X is *roughly R-definable* in $\mathbf{A}$ if $\underline{R}X \neq \emptyset$ and $\bar{R}X \neq U$

B2) X is *externally R-nondefinable* in $\mathbf{A}$ if $\underline{R}X \neq \emptyset$ and $\bar{R}X = U$

B3) X is *internally R-nondefinable* in $\mathbf{A}$ if $\underline{R}X = \emptyset$ and $\bar{R}X \neq U$

B4) X is *totally R-nondefinable* in $\mathbf{A}$ if $\underline{R}X = \emptyset$ and $\bar{R}X = U$.

Let us notice that if X is R-definable (roughly R-definable, totally R-nondefinable) so is $-X$. If X is externally (internally) R-nondefinable, then $-X$ is internally (externally) R-nondefinable.

6. Approximation of families of sets. Let $\mathbf{A} = (U, \mathbf{R})$, $\mathbf{F} = \{X_1, X_2, \ldots, X_m\}$ $X_i \subseteq U$ be a family of subsets of U and let R be an indiscernibility relation in **A**. For any family of subsets **F** and indiscernibility relation R we define the two families of sets

$$\underline{R}\mathbf{F} = \{\underline{R}X_1, \underline{R}X_2, \ldots, \underline{R}X_m\}$$
$$\bar{R}\mathbf{F} = \{\bar{R}X_1, \bar{R}X_2, \ldots, \bar{R}X_m\}$$

called the *R-lower* and *R-upper approximation* of **F** in **A**.

The *R-positive region* of the family **F** is the set

$$\mathrm{Pos}_R(\mathbf{F}) = \bigcup_{X\in\mathbf{F}} \underline{R}X$$

The *R-doubtful region* of the family **F** is the set

$$Bn_R(\mathbf{F}) = \bigcup_{X\in\mathbf{F}} Bn_R(X)$$

The *R-negative region* of the family **F** is the set

$$\mathrm{Neg}_R(\mathbf{F}) = \bigcup_{X\in\mathbf{F}} \mathrm{Neg}_R(X)$$

where $\mathrm{Neg}_R(X) = U - \bar{R}X$.

The numbers

$$\gamma_R(\mathbf{F}) = \frac{\mathrm{card}\,(\mathrm{Pos}_R(\mathbf{F}))}{\mathrm{card} \bigcup_{X\in\mathbf{F}} \bar{R}X}$$

and

$$\beta_R(\mathbf{F}) = \frac{\mathrm{card}\,(\mathrm{Pos}_R(\mathbf{F}))}{\sum_{X\in\mathbf{F}} \mathrm{card}\,\bar{R}X}$$

are called the *quality* and the *accuracy* of the family **F** with respect to R in **A**. Of course

$$0 \leqslant \gamma_R(\mathbf{F}) \leqslant \beta_R(\mathbf{F}) \leqslant 1.$$

7. Dependency of indiscernibility relations. Let P, R be the two indiscernibility relations in $\mathbf{A} = (U, \mathbf{R})$ and let P^* stands for the family of equivalence classes of P.

We say that P *depends* in *degree* k on R in **A**, in symbols $R \xrightarrow{k} P$, if $k = \gamma_R(P^*)$.

If $R \xrightarrow{k} P$ we shall also say that the dependency $R \xrightarrow{k} P$ is true in degree k in **A**. If $k = 1$ we say that $R \xrightarrow{1} P$ is *true* in **A** and we shall write $R \to P$. If $0 < k < l$ we say that $R \xrightarrow{k} P$ is *roughly true* in **A** and if $k = 0$ we say that $R \xrightarrow{0} P$ is *false* in **A**.

If $R \xrightarrow{k} P$ is true in **A**, then $(x, y) \in R$ implies $(x, y) \in P$ for every

$x, y \in U$. If $R \xrightarrow{k} P$ is roughly true in $\mathbf{A}$, then $(x, y) \in R$ implies $(x, y) \in P$ for some $x, y \in U$ and if $R \xrightarrow{k} P$ is false in $\mathbf{A}$ then $(x, y) \in R$ does not imply $(x, y) \in P$ for any $x, y \in U$.

Property 1. $R \xrightarrow{k} P$ in $\mathbf{A}$ iff $R \xrightarrow{1} P$ in $\mathbf{A}/\mathrm{Pos}_R(P^*)$ and $R \xrightarrow{0} R$ in $\mathbf{A}/Bn_R(P^*)$.

Property 2. The following conditions are equivalent

1) $R \to P$
2) $\tilde{R} \subseteq \tilde{P}$
3) $\widetilde{R \cap P} = \tilde{R}$
4) $\gamma_R(P^*) = 1$
5) $\beta_R(P^*) = 1$
6) $\underline{R}(P^*) = \bar{R}(P^*)$

Property 3.

1) If $R \xrightarrow{k} P$ and $Q \xrightarrow{l} P$, then $R \cap Q \xrightarrow{m} P$, where $m \geqslant \max(k, l)$
2) If $R \cap P \xrightarrow{k} Q$, then $R \xrightarrow{l} Q$ and $P \xrightarrow{m} Q$, where $l, m \leqslant k$
3) If $R \xrightarrow{k} Q$ and $R \xrightarrow{l} P$, then $R \xrightarrow{m} Q \cap P$, where $m \leqslant \min(k, l)$
4) If $R \xrightarrow{k} Q \cap P$, then $R \xrightarrow{l} Q$ and $R \xrightarrow{m} P$, where $l, m \geqslant k$
5) If $R \xrightarrow{k} P$ and $R \xrightarrow{l} Q$, then $R \xrightarrow{m} Q$, where $m \geqslant \max(k, l)$.

8. Approximation space and information systems. An *information system* is 4-tuple $S = (U, A, V, f)$
where

U — is a set called the *universe*

A — is a set of *attributes*

$V = \bigcup_{a \in A} V_a$ — is the set of *values* of an attribute a

$f: U \times A \to V$ — is an *information function*.

With every information system S we can associate the approximation space $\mathbf{A}_S = (U, \mathbf{R}_S)$, defined thus: every subset of attributes $B \subseteq A$ defines a binary relation

$$\tilde{B} = \{(x, y): f(x, a) = fy, a) \quad \text{for every} \quad a \in B\}.$$

Of course $\tilde{B}$ is an equivalence relation and can be regarded as an indiscernibility relation in $\mathbf{A}_S$. Certainly $\tilde{a}$ is a primitive indiscernity relation for any $a \in A$.

Conversely, with every approximation space $\mathbf{A} = (U, \mathbf{R})$ we can associate an information system $S_{\mathbf{A}} = (U, A, V, f)$ defined as follows: with every primitive indiscernibility relation $R \in \mathbf{R}$ we associate uniquelly a name a_R of R and we define the information function f in such a way that $f(x, a_R) = f(y, a_R)$ iff x and y belong to the same equivalence class of the relation R.

Thus the concepts of an approximation space and that of information system are isomorphic and can be mutally replaced.

INSTITUTE OF COMPUTER SCIENCE, POLISH ACADEMY OF SCIENCES, P.O. Box 22, PKIN, 00-901 WARSAW
DEPARTMENT OF COMPUTER SCIENCE, UNIVERSITY OF NORTH CAROLINA AT CHARLOTTE, CHARLOTTE, N.C. 28223

REFERENCES

[1] Z. Pawlak, *Rough sets*, Int. J. Inf. Comp. Sci., **11** (1982), 341–356.
[2] L. A. Zadeh, *Fuzzy logic and approximate reasoning*, Synthese, **30** (1975), 407–428.

З. Павляк, **Несколько замечаний по приближенным множествам**

В настоящей работе дается несколько более общая формулировка понятия приближенного множества, чем в предыдущих работах.