Bulletin of the Section of Logic
Volume 12/4 (1983), pp. 194–199
reedition 2008 [original edition, pp. 194–201]

Zdzisław Pawlak

PRINCIPLES OF KNOWLEDGE REPRESENTATION

1. Introduction

In many computer applications we face the following situation: we are given set of objects which are characterized by means of some features like: temperature, height, weight etc., and we want to define in terms of those features some concepts. For example we have a data file concerning patients suffering from a certain disease. State of each patient is characterized by some symptoms like, for example: blood pressure, temperature etc. The question arises whether wa are able to define this disease in terms of available symptoms. Notice that we identify here concept of a specific disease with a set of patients suffering from this disease. Hence the definition of a disease depends upon specific example of a disease under consideration. This approach refers to inductive reasoning, when a concept is built upon the basis of a finite number of examples of this concept.

More detailed discussion of problems considered here one – can find in Pawlak (1982), Pawlak (1983) and Orłowska and Pawlak (1983b).

2. Knowledge representation system

By a knowledge representation system we mean a $S = (U, A, V, \varrho)$ system where

U – is a finite set of *objects*
A – is a finite set of *attributes*
$V = \bigcup V_a$ – is a finite set of values of attributes;

V_a is referred to as a *domain* of attribute a, $a \in A$.
$\varrho : U \times A \longrightarrow V$ – is an *information function*.

We shall use also function $\varrho_x : A \longrightarrow V$ such that $\varrho_x(a) = \varrho(x, a)$ for every $a \in A, x \in X$. We will call ϱ_x the *information* about x in S.

3. Indiscernibility relation

Given a knowledge representation system $S = (U, A, V, \varrho)$, objects $x, y \in U$ and subset $B \subset A$ of attributes: we introduce binary relation between objects as follows:

$x \sim_B y$ iff $\varrho_x(a) = \varrho_y(a)$

for every $a \in B$.

If $x \sim_B y$ we say that x and y are *indiscernible* by B is S. Obviously $\sim_B$ is an equivalence relation for every $B \subset A$. Equivalence class of the relation B containing element $x \in U$ will be denoted by $[x]_{\sim B}$.

Any union of equivalence classes of relation B (and the empty set) will be called B-*definable* set in S.

Let $S = (U, A, V, \varrho)$ and $S' = (U, A', V', \varrho')$ be two knowledge representation systems, with the same set of objects U.

We say that S is *finer* than S ($S < S'$) iff $\tilde{A} \subsetneq \tilde{A}'$: system S and S' are *equivalent* iff $A = A'$.

4. Dependency of attributes

Suppose we are given knowledge representation system $S = (U, A, V, \varrho)$, and two subsets $B, C \subset A$ of attributes.

We say that B *depends* upon C ($C \rightarrow B$) iff $\tilde{B} \supset \tilde{C}$; in particular $a \rightarrow b$, $a, b \in A$ is to mean that attribute b depends upon attribute a, i.e. if we know the value of attribute a we can also compute the value of attribute b.

Let $S = (U, A, V, \varrho)$ be a knowledge representation system. We say that the set of attributes A is *independent* in S iff for every $B \subset A$, $\tilde{B} \supset \tilde{A}$; if there exists subset $B \subset A$ such that $\tilde{B} = (\tilde{A})$ we say that set A of attributes is *dependent* in S.

If A is dependent set of attributes in S, then the least set $B \subset A$ such that $B = A$ will be called *reduct* of A in S. Thus reduct of A is the least subset of attributes of A, such that systems $S = (U, A, V, \varrho)$ and $S' = (U, B, V', \varrho')$ are equivalent. $(\varrho' = \varrho/U \times B)$.

Every system $S = (U, A, V, \varrho)$ with independent set of attributes will be called *reduct* knowledge representation system; otherwise system is *not reduced.*

Thus from every reduced system we may remove some attributes without loss of information in the system.

Of course knowledge representation system can have more than one reduct!

5. Approximation of sets

Suppose we are given a knowledge representation system $S = (U, A, V, \varrho)$ and subset $X \subset U$ of objects.

Lower approximation of X with respect to B' in S we call set

$\underline{B}(X) = \{x \in U : [x]_{\sim B} \subset X\}$.

Upper approximation o X with respect to B in S, we call set

$\overline{B}(X) = \{x \in U : [x]_{\sim B} \cap X \neq 0\}$.

Thus lower approximation of X with respect to B is the greatest B-definable set in S, contained in set X, and upper approximation of X with respect to B is the least B-definable set in S containing set X. Set $Fr_B(X) = \overline{B}(X) - \underline{B}(X)$ will be called *boundary* of X with respect to B in S.

The following property is obvious: set $X \subset U$ is B-definable in S iff $\underline{B}(X) = \overline{B}(X)$. That is to mean that only B-definable sets in S can be defined by means of attributes available in the system.

If $\underline{B}(X) \neq \overline{B}(X)$ then set X is not definable by B in S.

We can introduce the following classes of not definable sets in S.

1) If $\underline{B}(X) = \emptyset$, X is called *internally* not definable by B in S.
2) If $\overline{B}(X) = U$, X is called *externally* not definable by B in S.
3) If X is both internally and externally not definable by B in S, X will be called *totally* not definable by B in S.

4) If $\overline{B}(X) = U$ and $\underline{B}(X) = \emptyset$ we say that X is *roughly* definable by B in S.

Thus if X is internally not definable by B, it means that we are unable to decide by means of features B whether any object is for sure member of set X: if set X is externally not definable by B we are unable to exclude by means of features B any object $x \in U$ being an element of X: if set X is totally not definable by B in S, that is to mean that we can not for sure by means of features B decide whether object x belongs to set X or not and we can not exclude any object x being an element of X.

If set X is roughly definable by B in S it means that for some objects we can decide that they belong or not to set X examining their properties expressed by attributes from B.

In order to measure the degree in "which set X can be defined" by set of attributes B, we introduce "uncertainty" coefficient $\mu_B(X)$ which to each set X and set of attributes B associates a number from the interval $< 0, 1 >$ in the following way

$$\mu_B(X) = \frac{\mid \underline{B}(X) \mid}{\mid \overline{B}(X) \mid}$$

If $\mu_B(X) = 1$, then X is definable by B: if $\mu_B(X) = 0$ is internally not definable ($\mid X \mid$ – denotes cardinality of set X). The coefficient $\mu_B(X)$ can be also treated as an "accuracy" measure of approximation.

We can also introduce another measure of approximations, namely an interval $\Pi_B(X)$ defined as

$$\Pi_B(X) = < \frac{\mid \underline{B}(X) \mid}{\mid X \mid}, \frac{\mid B(X) \mid}{\mid X \mid} >$$

or

$$\Pi_B^*(X) = < \frac{\mid \underline{B}(X) \mid}{\mid U \mid}, \frac{\mid \overline{B}(X) \mid}{\mid U \mid} >$$

The meaning of these intervals is obvious.

6. Sets decidable by attributes

In this section we shall discuss how subsets of attributes "contribute" to the accuracy of approximation of a given subset X of objects in a system S.

In order to do this we introduce the notion of a set of objects *decidable* by set of attributes B in S.

Let $S = (U, A, V, \rho)$ be a knowledge representation system, $B \subset A$ - subset of attributes and $X \subset U$ subset of objects.

With set X and set of attributes B we associate set X_B defined as follows:

$X_B = Fr_{A-B}(X) - Fr_A(X)$

Set X_B is set of objects which membership to X is decided by set of attributes B.

Set X_B can be split into two sets X_B^+ and X_B^- where

$$X_B^+ = \underline{A}(X) - \underline{A-B}(X)$$

$$X_B^- = \overline{A-B}(X) - \overline{A}(X)$$

Set X_B^+ is set of those objects which are positively decided being member of X and set X_B^- is set of those objects which are negatively decided being members of X.

Set of attributes B is superfluous for set of objects $X \subset U$ in S iff $X_B = \emptyset$.

That means that set B of attributes is not necessary for defining set of objects X U, by means of features available in S.

--- . ---

Problems considered in this article also have logical flavour, however, we shall not discuss them here. They are discussed in detail in Orłowska (1981), Orłowska (1982 b), Orłowska (1982 c) and Orłowska and Pawlak (1983 b).

References

Orłowska E. (1981), *Logic approach to information systems*, **ICS PAS Reports** 437.

Orłowska E. (1982 a), *Dynamic information systems*, **Fundamenta Informaticae**, V. 1, pp. 101–118.

Orłowska E. (1982 b), *Languages of approximate information*. **ICS PAS Report** 479.

Orłowska E., Pawlak Z. (1981 b). *Expressive power of knowledge representation systems*, **Internal Journal of Man-Machine Studies** (to appear). **ICS PAS Report** 432.

Orłowska E. (1982 c). *Semantics of vague concepts*. **ICS PAS Reports** 469.

Orłowska E., Pawlak Z. (1981). *Representation of non-deterministic information*, **Theoretical Computer Science** (to appear). **ICS PAS Reports** 450.

Orłowska E. (1983 a), *Representation of vague information*, **ICS PAS Reports** 503.

Orłowska E., Pawlak Z. (1983 b). **Foundation of knowledge representation**. submitted to Springer-Verlag.

Pawlak Z. (1982), *Rough sets*. **International Journal of Computer and Information Science**, No. 11, pp. 341–350.

Pawlak Z. (1983), *Rough classification*, **International Journal of Man-Machine Studies** (to appear).

Polish Academy of Sciences
Institute of the Foundations
of Computer Science