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

*VARIA*

*COMPUTER SCIENCE*

# On Rough Dependency of Attributes in Information Systems

by

Zdzisław PAWLAK

*Presented by Z. PAWLAK on April 25, 1985*

**Summary.** In this note the concept of the rough dependency of attributes in an information system is introduced and it is shown that this concept is equivalent to that of approximation of sets.

**1. Introduction.** In this note we introduce the concept of a rough dependency of attributes, which can be viewed as a generalization of attribute dependency, considered previously in connection with information systems and rough sets theory (see [1]).

**2. Information system.** By an information system we mean the 4-tuple

$$S = (U, A, V, \varrho)$$

where

$U$ – is a finite set of **objects**, called the **universe**

$A$ – is a finite set of **attributes**

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

$\varrho : U \times A \rightarrow V$ – is a total function, such that $\varrho(x, a) \in V_a$ for every $a \in A$ and $x \in U$ – called the **information function**.

The function $\varrho_x : A \rightarrow V$ such that $\varrho(a) = \varrho(x, a)$ for every $a \in A$ and $x \in U$ will be called **information about** $x$ in $S$.

We say that objects $x, y \in U$ are **indiscernible** with respect to the subset of attributes $B \subseteq A$ in $S$ $(x \underset{B}{\sim} y)$ if $\varrho_x(a) = \varrho_y(a)$ for every $a \in B$.

Obviously $\widetilde{B}$ is an equivalence relation in $U$ for any $B \subseteq A$. The equivalence classes of the relation $\widetilde{B}$ are called **$B$-elementary** sets in $S$. The partition generated by the equivalence relation $\widetilde{B}$ is denoted by $B^*$.

Let $S = (U, A, V, \varrho)$ be an information system. By the **$B$-representation** of $S$ we mean the information system defined thus:

$$S_B = (B^*, B, V_B, \varrho_B)$$

where

$B^*$ – is the family of all equivalence classes of the relation $\widetilde{B}$

$V_B = \bigcup_{a \in B} V_a$

$\varrho_B : B^* \times B \to V_B$, and $\varrho_B([x]_{\widetilde{B}}, a) = \varrho(x, a)$, for every $x \in U, a \in A$, ($[x]_{\widetilde{B}}$ – denotes an equivalence class containing the object $x$).

If $S = (U, A, V, \varrho)$ is an information system and $X \subseteq U$ then by the **$X$-restriction** of $S$ we mean the information system defined as follows:

$$S/X = (X, A, V', \varrho')$$

where

$V' = \bigcup_{a \in A} V_{a,X}$, and $V_{a,X} = \{v \in V_a : \text{there exists } x \in X \text{ such that } \varrho(x, a) = v\}$.

**3. Approximations of sets and families of sets.** Let $S = (U, A, V, \varrho)$ be an information system, $B \subseteq A$ and $X \subseteq U$.

The **$B$-lower** and **$B$-upper approximation** of $X$ in $S$ are sets defined thus:

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

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

The set

$$Bn_B(X) = \bar{B}X - \underline{B}X$$

is called the **$B$-boundary** of $X$ in $S$.

If $\underline{B}X = \bar{B}X$ we say that the set $X$ is **$B$-definable** in $S$, otherwise the set $X$ is **$B$-nondefinable**.

Let $\mathscr{X} = \{X_1, X_2, \ldots, X_n\}$, $X_i \subseteq U$ be a finite family of subset of $U$. By **$B$-lower** and **$B$-upper approximation** of $\mathscr{X}$ in $S$ we mean sets

$$\underline{B}\mathscr{X} = \{\underline{B}X_1, \underline{B}X_2, \ldots, \underline{B}X_n\}$$

$$\bar{B}\mathscr{X} = \{\bar{B}X_1, \bar{B}X_2, \ldots, \bar{B}X_n\}$$

respectively.

In what follows we assume that $\mathscr{X}$ is classification (partition) of $U$, i.e. $X_i \cap X_j = \emptyset$ and $\bigcup_{i=1}^{n} X_i = U$ for $i \neq j$ and $0 \leqslant i, j \leqslant n$.

If $\underline{B}\mathscr{X} = \bar{B}\mathscr{X}$ we say that the classification if $B$-**definable** in $S$; otherwise the classification is $B$-**nondefinable** in $S$.

Now let us introduce the two following notions:

1) $\mathrm{Pos}_B(\mathscr{X}) = \bigcup_{i=1}^{n} \underline{B}X_i$ – the $B$-**positive region** of $\mathscr{X}$ in $S$,

2) $Bn_B(\mathscr{X}) = \bigcup_{i=n}^{n} Bn_B X_i$ – the $B$-**doubtful region** of $\mathscr{X}$ in $S$.

Certainly

$$\mathrm{Pos}_B(\mathscr{X}) \dot{\cup} Bn_B(\mathscr{X}) = U.$$

The $B$-positive region of the classification $\mathscr{X}$ is the subset of objects from the universe $U$, which can be positively classified (i.e. uniquely assigned to one class of the classification $\mathscr{X}$) employing all attributes from $B$. The $B$-doubtful region of the classification $\mathscr{X}$ is the set of objects which cannot be classified using attributes from $B$.

The number

$$\gamma_B(\mathscr{X}) = \frac{\mathrm{card}\ \mathrm{Pos}_B(\mathscr{X})}{\mathrm{card}\ U}$$

will be reffered to as a **quality of the approximation** of $\mathscr{X}$ by $B$ in $S$.

Of course

$$0 \leqslant \gamma_B(\mathscr{X}) \leqslant 1.$$

**4. Rough dependency of attributes.** Let $S = (U, A, V, \varrho)$ be an information system and let $B, C \subseteq A$.

We say that $C$ **depends in degree** $k$ **(**$k$**-depends) on** $B$ **in** $S$, in symbols $B \xrightarrow[S]{k} C$, or $B \xrightarrow{k} C$ when $S$ is understood, if $k = \gamma_B(C^*)$.

If $k = 1$ we say that $C$ **totally depends** on $B$ in $S$; instead of $B \xrightarrow{1} C$ we shall also write $B \to C$.

If $0 < k < 1$ we say that $C$ **roughly depends** on $B$ in $S$.

If $k = 0$ we say that $C$ is **totally independent** on $B$ in $S$.

The following two properties show the relationship between the rough dependency and approximations.

*Property* 1. The following conditions are equivalent:

1) $B \to C$

2) $B \xrightarrow[S_{S \cup C}]{S} C$

3) $\tilde{B} \subseteq \tilde{C}$

4) $B \dot{\cup} C = \tilde{B}$

5) $\underline{B}(C^*) = \bar{B}(C^*)$

6) $\gamma_B(C^*) = 1$.

*Property* 2. $B \xrightarrow{k} C$ in $S$ if and only if $B \xrightarrow{1} C$ in $S/\mathrm{Pos}_B(C^*)$ and $B \xrightarrow{0} C$ in $S/Bn_B(C^*)$.

**5. Example.** Consider an information system as shown in the Table:

| $U$ | $a$ | $b$ | $c$ |
|---|---|---|---|
| $x_1$ | 1 | 0 | 2 |
| $x_2$ | 0 | 1 | 1 |
| $x_3$ | 2 | 0 | 0 |
| $x_4$ | 1 | 0 | 0 |
| $x_5$ | 1 | 0 | 2 |
| $x_6$ | 2 | 0 | 0 |
| $x_7$ | 0 | 1 | 1 |
| $x_8$ | 1 | 0 | 0 |
| $x_9$ | 1 | 0 | 2 |
| $x_{10}$ | 0 | 1 | 1 |

Let us compute $\{a, b\} \xrightarrow{k} \{c\}$ in this system. Denote $\{a, b\} = B$. Obviously

$$\{c\}^* = \{X_1, X_2, X_3\}$$

where

$$X_1 = \{x_1, x_5, x_9\}$$
$$X_2 = \{x_2, x_7, x_{10}\}$$
$$X_3 = \{x_3, x_4, x_6, x_8\}$$

and

$$B^* = \{Y_1, Y_2, Y_3\}$$

where

$$Y_1 = \{x_1, x_4, x_5, x_7, x_8, x_9\}$$
$$Y_2 = \{x_2, x_7, x_{10}\}$$
$$Y_3 = \{x_3, x_6\}.$$

Hence

$$\underline{B}X_1 = \emptyset$$
$$\underline{B}X_2 = Y_2$$
$$\underline{B}X_3 = Y_3$$

and

$$\mathrm{Pos}_B(\{c\}^*) = \underline{B}X_1 \dot{\cup} \underline{B}X_2 \dot{\cup} \underline{B}X_3 = Y_2 \cup Y_3.$$

Thus

$$\gamma_B(\{c\}^*) = 5/10 = 0.5 = k.$$

It means that for five elements only $(x_2, x_3, x_6, x_7, x_{10})$ the total dependency $\{a, b\} \rightarrow \{c\}$ holds, i.e. the value of the attribute $c$ can be uniquely determined when values of attributes $a$ and $b$ are known.

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

REFERENCES

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

З. Павляк, **Приближенная зависимость характерных признаков в информационных системах**

В настоящей работе вводится понятие приближенной зависимости характерных признаков в информационной системе, а также доказывается, что это понятие эквивалентно понятию приближения множеств.