Annales Societatis Mathematicae Polonae
Series IV: Fundamenta Informaticae VII.1 (1984)

ROUGH SETS AND INFORMATION SYSTEMS

Wiktor M a r e k
University of Warsaw
Zdzisław P a w l a k
Institute of Computer Science PAS
Received October 25, 1982
AMS Categories: 68H05

A b s t r a c t: We apply rough sets to characterize definable subsets of the universe of the information system.

K e y w o r d s: approximation space, fuzzy sets, information systems, tolerance theory.

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 our [3] . The approximation approach allow to explain some facts concerning the embedding 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 universum and $S = \langle X_S, D, A, U \rangle$ be an information system on the set X_S . With the system S we adjoin a language L_S . This formal language allows us to describe some subsets of the set X . The describable sets form a Boolean algebras $B(S)$. The atoms of this algebra are called constituents of the system S or elementary sets in S . These sets are nonempty and pairwise disjoint. Every describable set is a finite union of elementary sets. For details see [3] .

It is easy to check that the information system $S = \langle X_S, D, A, U \rangle$ defines uniquely the topological space $T_S = \langle X_S, B(S) \rangle$, where the family of all describable sets in S is a family of open and closed sets in T_S , and the family of all elementary sets in S is a base for T_S .

Let Z be a subset of the set X . The least describable subset $T \subseteq X$ such that $Z \subseteq T$ is called best upper approximation of Z in S or closure of Z and denoted by $\bar{A}Z$. Analogously the largest describable subset $T \subseteq X$ such that $T \subseteq Z$ is called best lower approximation of Z in S or interior of Z and is denoted by $\underline{A}Z$.

The set $Fr(Y) = \overline{A}Y - \underline{A}Y$ is called the *boundary* of Y in S.

The set $E(Y) = \overline{A}Y - Y$ is called the *edge* of Y in S.

As is easily seen the operation $\overline{A}$ is a closure operation in the sense of Kuratowski, $\underline{A}$ is the adjoint interior operation, in the topological space T_S.

This means that the following facts are true:

1° $\overline{A}Z \supseteq Z \supseteq \underline{A}Z$

2° $\overline{A}X = \underline{A}X = X$

3° $\overline{A}\emptyset = \underline{A}\,\emptyset = \emptyset$

4° $\overline{A}\overline{A}Z = \overline{A}Z$

5° $\underline{A}\underline{A}Z = \underline{A}Z$

6° $\overline{A}(Y \cup Z) = \overline{A}Y \cup \overline{A}Z$

7° $\underline{A}(Y \cap Z) = \underline{A}Y \cap \underline{A}Z$

8° $\overline{A}Z = -\underline{A}(-Z)$

9° $\underline{A}Z = -\overline{A}(-Z)$

In the tological space T_S we have also the following properties:

10° $\underline{A}\overline{A}Z = \overline{A}Z$

11° $\overline{A}\underline{A}Z = \underline{A}Z$

Moreover

12° For every set $Z \in B(S)$ we have $\overline{A}Z = \underline{A}Z = Z$.

The following facts would be of interest

13° $\overline{A}(Y \cap Z) \subseteq \overline{A}Y \cap \overline{A}Z$

14° $\underline{A}(Y \cup Z) \supseteq \underline{A}Y \cup \underline{A}Z$

15° $\overline{A}Y - \overline{A}Z \subseteq \overline{A}(Y - Z)$

16° $\underline{A}Y - \underline{A}Z \supseteq \underline{A}(Y - Z)$

17° $\underline{A}Y \cup \overline{A}(-Y) = X$

$19^\circ \quad \underline{A}Y \cup \underline{A}(-Y) = -Fr(Y)$

$20^\circ \quad \bar{A}Y \cap \bar{A}(-Y) = Fr(Y)$

Classical de Morgan laws have various cornterparts here, for instance

$21^\circ \quad -(\underline{A}Y \cup \underline{A}Z) = \bar{A}(-Y) \cap \bar{A}(-Z)$ etc.

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

$y \,\underline{\in}\, Y$ iff $y \in \underline{A}Y$ and $y \,\overline{\in}\, Y$ iff $y \in \bar{A}Y$.

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

Now we have the following 3 equivalence relations in $P(X)$:

$1^\circ \quad Z \simeq T \Leftrightarrow \underline{A}Z = \underline{A}T$

$2^\circ \quad Z \eqsim T \Leftrightarrow \bar{A}Z = \bar{A}T$

$3^\circ \quad Z \approx T \Leftrightarrow Z \simeq T \wedge Z \eqsim T$.

The following properties of these relations are provable:

$4^\circ \quad Y \simeq \emptyset \Leftrightarrow \underline{A}Y = \emptyset$ $\qquad 5^\circ \quad Z \eqsim \emptyset \Leftrightarrow Z = \emptyset$

$6^\circ \quad Y \simeq X \Leftrightarrow Y = X$ $\qquad 7^\circ \quad Z \eqsim X \Leftrightarrow \bar{A}Z = 1$

The sets with the property $Y \simeq 0$ are called <u>loose</u> sets, the boundary of X is always loose.

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

8° If $Z \simeq T$ then $Z \cap T \simeq Z \simeq T$

9° If $Z \eqsim T$ then $Z \cup T \eqsim Z \eqsim T$

10° If $Z \eqsim Z'$ and $T \eqsim T'$ then $Z \cup T \eqsim Z' \cup T'$

11° If $Z \simeq Z'$ and $T \simeq T'$ then $Z \cap T \simeq Z' \cap T'$

One introduces the corresponding notions for inclusion:

$12^\circ \quad Z \underset{\sim}{\subseteq} T \Leftrightarrow \underline{A}Z \subseteq \underline{A}T$

13° $Z \widetilde{\subset} T \Leftrightarrow \bar{A}Z \subset \bar{A}T$

14° $Z \widetilde{\underset{\sim}{\subseteq}} T \Leftrightarrow Z \underset{\sim}{\subseteq} T \wedge Z \widetilde{\subset} T$

The following holds

15° If $Z \underset{\sim}{\subseteq} T$ and $T \underset{\sim}{\subseteq} Z$ then $Z \simeq T$

16° If $Z \widetilde{\subset} T$ and $T \widetilde{\subset} Z$ then $Z \eqsim T$

17° If $Z \widetilde{\underset{\sim}{\subseteq}} T$ and $T \widetilde{\underset{\sim}{\subseteq}} Z$ then $Z \approxeq T$

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

In this section we use the notation introduced in [3] and we assume that the reader is familiar with this paper.

Proposition 1: If t is a primitive term of the language L_S and Z is a describable subset of the set X then

$\|t\|_{S\restriction Z} \neq \emptyset \supset \|t\|_{S\restriction Z} = \|t\|_S$.

Proof: Let $Z = \|s\|_S$. Notice that for any term t

$\|t\|_{S\restriction Z} = \|t\|_S \cap Z = \|t\|_S \cap \|s\|_S$.

By the assumption t is a primitive term. Thus

$\| s.t = t \vee s.t = F \|_S = X$.

So here have

1° $\|t\|_S \cap \|s\|_S = \emptyset$, then $\|t\|_{S\restriction Z} = \emptyset$, a contradiction.

2° $\|t\|_S \subset \|s\|_S$. Then $\|t\|_{S\restriction Z} = \|t\|_S$

3° $\|s.t = t\|_S = -\|s.t = F\|_S$ thus

$\|t\|_S = X$; a contradiction because t is a primitive term.

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

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

Theorem 2: If $Z \subsetneq X_S$ then there is an isomorphism φ between $B(S \restriction Z)$ and $B(S \restriction \bar{A}Z)$, i.e. $B(S \restriction Z) \cong B(S \restriction \bar{A}Z)$, moreover the isomorphism is given by:

$$\varphi \| t \|_{S \restriction Z} = \| t \|_{S \restriction \bar{A}Z}$$

Proof. Let $Z \subsetneq X_S$ and let us take $x \in \bar{A}Z$. Notice that $\bar{A}Z$ is describable in S . So $\bar{A}Z$ in the disjoint union of all constituents T in S contained in $\bar{A}Z$. Thus x belongs to one of them, say $x \in T$. Since $T \cap Z \neq \emptyset$ it follows from the definition of constituents in S and the valuation of terms in S that every constituent in S is a value of a primitive term. Thus $T = \| t \|_{S \restriction \bar{A}Z} \neq \emptyset$, where t is a primitive term.

Notice that for a primitive term t

(*) $\| t \|_{S \restriction Z} \neq \emptyset$ and $\| t \|_{S \restriction \bar{A}Z} \neq 0$.

Now the algebras $B(S \restriction Z)$ and $B(S \restriction \bar{A}Z)$ are generated by non-empty constituentes of the system $S \restriction Z$ and $S \restriction \bar{A}Z$ respectively. By the proposition 1 and the (*) φ is injection. It follows from the proposition 1, the (*) , and the valuation of terms that φ is homomorphism.

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

Proposition 3: If S is a selective system then:

(i) for every $Z \subsetneq X_S$, $\bar{A}Z = \underline{A}Z = Z$

(ii) $B(S) = P(X_S)$.

Now let us investigate how the operations $\bar{A}$ and $\underline{A}$ behave with respect to subsystem and extensions. Notice first that if $S_1 \subseteq S_2$ then the operation $\underline{\Lambda}$ in S_1 is not the trace of $\underline{\Lambda}$ in S_2. One can show a "drastic" example namely it is easy to construct $S_1 \subseteq S_2$ and $Z \subseteq X_{S_1}$ such that $\underline{\Lambda}_{S_1} Z = Z$ and $\underline{\Lambda}_{S_1} Z = \emptyset$.

The operation $\bar{\Lambda}\cdot$ behave nicer; since for every term, $\|t\|_{S_1} = \|t\|_{S_2} \cap X$ therefore, for $Z \subseteq X_{S_1}$, $\bar{\Lambda}_{S_2} Z \cap X_{S_1} = \bar{\Lambda}_{S_1} Z$.
/This asymetry is related to the difference in the behaviour of interior and closure operation/.

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

Theorem 4: The following are equivalent:

(i) $Z \subseteq X_S$ is discribable

(ii) $B(S) \cong B(S\upharpoonright Z) \times B(S\upharpoonright (X_S - Z))$

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 S possesses k non-empty constituents then $B(S) \cong 2^{\{1,\ldots,k\}}$. Now let Z be the union of l, $l \leq k$ among them. Then $X_S \setminus Z$ is the union of the remaining $k-l$ constituents. Now by the proposition 1,

$B(S\upharpoonright Z) \cong 2^{\{1,\ldots,l\}}$, $B(S\upharpoonright(X_S \setminus Z)) \cong 2^{\{1,\ldots,k-l\}}$

This gives our implication.

(ii) $\Rightarrow$ (i) Assume Z is not describable. This means $\underline{\Lambda}Z \neq Z \neq \bar{\Lambda}Z$.

Assume again that S has k nonempty constituents, $S\upharpoonright Z$ has l, $l \leq k$ nonempty constituents. Consider set $FrZ = \bar{\Lambda}Z - \underline{\Lambda}Z$. Since it is nonempty and describable it is the union of at least one constituent. Assume it is union of m constituents. Now $B(S \upharpoonright Z)$ has l constituents and $B(S \upharpoonright (X_S \setminus Z))$ has $k-l+m$ constituents. Thus the product $B(S \upharpoonright Z) \times B(S \upharpoonright (X_S \setminus Z))$ has $k+m$ generators and so is not isomorphic with $B(S)$ since the latter has k generators.

Theorem 5: Let $Z \subsetneq X_S$. Then

$$B(S) \cong B(S \upharpoonright Z) \times B(S \upharpoonright \underline{A}(X_S \setminus Z))$$

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

Let us call <u>definable restriction</u> of S every subsystem $S \upharpoonright Z$ where $Z \in B(S)$.
The following is useful:

Proposition 6: There exists the largest selective definable restriction of S .

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

The following important property of definable restrictions holds:

Proposition 7:
Let S_o be a definable restriction of S_2 and let S_1 be $S_2 \upharpoonright (X_{S_2} \setminus X_{S_o})$ then for every $Z \subseteq X_{S_2}$ we have

$$\bar{A}_{S_2} Z = \bar{A}_{S_o}(Z \cap X_{S_o}) \cup \bar{A}_{S_1}(Z \cap X_{S_1})$$

The proof of this fact follows from the proposition 1.

Now let S_o be the largest selective definable restriction of S and let S_1 be $S\upharpoonright(X_S \setminus X_{S_o})$. According to the definition of S_o and the proposition 1 every constituent of S_1 consists of at least two elements. Moreover $B(S) \cong B(S_o) \times B(S_1)$. The system S_1 is called <u>totally nonselective</u>. Our remarks boil down to the following:

Proposition 8: (i) There is a unique decomposition of the system S into selective definable restriction S' and totally nonselective definable restriction S''.

(ii) In the above situation we have, for every $Z \subseteq X_S$

$$\bar{A}_S Z = (Z \cap X_{S'}) \cup \bar{A}_{S''}(Z \cap X_{S''}),$$

(iii) Similarly

$$\underline{A}_S Z = (Z \cap X_{S'}) \cup \underline{A}_{S''}(Z \cap X_{S''}).$$

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

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

It follows from the proposition 8, that the operation $\bar{A}$ and $\underline{A}$ are of interest only for the totally nonselective systems /which just means that Hausdorff part of the <u>corresponding topological space is empty/.</u>

Proposition 9: If S is totally nonselective then there exist sets Z and T with the following properties

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

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

(iii) $\bar{A}Z = \bar{A}T = X_S$

Proof: Choose Z to be a selector of constituents of S and T its complement.

Theorem 10: Let S be totally nonselective and C,D

describable subsets of X_S , moreover $C \subseteq D$. Then there exists $Z \subseteq X$ such that $\underline{A}Z = C$ and $\bar{A}Z = D$.

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 $\bar{A}(C \cup T) = D$ whereas $\underline{A}(C \cup T) = C$.

The result of the theorem 10 is used to characterize the algebra $P(X_S)/\approx$. Since $Z_1 \approx Z_2 \Leftrightarrow \underline{A}Z_1 = \underline{A}Z_2$ & $\bar{A}Z_1 = \bar{A}Z_2$ 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 $\underline{A}Z = C$ and $\bar{A}Z = D$/. 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 $\underline{A}Z$ /i.e. those x's which $\underline{\in}$ -belong to Z/ are in Z with value 1, the elements of $X \setminus \bar{A}Z$ /i.e. those x's which $\bar{\in}$ -do not belong to Z/ are in Z with value 0 whereas the elements of $Fr(Z)$ belong to Z with value 1/2 since they are undistinguishable /from the point of view of S / from some elements which do not belong to Z .

Let us finally note that the algebras $P(X)/\simeq$ and $P(X)/\approx$ are isomorphic to $B(S)$.

A c k n o w l e d g m e n t   Thanks are due to doc. C.Rauszer for valuable remarks and comments.

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, International Journal of Computer Sciences Vol. 11, No 5 /1982/ pp. 341-356

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. P.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