BULLETIN OF THE POLISH
ACADEMY OF SCIENCES
TECHNICAL SCIENCES
Vol. 33, No. 9–10, 1985

VARIA
COMPUTER SCIENCE

Rough Concept Analysis

by

Zdzisław PAWLAK

Presented by Z. PAWLAK on April 25, 1985

Summary. Wille [5] has proposed the set theoretical approach to the concept analysis. In this article there is shown that his ideas can be easily formulated and generalized within the rough set theory (see [3, 4]). In particular the notion of a rough (vague) concept is defined and we focus attention on this notion.

1. Introduction. Wille [5] defines a concept in a context as a pair which consists of two sets, set of objects and set of features of objects and such that the set of features uniquelly defines the set of object considered.

The context may be regarded as a special case of information system (see [2]) and then the concept may be understood as a definable set with respect to some subset of attributes. Thus, definable sets are generalization of concepts.

The question arises: what are the nondefinable sets? Certainly the nondefinable sets may be regarded as rough (vague) concepts, i.e. concepts meaning of which cannot be defined precisely by a given set of attributes. We will consider this problem in detail below.

2. Preliminaries. An information system is the 4-tuple

$$S = (U, A, V, f)$$

where

U – is the **universe**

A – is the set of **attributes**

$V = \bigcup_{a \in A} V_a$ – is the **domain** of a

$f: U \times A \to V$ – is an **information function** (total)

Any subset X of U is called **concept** in S, and any subset B of A is called **content** in S.

With each $B \subseteq A$ we associate the **indiscernibility relation** $\tilde{B}$ in U defined thus

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

If $(x, y) \in \tilde{B}$ we say that x and y are **indiscernible** by B in S.

Certainly $\tilde{B}$ is an equivalence relation in U for every $B \subseteq A$. Equivalence classes of B are called B-**elementary** sets in S.

For any $B \subseteq A$ and $X \subseteq U$ we define two sets

$$\underline{B}X = \{x \in U: [x]_{\tilde{B}} \subseteq X\}$$

$$\bar{B}X = \{x \in U: [x]_{\tilde{B}} \cap X \neq \emptyset\}$$

called B-**lower** and B-**upper** approximation of X in S.

The set $Bn_B(X) = \bar{B}X - \underline{B}X$ is called the B-**boundary** of X in S.

The following definitions are employed

A1) $\underline{B}X$ is the B-**positive region** of X in S

A2) $Bn_B(X)$ is the B-**doubtful region** of X in S

A3) $U - \bar{B}X$ is B-**negative region** of X in S.

3. Rough concepts. A concept $X \subseteq U$ is **precise** with respect to content B, or B is **precise content** of X in S if $\bar{B}X = \underline{B}X$. If $\bar{B}X \neq \underline{B}X$ the concept X is **rough (vague)** with respect to B or B is the **rough content** of X in S.

Thus, precise concepts are definable sets in S and rough concepts are nondefinable sets in S (see [3]).

The number

$$\mu_B(X) = \frac{\text{card } \underline{B}X}{\text{card } \bar{B}X}$$

is called the **accuracy** of X with respect to B in S, and the number

$$\eta_B(X) = 1 - \mu_B(X)$$

is called the **roughness** of X with respect to B in S.

4. Classification of rough concepts. Rough concepts can be classified as follows

B1) If $\underline{B}X \neq \emptyset$ and $\bar{B}X \neq U$ then the concept X is **roughly definable** by content B in S

B2) If $\underline{B}X \neq \emptyset$ and $\bar{B}X = U$ then the concept X is **externally nondefinable** by content B in S

B3) If $\underline{B}X = \emptyset$ and $\bar{B}X \neq U$ then the concept X **internally is nondefinable** by content B in S

B4) If $\underline{B}X = \emptyset$ and $\bar{B}X = U$ then the concept X is **totally nondefinable** by content B in S.

Let us notice that

C1) If X is precise (roughly definable, totally nondefinable), so is $-X$

C2) If X is internally (externally) nondefinable then $-X$ is externally (internally) nondefinable.

Concepts X and Y are **surely disjojnt** with respect to B in S if $\bar{B}X \cap \bar{B}Y = \emptyset$; concepts X and Y are **possibly disjojnt** with respect to B in S if $\underline{B}X \cap \underline{B}Y = \emptyset$.

More complicated relation betwen concepts can be defined on the basis of ideas presented in [1]. The meaning of the above definitions is obvious.

5. Rough equality of concepts. Let X and Y be two concepts in S and let B be a content in S. We introduce the following definitions.

D1) $X \eqsim_B Y$ if $\underline{B}X = \underline{B}Y$ (X and Y are **bottom equal** with respect to B in S)

D2) $X \simeq_B Y$ if $\bar{B}X = \bar{B}Y$ (X and Y are **top equal** with respect to B in S)

D3) $X \approx_B Y$ if $X \eqsim_B Y$ and $X \simeq_B Y$ (X and Y are **roughly equal** with respect to B in S).

Bottom equality of concepts preserves the positive regions of concepts; top equality of concepts preserves negative regions of concepts and rough equality of concepts preserves the doubtful region of concepts.

It is easy to see that all the above introduced relations are equivalence relations.

6. Rough inclusion of concepts. Concepts can be roughly ordered by a rough inclusion of concepts defined as below.

E1) $X \underset{\sim}{\subset}_B Y$ if $\underline{B}X \subseteq \underline{B}Y$ (X is **roughly bottom included** in Y with respect to B in S)

E2) $X \tilde{\subset}_B Y$ if $\bar{B}X \subseteq \bar{B}Y$ (X is **roughly top included** in Y with respect to B in S)

E3) $X \underset{\sim}{\tilde{\subset}}_B Y$ if $X \underset{\sim}{\subset}_B Y$ and $X \tilde{\subset}_B Y$ (X is **roughly included** in Y with respect to B in S).

The meaning of the above definitions is also obvious. One can also define other kinds of hierarhies between objects, for example: X is finer then Y with respect to B in S, in symbols $X <_B Y$, if $\underline{B}X \supset \underline{B}Y$ and $\bar{B}X \subset \bar{B}Y$.

7. Algebra of concepts. To this end, let us remark that performing boolean operations on concepts, we get the possibilities listed in the table below.

X	Y	$X \cup Y$	$X \cap Y$	$-x$
p	p	p	p	p
r	p	r or p	r or p	r
p	r	r or p	r or p	p
r	r	r or p	r or p	r

where p stands for precise and r for rough.

These properties cause some dificulties when defining operations on rough concepts and we shall disscuss this problem in some details in another paper.

INSTITUTE OF COMPUTER SCIENCE, POLISH ACADEMY OF SCIENCES, PKIN, PO BOX 22, 00-901 WARSAW
(INSTYTUT PODSTAW INFORMATYKI PAN)
UNIVERSITY OF NORTH CAROLINA, DEPARTMENT OF COMPUTER SCIENCE, NC 28223 USA

REFERENCES

[1] G. Epstein, *Decisive venn diagrams*, Int Symp. on Multiple-Valued Logic, Logan, Utah 1976.

[2] Z. Pawlak, *Information systems: mathematical foundations*, Inf. Syst., **6** (1981) 205–218.

[3] Z. Pawlak, *Rough sets*, Int. J. Inf. Comp. Sci., **11** (1982) 341–356.

[4] Z. Pawlak, *Rough classification*, Int. J. Man-Machine Studies, **20** (1984) 469–483.

[5] R. Wille, *Liniendiagramme hierarhischer begriffssysteme*, Technische Hochschule, Preprint No. 812, Darmstadt 1984.

З. Павляк, **Анализ приближенного понятия**

Р. Вилле предложил применение теоретико-множественного подхода к анализу понятий. В этой работе доказывается, что его идеи можно легко сформулировать и обобщить в рамах теории приближенных множеств. В особенности, дается определение „приближенного понятия" и именно оно является предметом подробного исследования.