PRACE IPI PAN • ICS PAS REPORTS

Wiktor Marek, Zdzisław Pawlak

Rough sets and information systems

441

July 1981

WARSZAWA

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

Wiktor Marek, Zdzisław Pawlak

ROUGH SETS AND INFORMATION SYSTEMS

441

Warsaw, July 1981

Pracę zgłosił Witold Lipski

Mailing address: Wiktor Marek
University of Warsaw
00-901 Warszawa PKiN IX floor

Zdzisław Pawlak
Institute of Computer Science PAS
P.O. Box 22
00-901 Warszawa PKiN

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 700 egz. Ark. wyd. 0,45; ark. druk. 1,00. Papier offset. kl. III, 70 g, 70 x 100. Oddano do druku w lipcu 1981 r. W. D. N. Zam. nr 390/81

Streszczenie . Abstract . Содержание

Stosujemy zbiory przybliżone do scharakteryzowania zbiorów definiowalnych w systemie informacyjnym.

Rough sets and information systems

We apply rough sets to characterize definable subsets of the universe of the information system.

Приближенные множества и информационные системы

В работе применяется приближенные множества к характеристике определимых множеств в информационных системах.

0. INTRODUCTION

Approximate classification of objects is an important task in various fields. Formal tools dedicated to deal with such class of problems are offered by fuzzy sets theory of Zadeh ([6]) and tolerance theory of Zeeman ([7]). Another proposal for approximate classification has been considered in [4] where the notion of a rough (approximate) set is the departure point of the proposed method. The method is based on the upper and lower approximation of a set.

The approximation operations on sets are closely related to the theory of subsystems of the information system as developed in ours [3] . The approximation approach allows to explain some facts concerning the embeddings of algebras of describable sets. This problem is considered in detail in this note. For the completeness sake we recapitulate the basic properties of rough sets in the section 1.

1. PRELIMINARIES

Let X be a set, called an <u>universum</u> and $\mathcal{S} = \langle X,D,A,U \rangle$ be an information system on the set X. With the system $\mathcal{S}$ we adjoin a language $\mathcal{L}_{\mathcal{S}}$. This formal language allows us to describe some subsets of the sets X. The describable sets form a Boolean algebra $\mathcal{B}(\mathcal{S})$. The atoms of this algebra are called <u>constituents</u> of the system $\mathcal{S}$ or <u>elementary sets</u> in $\mathcal{S}$. These sets are nonempty and pairwise disjoint. Every describable set is a finite union of elementary sets.

Let Z be a subset of the set X. The least describable subset $T \subseteq X$ such that $Z \subseteq T$ exists and is called <u>best upper approximation</u> of Z in $\mathcal{A}$ or closure of Z and denoted by $\overline{Z}$. Analogously the largest describable subset $T \subseteq X$ such that $T \subseteq Z$ again exists and is called <u>best lower approximation</u> of Z in $\mathcal{A}$ or <u>interior</u> of Z and is denoted by $\underline{Z}$.

The set $Fr(X) = \overline{X} - \underline{X}$ is called the <u>boundary of</u> X in $\mathcal{A}$.

The set $E(X) = \overline{X} - X$ is called <u>the edge of</u> X in $\mathcal{A}$.

As is easily seen the operation $\overline{\ \cdot\ }$ is a closure operation in the sense of Kuratowski, $\underline{\ \cdot\ }$ is the adjoint interior operation. This means that the following facts are true:

1° $\overline{Z} \supseteq Z \supseteq \underline{Z}$. 2° $\overline{1} = \underline{1} = 1$. 3° $\overline{0} = \underline{0} = 0$

4° $\overline{\overline{Z}} = \overline{Z}$ 5° $\underline{\underline{Z}} = \underline{Z}$ 6° $\overline{Z \cup T} = \overline{Z} \cup \overline{T}$

7° $\underline{Z \cap T} = \underline{Z} \cap \underline{T}$ 8° $\overline{Z} = -(\underline{-Z})$ 9° $\underline{Z} = -(\overline{-Z})$

The following is also of interest:

10° $\underline{(\overline{Z})} = \overline{Z}$ 11° $\overline{(\underline{Z})} = \underline{Z}$

12° Moreover, for every set $Z \in \mathcal{B}(\mathcal{A})$, $\overline{Z} = \underline{Z} = Z$

The following facts could be of interest

13° $\overline{X \cap Y} \subseteq \overline{X} \cap \overline{Y}$

14° $\underline{X \cup Y} \supseteq \underline{X} \cup \underline{Y}$

15° $\overline{X} - \overline{Y} \subseteq \overline{X - Y}$

16° $\underline{X} - \underline{Y} \supseteq \underline{X - Y}$

17° $\underline{X} \cup \overline{-X} = 1$

18° $\overline{X} \cup \underline{-X} = 1$

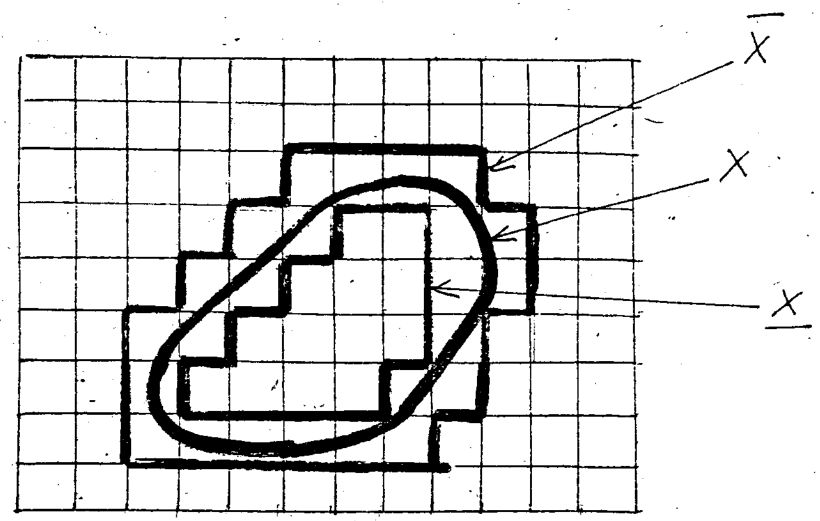
19° $\underline{X} \cup \underline{(-X)} = -Fr(X)$

20° $\overline{X} \cap \overline{(-X)} = Fr(X)$

Classical de Morgan laws have various counterparts here, for instance

21° $\overline{-(\underline{X} \cup \underline{Y})} = \overline{(-X)} \cap \overline{(-Y)}$ etc.

The following picture depicts our situation



We define two additional membership relations $\underline{\in}$ and $\overline{\in}$ called <u>strong</u> and <u>weak</u> membership as follows:

$x \underline{\in} X$ iff $x \in \underline{X}$ and $x \overline{\in} X$ iff $x \in \overline{X}$

Those have clear meanings: "x definitely is in X" and "x is possibly in X". They may be interpreted as $\Diamond x \in X$ and $\Box x \in X$ in the sense of modal logic.

Now we have the following 3 equivalence relations in $\mathcal{P}(X)$:

1° $Z \simeq T \iff \underline{Z} = \underline{T}$

2° $Z \eqsim T \iff \overline{Z} = \overline{T}$

3° $Z \approx T \iff Z \simeq T \wedge Z \eqsim T$.

The following properties of these relations are provable:

4° $X \simeq 0 \iff \underline{X} = 0$ 5° $Z \eqsim 0 \iff Z = 0$

6° $X \simeq 1 \iff X = 1$ 7° $Z \eqsim 1 \iff \overline{Z} = 1$

8° $\mathrm{Fr}X \simeq 0$.

The sets with the property $X \simeq 0$ are called loose sets, the boundary of X is always loose.

We list below a couple more properties of $\simeq$ and $\eqsim$:

9° If $Z \simeq T$ then $Z \cap T \simeq Z \simeq T$

10° If $Z \eqsim T$ then $Z \cup T \eqsim Z \eqsim T$

11° If $Z \eqsim Z'$ and $T \eqsim T'$ then $Z \cup T \eqsim Z' \cup T'$

12° If $Z \simeq Z'$ and $T \simeq T'$ then $Z \cap T \simeq Z' \cap T'$

One introduces the corresponding notions of approximation:

1° $Z \underset{\sim}{\subset} T \iff \underline{Z} \le \underline{T}$

2° $Z \overset{\sim}{\subset} T \iff \overline{Z} \le \overline{T}$

3° $Z \underset{\sim}{\overset{\sim}{\subset}} T \iff Z \underset{\sim}{\subset} T \wedge Z \overset{\sim}{\subset} T$

The following holds

4° If $Z \underset{\sim}{\subset} T$ and $T \underset{\sim}{\subset} Z$ then $Z \simeq T$

5° If $Z \overset{\sim}{\subset} T$ and $T \overset{\sim}{\subset} Z$ then $Z \eqsim T$

6° If $Z \underset{\sim}{\overset{\sim}{\subset}} T$ and $T \underset{\sim}{\overset{\sim}{\subset}} Z$ then $Z \approx T$

7° If $Z \subset T$ then $Z \overset{\sim}{\subset} T$, $Z \underset{\sim}{\subset} T$, $Z \underset{\sim}{\overset{\sim}{\subset}} T$

8° If $Z \underset{\sim}{\subset} T$, $Z \simeq Z'$, $T \simeq T'$ then $Z' \underset{\sim}{\subset} T'$

There is a couple more of properties of the type of 8°.

2. APPLICATION OF THE CLOSURE PROPERTIES TO THE INVESTIGATIONS OF SUBSYSTEMS OF INFORMATION SYSTEMS

Proposition 1: If t is a primitive term of the language $\mathcal{L}_{\mathcal{S}}$ and Z is a describable subset of the set $X (= X_{\mathcal{S}})$ then

$$\|t\|_{\mathcal{S}\restriction Z} \neq \emptyset \Rightarrow \|t\|_{\mathcal{S}\restriction Z} = \|t\|_{\mathcal{S}}$$

Proof: Let s be a description of the set Z in $\mathcal{S}$. We have $\|s \cdot t = t \vee s \cdot t = 0\|_{\mathcal{S}} = V$, this means that:

$\|t\|_{\mathcal{S}} \le \|s\|_{\mathcal{S}}$ or $\|t\|_{\mathcal{S}} \cap \|s\|_{\mathcal{S}} = \emptyset$

If $\|t\|_{\mathcal{S}} \cap \|s\|_{\mathcal{S}} = \emptyset$ then $\|t\|_{\mathcal{S}\restriction Z} = \emptyset$ since $Z = \|s\|_{\mathcal{S}}$ and $\|t\|_{\mathcal{S}\restriction Z} = \|t\|_{\mathcal{S}} \cap Z = \|t\|_{\mathcal{S}} \cap \|s\|_{\mathcal{S}}$. So we are left with the case $\|t\|_{\mathcal{S}} \le \|s\|_{\mathcal{S}}$ If $\|t\|_{\mathcal{S}} = \emptyset$ then there is nothing to prove. Otherwise $\|t\|_{\mathcal{S}} \neq \emptyset$ and then $\|t\|_{\mathcal{S}\restriction Z} = \|t\|_{\mathcal{S}} \cap Z = \|t\|_{\mathcal{S}} \cap \|s\|_{\mathcal{S}} = \|t\|_{\mathcal{S}}$.

$\square$

This means that if we restrict our system to a definable subset then the constituent do not change or vanish.

We have the following relationship between the sets Z and $\overline{Z}$:

Theorem 2: If $Z \subseteq X_{\mathcal{S}}$ then

$$\mathfrak{B}(\mathcal{S}\restriction Z) \cong \mathfrak{B}(\mathcal{S}\restriction \overline{Z})$$

moreover the isomorphism is given by:

$$\varphi(\|t\|_{\mathcal{S}\restriction Z}) = \|t\|_{\mathcal{S}\restriction \overline{Z}}$$

Proof: Let us investigate $\overline{Z}$. It happens that $x \in \overline{Z}$ exactly in the case when there exists a constituent T oif the system $\mathcal{S}$ (i.e. the value of a primitive term) such that $x \in T$ and $T \cap Z \neq \emptyset$. Since in general it holds: $\|t\|_{\mathcal{S}\upharpoonright Z} = \|t\|_{\mathcal{S}} \cap Z$ (cf. [ ]) we have, for primitive terms t:

(*) $\quad \|t\|_{\mathcal{S}\upharpoonright Z} \neq 0 \iff \|t\|_{\mathcal{S}\upharpoonright \overline{Z}} \neq \emptyset$

Now the algebras $\mathcal{B}(\mathcal{S}\upharpoonright Z)$ and $\mathcal{B}(\mathcal{S}\upharpoonright \overline{Z})$ are generated by non-empty constituents of the system $\mathcal{S}\upharpoonright Z$ and $\mathcal{S}\upharpoonright \overline{Z}$ respectively. It follows from (*) that both algebras have same number of generators. Thus they are isomorphic. The map $\|t\|_{\mathcal{S}\upharpoonright Z} \longmapsto \|t\|_{\mathcal{S}\upharpoonright \overline{Z}}$ is an injection (again by (*)) and uniquely extends to algebras $\mathcal{B}(\mathcal{S}\upharpoonright Z)$ and $\mathcal{B}(\mathcal{S}\upharpoonright \overline{Z})$. $\square$

We discuss for a moment selective systems. The following is fairly simple.

Proposition 3: If $\mathcal{S}$ is a selective system then:

(i) for every $Z \subseteq X_{\mathcal{S}}$, $\overline{Z} = \underline{Z} = Z$

(ii) $\mathcal{B}(\mathcal{S}) = \mathcal{P}(X_{\mathcal{S}})$.

Now let us investigate how the operations $\overline{\cdot}$ and $\underline{\cdot}$ behave with respect to subsystem and extensions. Notice first that if $\mathcal{S}_1 \leq \mathcal{S}_2$ then the operation $\underline{\cdot}$ in $\mathcal{S}_1$ is not the trace of $\underline{\cdot}$ in $\mathcal{S}_2$. One can show a "drastic" example namely it is easy to construct $\mathcal{S}_1 \leq \mathcal{S}_2$ and $Z \leq X_{\mathcal{S}_1}$ such that $\underline{Z}_{\mathcal{S}_1} = Z$ and $\underline{Z}_{\mathcal{S}_2} = \emptyset$.

The operation $\overline{\cdot}$ behave nicer; since for every term, $\|t\|_{\mathcal{S}_1} = \|t\|_{\mathcal{S}_2} \cap X$ therefore, for $Z \subseteq X_{\mathcal{S}_1}$, $\overline{Z}^{\mathcal{S}_2} \cap X_{\mathcal{S}_1} = \overline{Z}^{\mathcal{S}_1}$.

(This asymetry is related to the difference in the behavior of interior and closure operation).

We investigate the relationship between the algebras $\mathcal{B}(\mathcal{S}\upharpoonright Z)$ and $\mathcal{B}(\mathcal{S})$. The following characterization result was proved independently by M. Jaegermann (oral communication):

Theorem 4: The following are equivalent:

(i) $Z \subseteq X_{\mathcal{S}}$ is describable

(ii) $\mathcal{B}(\mathcal{S}) \cong \mathcal{B}(\mathcal{S}\upharpoonright Z) \times \mathcal{B}(\mathcal{S}\upharpoonright X_{\mathcal{S}} - Z)$

Proof: (i) $\Rightarrow$ (ii). This follows from calculating the number of generators of the algebras under consideration, i.e. the number of non-empty constituents. If $\mathcal{S}$ possesses k nonempty constituents then $\mathcal{B}(\mathcal{S}) \cong 2^{\{1,\dots,k\}}$. Now let Z be the union of l among them. Then $X_{\mathcal{S}} \setminus Z$ is the union of remaining $k-l$ constituents. Now by the proposition 1,

$$\mathcal{B}(\mathcal{S}\upharpoonright Z) \simeq 2^{\{1,\dots,l\}}, \quad \mathcal{B}(\mathcal{S}\upharpoonright (X_{\mathcal{S}}\setminus Z)) \cong 2^{\{1,\dots,k-l\}}$$

This gives our implication.

(ii) $\Rightarrow$ (i) Assume Z is not describable. This means $\underline{Z} \neq Z \neq \overline{Z}$. Assume again that $\mathcal{S}$ has k nonempty constituents, $\mathcal{S}\upharpoonright Z$ has l nonempty constituents. Consider set $\mathrm{Fr}Z = \overline{Z} - \underline{Z}$. Since it is nonempty and describable it is the union of at least one constituent. Assume it is union of m constituents. Now $\mathcal{B}(\mathcal{S}\upharpoonright Z)$ has l constituents and $\mathcal{B}(\mathcal{S}\upharpoonright (X_{\mathcal{S}}\setminus Z))$ has $k-l+m$ constituents. Thus the product $\mathcal{B}(\mathcal{S}\upharpoonright Z) \times \mathcal{B}(\mathcal{S}\upharpoonright (X_{\mathcal{S}}\setminus Z))$ has $k+m$ generators and so is not isomorphic with $\mathcal{B}(\mathcal{S})$ since the latter has k generators. $\square$

Theorem 5: Let $Z \subseteq X_{\mathcal{S}}$. Then

$$\mathcal{B}(\mathcal{S}) \cong \mathcal{B}(\mathcal{S}\upharpoonright Z) \times \mathcal{B}(\mathcal{S}\upharpoonright (\underline{X_{\mathcal{S}} - Z}))$$

Proof: By the theorem 2, $\mathcal{B}(\mathcal{S}\upharpoonright Z) \cong \mathcal{B}(\mathcal{S}\upharpoonright \overline{Z})$. Now, by the theorem 4, $\mathcal{B}(\mathcal{S}) \cong \mathcal{B}(\mathcal{S}\upharpoonright \overline{Z}) \times \mathcal{B}(\mathcal{S}\upharpoonright (X_{\mathcal{S}}\setminus \overline{Z}))$. Since $X_{\mathcal{S}} \setminus \overline{Z} = \underline{X_{\mathcal{S}} \setminus Z}$ the result follows.

Let us call definable restriction of $\mathfrak{A}$ every subsystem $\mathfrak{A}\restriction Z$ where $Z \in \mathcal{B}(\mathfrak{A})$.

The following is useful:

Proposition 6: There exists largest selective definable restriction of $\mathfrak{A}$.

Proof: Its universe consists of those x's for which $\|t_x\| = \{x\}$ □

The following important property of definable restrictions holds:

Proposition 7:

Let $\mathfrak{A}_0$ be a definable restriction of $\mathfrak{A}_2$ and let $\mathfrak{A}_1$ be $\mathfrak{A}_2\restriction(X_{\mathfrak{A}_2}\setminus X_{\mathfrak{A}_0})$ then for every $Z \subseteq X_{\mathfrak{A}_2}$ we have

$$\overline{Z}^{\mathfrak{A}_2} = \overline{Z\cap X_{\mathfrak{A}_0}}^{\mathfrak{A}_0} \cup \overline{Z\cap X_{\mathfrak{A}_1}}^{\mathfrak{A}_1}$$

The proof of this fact follows from the proposition 1. □

Now let $\mathfrak{A}_0$ be the largest selective definable restriction of $\mathfrak{A}$ and let $\mathfrak{A}_1$ be $\mathfrak{A}\restriction(X_{\mathfrak{A}}\setminus X_{\mathfrak{A}_0})$. According to the definition of $\mathfrak{A}_0$ and proposition 1 every constituent of $\mathfrak{A}_1$ consists of at least two elements. Moreover $\mathcal{B}(\mathfrak{A}) \cong \mathcal{B}(\mathfrak{A}_0)\times\mathcal{B}(\mathfrak{A}_1)$. The system $\mathfrak{A}_1$ is called totally nonselective. Our remarks boil down to the following:

Proposition 8: (i) There is a unique decomposition of the system $\mathfrak{A}$ into selective definable restriction $\mathfrak{A}'$ and totally nonselective definable restriction $\mathfrak{A}''$

(ii) In the above situation we have, for every $Z \subseteq X_{\mathfrak{A}}$

$$\overline{Z}^{\mathfrak{A}} = (Z\cap X_{\mathfrak{A}'}) \cup (\overline{Z\cap X_{\mathfrak{A}''}}^{\mathfrak{A}''})$$

(iii) Similarly

$$\underline{Z}_{\mathfrak{A}} = (Z\cap X_{\mathfrak{A}'}) \cup (\underline{Z\cap X_{\mathfrak{A}''}}_{\mathfrak{A}''})$$

Proof: (i) Uniqueness follows from the proposition 1 and existence from the proposition 6.

(ii) and (iii) follow by propositions 6, 7 and 3 □

It follows from the proposition 8, that the operations $\overline{}$ and $\underline{}$ are of interest only for the totally nonselective systems (which just means that Hausdorff part of the corresponding topological space is empty).

Proposition 9: If $\mathfrak{A}$ is totally nonselective then there exist sets Z and T with the following properties

(i) $Z \cap T = \emptyset$

(ii) $\underline{Z} = \underline{T} = \emptyset$

(iii) $\overline{Z} = \overline{T} = X_{\mathfrak{A}}$

Proof: Choose Z to be a selector of constituents of $\mathfrak{A}$ and T its complement. □

Theorem 10: Let $\mathfrak{A}$ be totally nonselective and C, D describable subsets of $X_{\mathfrak{A}}$, moreover $C \subseteq D$. Then there exists $Z \subseteq X$ such that $\underline{Z} = C$ and $\overline{Z} = D$.

Proof: We follow the construction of the proposition 9. Consider $D\setminus C$ split it into constituents and let T be a selector of these. By total nonselectiveness none of the constituents of $D\setminus C$ is included in T. Thus $\overline{C\cup T} = D$ whereas $\underline{C\cup T} = C$. □

The result of the theorem 10 is used to characterize the algebra $\mathcal{P}(X_{\mathfrak{A}})/\approx$. Since $Z_1 \approx Z_2 \iff \underline{Z}_1 = \underline{Z}_2 \ \&\ \overline{Z}_1 = \overline{Z}_2$ therefore an equivalence class of the relation $\approx$ is determined by the pair $\langle C,D\rangle$ of describable sets such that $C \subseteq D$. Now by

the theorem 10, in totally nonselective system each such pair
determines an equivalence class (of Z's such that $\underline{Z} = C$ and
$\overline{Z} = D$). Let us introduce now, in the set of pairs $\langle C,D \rangle$
such that $C \subseteq D$ the operations "coordinate-wise". By the results
of Traczyk [5] and Dwinger [1] [2] the resulting structure is
a Post algebra with 3 generators which is naturally related to
the three-valued logic. It is far from being strange since the
elements of $\underline{Z}$ (i.e. those x's which $\underline{\in}$ - belong to Z) are
in Z with value 1, the elements of $X \setminus \overline{Z}$ (i.e. those x's
which $\overline{\in}$ - do not belong to Z) are in Z with value 0 where-
as the elements of $Fr(Z)$ belong to Z with value 1/2 since
they are undistinguishable (from the point of view of $\mathfrak{S}$) from
some elements which do not belong to Z.

Let us finally note that the algebras $\mathcal{P}(X)/\underline{\sim}$ and
$\mathcal{P}(X)/\eqsim$ are isomorphic to $\mathcal{B}(\mathfrak{S})$.

REFERENCES

1. Ph. Dwinger, Note on Post Algebras I, Ind. Math. 281/1966/ pp. 464-468
2. Ph. Dwinger, Note on Post Algebras II, Ind. Math. 281/1966/, pp. 468-478
3. W. Marek, Z. Pawlak, Information Storage and retrieval systems, Theoretical Computer Sciences, 1/1976/ pp. 331-354
4. Z. Pawlak, Rough sets, ICS PAS Reports /1981/, No 431
5. T. Traczyk, On axioms and some properties of Post algebras, Bull. Acad. Pol. Sci., No 10/1962/ pp. 509-512
6. L.A. Zadeh, Fuzzy sets, Information and Control /1965/ 8 pp. 338-353
7. F.O. Zeeman, The Topology of the Brain and Visual Perception in M.K. Fort / ed./ Topology of 3 - Manifolds and related topics. Englewood Cliffs N.Y. 1962