BULLETIN OF THE POLISH
ACADEMY OF SCIENCES
MATHEMATICS
Vol. 36, No. 7-8, 1988

THEORETICAL COMPUTER SCIENCE

Partial Dependency of Attributes

by

Miroslav NOVOTNÝ and Zdzisław PAWLAK

Presented by Z. PAWLAK on December 8, 1987

Summary. Any set Z of attributes in an information system defines an equivalence $\tilde{Z}$ on the set X of all objects; the blocks of $\tilde{Z}$ are sets of objects that are indiscernible by means of attributes in Z. If Z, T are sets of attributes and if any block of $\tilde{Z}$ is a subset of a block of $\tilde{T}$, the set T is said to be totally dependent on Z. In the general case, only some blocks of $\tilde{Z}$ are subsets of blocks of $\tilde{T}$. Union of blocks of $\tilde{Z}$ such that any of them is a subset of a block of $\tilde{T}$ is a subset of X; the ratio of the cardinality of this set to the cardinality of X expresses the dependency degree of T on Z. The present paper includes some results concerning this dependency degree. A distance function is defined on the basis of dependency degree.

1. Introduction. Let (X, A, V, f) be an information system (see [1–3]), i.e. X, A, V are finite nonempty sets and f is a mapping of $X \times A$ into V. Elements in X are interpreted to be objects, elements in A–attributes, and elements in V are considered to be values of attributes; $f(x, a) = v$ means that the attribute a has the value v for the object x. For any $Z \subseteq A$, we put $\tilde{Z} = \{(x, y) \in X \times X;\ f(x, a) = f(y, a) \text{ for any } a \in Z\}$.

If $(x, y) \in \tilde{Z}$ then the objects x, y are indiscernible by means of attributes in Z. If $Z \subseteq A$, $T \subseteq A$ and $\tilde{Z} \subseteq \tilde{T}$, then any objects indiscernible by means of attributes in Z are indiscernible by means of attributes in T, i.e. the ability of Z to discern objects is greater or as great as the same ability of T. The set T is said to be dependent on Z. We can express it by saying that any indiscernibility block of $\tilde{Z}$ is a subset of an indiscernibility block of $\tilde{T}$.

Such a situation occurs only exceptionally. In a general case, only some indiscernibility blocks of $\tilde{Z}$ are subsets of indiscernibility blocks of $\tilde{T}$. Union of indiscernibility blocks of $\tilde{Z}$ such that any of them is included in an indiscernibility block of $\tilde{T}$ is a subset of X; the ratio of the cardinality of this set to the cardinality of X expresses the dependency degree of T on Z.

The present paper includes analysis of this partial dependency of attributes. This analysis is based on some operations and relations on the set of all systems of nonempty subsets of a fixed set U. Some theorems concerning dependency degree can be considered to be the main results of the paper. An example completes the investigation. A distance function is defined on the basis of dependency degree.

2. Operations with set systems. Let U be a set. We denote by $\mathfrak{M}$–the set whose elements are all systems of nonempty subsets of the set U;

$\mathfrak{D}$–the set whose elements are all disjoint systems of nonempty subsets of the set U;

$\mathfrak{R}$–the set whose elements are all decompositions of the set U.

Clearly, $\mathfrak{R} \subseteq \mathfrak{D} \subseteq \mathfrak{M}$.

We define an operation $*$ on the set $\mathfrak{M}$ as follows: for any $P \in \mathfrak{M}$ and any $Q \in \mathfrak{M}$, we put $P*Q = \{p \in P;$ there exists $q \in Q$ with $p \subseteq q\}$. Clearly, $P*Q \in \mathfrak{M}$.

2.1. Example. Put $U = \{p, q, r\}$ where $p \neq q \neq r \neq p$, $P = \{\{p\}\}$, $Q = \{\{p, q\}\}$, $R = \{\{p, r\}\}$. Then $P*Q = P$, $P*R = P$ and, therefore, $(P*Q)*R = P$. On the other hand, $Q*R = \emptyset$, $P*\emptyset = \emptyset$ and, hence, $P*(Q*R) = \emptyset$. It follows that the operation $*$ is not associative. □

2.2. Example. It is easy to see that $P*Q \subseteq P$ for any $P \in \mathfrak{M}$ and any $Q \in \mathfrak{M}$. □

For any $P \in \mathfrak{M}$ and any $Q \in \mathfrak{M}$, we put $P \leqslant Q$ if and only if $P*Q = P$. By 2.2 we obtain

2.3. Lemma. *For any $P \in \mathfrak{M}$ and any $Q \in \mathfrak{M}$ the conditions $P \leqslant Q$ and $P \subseteq P*Q$ are equivalent.* □

2.4. Lemma. *If P, Q, R are in $\mathfrak{M}$ and $Q \leqslant R$ holds, then $P*Q \subseteq P*R$.*

Proof. If $t \in P*Q$, then $t \in P$ and there exists $q \in Q$ such that $t \subseteq q$. Since $Q = Q*R$, there exists $r \in R$ such that $q \subseteq r$ and, hence, $t \subseteq r$. Thus, $t \in P*R$.

For any $P \in \mathfrak{M}$ and any $Q \in \mathfrak{M}$, we set $P \wedge Q = \{p \cap q;\, p \in P, q \in Q, p \cap q \neq \emptyset\}$. Clearly, $P \wedge Q \in \mathfrak{M}$. □

2.5. Theorem. *Relation $\leqslant$ is an ordering on $\mathfrak{D}$ and $\inf\{P, Q\} = P \wedge Q$ holds for any $P \in \mathfrak{D}$ and any $Q \in \mathfrak{D}$.*

Proof. Reflexivity of $\leqslant$ on $\mathfrak{D}$ is obvious. If P, Q, R are in $\mathfrak{D}$ and $P \leqslant Q$, $Q \leqslant R$ hold, then $P = P*Q \subseteq P*R$ by 2.4 which implies $P \leqslant R$ by 2.3. Thus, $\leqslant$ is transitive.

If $P \leqslant Q$, $Q \leqslant P$ hold, then $P \subseteq P*Q$, $Q \subseteq Q*P$ by 2.3. Thus, to any $p \in P$ there exists $q \in Q$ such that $p \subseteq q$ and to $q \in Q$ there exists $p' \in P$ such that $q \subseteq p'$. Thus, $p \subseteq q \subseteq p'$. Since p, q, p' are nonempty and $P \in \mathfrak{D}$, we obtain $p = p'$ which implies $p = q$. We have proved $P \subseteq Q$. Similarly, we prove $Q \subseteq P$ which implies that $P = Q$. Therefore, $\leqslant$ is an antisymmetric relation.

For any $P \in \mathfrak{D}$ and any $Q \in \mathfrak{D}$, we have $P \wedge Q \subseteq (P \wedge Q)*P$ because $p \cap q \subseteq p$ holds for any $p \in P$ and any $q \in Q$. Thus, $P \wedge Q \leqslant P$ by 2.3. Similarly $P \wedge Q \leqslant Q$ holds and, hence, $P \wedge Q$ is a lower bound of the set $\{P, Q\}$. If $R \leqslant P$ and $R \leqslant Q$

hold, then $R \subseteq R*P$, $R \subseteq R*Q$ hold by Lemma 2.3. Thus, for any $r \in R$, there are $p \in P$ and $q \in Q$ such that $r \subseteq p$, $r \subseteq q$ which implies that $r \subseteq p \cap q \in P \wedge Q$. Thus, $R \subseteq R*(P \wedge Q)$ which implies that $R \leqslant P \wedge Q$ by Lemma 2.3. It follows that $P \wedge Q$ is the greatest lower bound of the set $\{P, Q\}$. □

2.6. Corollary. *If P, Q, R are in $\mathfrak{D}$, then the following assertions hold*:

$$\text{(i)}\quad P*(Q \wedge R) \subseteq P*Q,$$

$$\text{(ii)}\quad Q \wedge (R*P) \subseteq (Q \wedge R)*P.$$

Proof. By 2.5 we have $Q \wedge R \leqslant Q$ and (i) follows by 2.4.

If $t \in Q \wedge (R*P)$, there exist $q \in Q$ and $r \in R*P$ such that $t = q \cap r$. Furthermore, $r \in R$ and there exists $p \in P$ such that $r \subseteq p$. It follows that $t \in Q \wedge R$ and $t \subseteq r \subseteq p$ which means $t \in (Q \wedge R)*P$ and (ii) holds. □

If α, β are equivalences on U, then $\alpha \cap \beta$ is an equivalence on U as well. It follows that U/α, U/β, $U/\alpha \cap \beta$ are in $\mathfrak{R}$. Then following Lemma is obvious:

2.7. Lemma. *If α, β are equivalences on U, then $U/\alpha \cap \beta = (U/\alpha) \wedge (U/\beta)$.* □

3. Attributes and their properties. Let (X, A, V, f) be an information system. In the Introduction we have defined the equivalence relation $\tilde{Z}$ on the set X to any set $Z \subseteq A$. Thus, $\sim$ is a mapping assigning an equivalence relation on the set of all objects to any set of attributes. We now present some properties of this mapping.

The following property is well-known:

3.1. Lemma. *For any $P \subseteq A$ and any $Q \subseteq A$, we have $\widetilde{P \cup Q} = \tilde{P} \cap \tilde{Q}$.* □

3.2. Theorem. *If P, Q, R are subsets of A, then the following assertions hold*:

$$\text{(i)}\quad X/\tilde{P}*X/\widetilde{Q \cup R} \subseteq X/\tilde{P}*X/\tilde{Q},$$

$$\text{(ii)}\quad X/\tilde{Q} \wedge (X/\tilde{R}*X/\tilde{P}) \subseteq X/\widetilde{Q \cup R}*X/\tilde{P}.$$

Proof. We have $X/\tilde{P}*X/\widetilde{Q \cup R} = X/\tilde{P}*X/\tilde{Q} \cap \tilde{R} = X/\tilde{P}*(X/\tilde{Q} \wedge X/\tilde{R})$ $\subseteq X/\tilde{P}*X/\tilde{Q}$ by Lemmas 3.1, 2.7, and 2.6(i) which is (i). Furthermore, $X/\tilde{Q} \wedge$ $\wedge (X/\tilde{R}*X/\tilde{P}) \subseteq (X/\tilde{Q} \wedge X/\tilde{R})*X/\tilde{P} = X/(\tilde{Q} \cap \tilde{R})*X/\tilde{P} = X/\widetilde{Q \cup R}*X/\tilde{P}$ by Lemmas 2.6(ii), 2.7, and 3.1 which is (ii). □

4. Partial dependency of attributes. If P is a system of sets, we denote by $\bigcup P$ the union of P, i.e. $\bigcup P = \bigcup \{p;\ p \in P\}$.

4.1. Lemma. *If $Q \in \mathfrak{R}$ and $P \in \mathfrak{D}$, then $\bigcup (Q \wedge P) = \bigcup P$.*

Proof. For any $x \in U$, the condition $x \in \bigcup (Q \wedge P)$ is equivalent with the existence of $p \in P$ such that $x \in p$ because an $q \in Q$ with $x \in q$ always exists. Thus, $x \in \bigcup (Q \wedge P)$ is equivalent with $x \in \bigcup P$ which is our assertion. □

4.2. LEMMA. *If* $P \in \mathfrak{M}$, $Q \in \mathfrak{D}$, $R \in \mathfrak{M}$, *then* $\bigcup(P*Q) \cap \bigcup(Q*R) \subseteq \bigcup(P*R)$.

Proof. If $x \in \bigcup(P*Q) \cap \bigcup(Q*R)$, there are $p \in P$, $q \in Q$, $q' \in Q$, $r \in R$ such that $x \in p \subseteq q$, $x \in q' \subseteq r$. Since $Q \in \mathfrak{D}$, we have $q = q'$ and, thus, $x \in p \subseteq q \subseteq r$ which implies that $x \in \bigcup(P*R)$. □

For any set Y, we denote by card Y the cardinality of Y. Let (X, A, V, f) be an information system and $P \subseteq A$, $Q \subseteq A$. If

$$k = \frac{\operatorname{card}(\bigcup(X/P*X/Q))}{\operatorname{card} X},$$

then we put $P \to^k Q$ and the set Q is said to depend in the degree k on the set P.

4.3. EXAMPLE. $P \to^1 Q$ holds if and only if $\tilde{P} \subseteq \tilde{Q}$. Indeed, $P \to^1 Q$ holds if and only if $X = \bigcup(X/\tilde{P}*X/\tilde{Q})$ which means that for any $p \in X/\tilde{P}$, there exists $q \in X/\tilde{Q}$ with $p \subseteq q$ which is equivalent with $\tilde{P} \subseteq \tilde{Q}$. □

4.4. EXAMPLE. $P \to^0 Q$ holds if and only if $X/\widetilde{P \cup Q} \cap X/\tilde{P} = \varnothing$. Indeed, $p \to^0 Q$ holds if and only if $\bigcup(X/\tilde{P}*X/\tilde{Q}) = \varnothing$ which means that for any $p \in X/\tilde{P}$ there exists $q \in X/\tilde{Q}$ with $p \subseteq q$, i.e. for any $p \in X/\tilde{P}$ and any $q \in X/\tilde{Q}$ the condition $p \cap q \neq p$ holds. This means that $(X/\tilde{P} \wedge X/\tilde{Q}) \cap X/\tilde{P} = \varnothing$; the last condition can be expressed in the form $(X/\tilde{P} \cap \tilde{Q}) \cap X/\tilde{P} = \varnothing$ by 2.7 and in the form $X/\widetilde{P \cup Q} \cap X/\tilde{P} = \varnothing$ by 3.1. □

For any $P \subseteq A$ and any $Q \subseteq A$ there exists exactly one k with $0 \leqslant k \leqslant 1$ such that $P \to^k Q$. If $P \to^1 Q$, the set Q is said to be totally dependent on the set P; if $P \to^0 Q$, the set Q is called totally independent on P.

5. Properties of dependency degrees

5.1. THEOREM. *If* (X, A, V, f) *is an information system and* P, Q, R *are subsets of* A *such that* $P \to^k Q$, $P \to^l R$, $P \to^m Q \cup R$, *then* $m \leqslant \min\{k, l\}$.

Proof. By 3.2(i), we have $\operatorname{card}(\bigcup(X/\tilde{P}*X/\widetilde{Q \cup R})) \leqslant \operatorname{card}(\bigcup(X/\tilde{P}*X/\tilde{Q}))$ which implies that $m \leqslant k$. Similarly, we obtain $m \leqslant l$ which implies the assertion. □

5.2. THEOREM. *If* (X, A, V, f) *is an information system and* P, Q, R *are subsets of* A *such that* $Q \to^k P$, $R \to^l P$, $Q \cup R \to^m P$, *then* $m \geqslant \max\{k, l\}$.

Proof. By lemmas 4.1 and 3.2(ii), we have $\operatorname{card}(\bigcup(X/\tilde{R}*X/\tilde{P})) = \operatorname{card}(\bigcup(X/\tilde{Q}(X/\tilde{R}*X/\tilde{P}))) \leqslant \operatorname{card}(\bigcup(X/\widetilde{Q \cup R}*X/\tilde{P}))$ which implies that $l \leqslant m$. Similarly, we obtain $k \leqslant m$ which implies the assertion. □

5.3. THEOREM. *If* (X, A, V, f) *is an information system and* P, Q, R *are subsets of* A *such that* $P \to^k Q$, $Q \to^l R$, $P \to^m R$, *then* $m \geqslant k+l-1$.

Proof. If B, C are arbitrary sets, then $\operatorname{card} B + \operatorname{card} C = \operatorname{card}(B \cup C) + \operatorname{card}(B \cap C)$. We put $B = \bigcup(X/\tilde{P}*X/\tilde{Q})$, $C = \bigcup(X/\tilde{Q}*X/\tilde{R})$. It follows by 4.2 that $B \cap C \subseteq \bigcup(X/\tilde{P}*X/\tilde{R})$. Thus, we obtain that $k+l \leqslant 1+m$ because $B \cup C \subseteq X$. □

6. Example. Let us have $X = \{\text{Burke, Clark, Jameson, Kellog, Newman}\}$, $A = \{\text{sex, age, hair}\}$, $V = \{\text{male, female, young, middle, old, black, blond}\}$. Suppose that f is given by Table 1:

TABLE 1

	Sex	Age	Hair
Burke	male	middle	blond
Clark	male	young	black
Jameson	female	young	blond
Kellog	male	old	black
Newman	female	middle	blond

We put $P = \{\text{age, hair}\}$, $Q = \{\text{age, sex}\}$, $R = \{\text{hair, sex}\}$. We denote any person in X by its initial letter; then

$$X/\tilde{P} = \{\{B, N\}, \{C\}, \{J\}, \{K\}\},$$

$$X/\tilde{Q} = \{\{B\}, \{C\}, \{J\}, \{K\}, \{N\}\},$$

$$X/\tilde{R} = \{\{B\}, \{C, K\}, \{J, N\}\}.$$

Clearly, $X/\tilde{Q} * X/\tilde{P}$, $X/\tilde{Q} * X/\tilde{Q}$, $X/\tilde{Q} * X/\tilde{R}$ are equal to $X/\tilde{Q}$ which implies that $\bigcup(X/\tilde{Q} * X/\tilde{P})$, $\bigcup(X/\tilde{Q} * X/\tilde{Q})$, $\bigcup(X/\tilde{Q} * X/\tilde{R})$ equal X. Thus, $Q \rightarrow^1 P$, $Q \rightarrow^1 Q$, $Q \rightarrow^1 R$. Furthermore, $X/\tilde{P} * X/\tilde{Q} = \{\{C\}, \{J\}, \{K\}\} = X/\tilde{P} * X/\tilde{R}$. It follows that $\bigcup(X/\tilde{P} * X/\tilde{Q}) = \{C, J, K\} = \bigcup(X/\tilde{P} * X/\tilde{R})$. Thus, $P \rightarrow^{0.6} Q$, $P \rightarrow^{0.6} R$, $P \rightarrow^1 P$. Finally, $X/\tilde{R} * X/\tilde{Q} = \{\{B\}\} = X/\tilde{R} * X/\tilde{P}$. Thus, $\bigcup(X/\tilde{R} * X/\tilde{Q}) = \{B\} = \bigcup(X/\tilde{R} * X/\tilde{P})$ and, therefore, $R \rightarrow^{0.2} Q$, $R \rightarrow^{0.2} P$, $R \rightarrow^1 R$. We obtain Table 2:

TABLE 2

k	P	Q	R
P	1.0	0.6	0.6
Q	1.0	1.0	1.0
R	0.2	0.2	1.0

Clearly, $X/\widetilde{Q \cup R} = \{\{B\}, \{C\}, \{J\}, \{K\}, \{N\}\} = X/\tilde{Q}$. It follows that $P \rightarrow^{0.6} Q \cup R$. Since $P \rightarrow^{0.6} Q$, $P \rightarrow^{0.6} R$, the inequality of Theorem 5.1 is satisfied. Similarly, $Q \cup R \rightarrow^1 P$. Since $Q \rightarrow^1 P$, $R \rightarrow^{0.2} P$, the inequality of Theorem 5.2 is satisfied. Finally, $P \rightarrow^{0.6} Q$, $Q \rightarrow^1 R$, $P \rightarrow^{0.6} R$ and the inequality of Theorem 5.3 is satisfied.

7. Distance function. Let (X, A, V, f) be an information system and suppose that $Z \subseteq A$, $T \subseteq A$. If $Z \rightarrow^k T$ holds, we set $\sigma(Z, T) = 1 - k$. Furthermore, we set $\varrho(Z, T) = 1/2(\sigma(Z, T) + \sigma(T, Z))$.

7.1. Theorem. *If (X, A, V, f) is an information system and P, P_1, Q, Q_1, R are arbitrary subsets of A, then the following assertions hold.*

(i) $0 \leqslant \varrho(P, Q) \leqslant 1$,

(ii) $\varrho(P, P) = 0$,

(iii) $\varrho(P, Q) = 0$ *if and only if* $\tilde{P} = \tilde{Q}$,

(iv) $\varrho(P, Q) = \varrho(Q, P)$,

(v) $\varrho(P, R) \leqslant \varrho(P, Q) + \varrho(Q, R)$,

(vi) *If* $\tilde{P} = \tilde{P}_1$, $\tilde{Q} = \tilde{Q}_1$, *then* $\varrho(P, Q) = \varrho(P_1, Q_1)$.

Proof. Since $P \to^k Q$ implies $0 \leqslant k \leqslant 1$, we have $0 \leqslant \sigma(P, Q) \leqslant 1$; similarly, we obtain $0 \leqslant \sigma(Q, P) \leqslant 1$ which implies (i). Furthermore, (ii) follows from the fact that $P \to^1 P$ holds. If $\varrho(P, Q) = 0$ then $\sigma(P, Q) = 0$, $\sigma(Q, P) = 0$ which implies that $P \to^1 Q$, $Q \to^1 P$ which means $\tilde{P} = \tilde{Q}$ by 4.3. On the other hand, $\tilde{P} = \tilde{Q}$ is equivalent with $P \to^1 Q$, $Q \to^1 P$ by 4.3 which implies that $\varrho(P, Q) = 0$. Thus, (iii) holds. Property (iv) follows from the definition of ϱ. Furthermore, if $\sigma(P, Q) = 1-k$, $\sigma(Q, R) = 1-l$, $\sigma(P, R) = 1-m$, then $m \geqslant k+l-1$ by 5.3. It follows that $1-m \leqslant 2-k-l$ which implies that $\sigma(P, R) \leqslant \sigma(P, Q) + \sigma(Q, R)$. Similarly, $\sigma(R, P) \leqslant \sigma(R, Q) + \sigma(Q, P)$; these inequalities imply (v). Finally, if $\tilde{P} = \tilde{P}_1$, $\tilde{Q} = \tilde{Q}_1$, then $\varrho(P_1, P) = 0$, $\varrho(Q, Q_1) = 0$ by (iii) and $\varrho(P_1, Q_1) \leqslant \varrho(P_1, P) + \varrho(P, Q_1) \leqslant \varrho(P_1, P) + \varrho(P, Q) + \varrho(Q, Q_1) = \varrho(P, Q)$ by (v). Similarly, we obtain $\varrho(P, Q) \leqslant \varrho(P_1, Q_1)$ and we have (vi). □

Put $\boldsymbol{B}(A) = \{Z; Z \subseteq A\}$. For any $Z \in \boldsymbol{B}(A)$ and any $T \in \boldsymbol{B}(A)$ we put $Z \equiv T$ if and only if $\tilde{Z} = \tilde{T}$. Then $\equiv$ is an equivalence on $\boldsymbol{B}(A)$. For any $\boldsymbol{Z} \in \boldsymbol{B}(A)/\equiv$ and any $\boldsymbol{T} \in B(A)/\equiv$, we set $\delta(\boldsymbol{Z}, \boldsymbol{T}) = \varrho(Z, T)$ where $Z \in \boldsymbol{Z}$ and $T \in \boldsymbol{T}$ are arbitrary. By Theorem 7.1(vi) this definition is correct and by Theorem 7.1, δ is a distance function on $\boldsymbol{B}(A)/\equiv$.

7.2. Example. If the information system (X, A, V, f) and the sets P, Q, R are the same as in Example 6, we have $\varrho(P, Q) = 0.2$, $\varrho(P, R) = 0.6$, $\varrho(Q, R) = 0.4$. □

MATHEMATICS INSTITUTE, ČSAV, MENDLOVO NAM. 1, 60300 BRNO (CZECHOSLOVAKIA)
DEPARTMENT OF COMPLEX CONTROL SYSTEMS, POLISH ACADEMY OF SCIENCES, BAŁTYCKA 10, 44-100 GLIWICE
(ZAKŁAD SYSTEMÓW AUTOMATYKI KOMPLEKSOWEJ PAN)

REFERENCES

[1] W. Marek, Z. Pawlak, *Information storage and retrieval systems – mathematical foundations*, Theor. Comp. Sci., **1** (1976), 331–354.

[2] M. Novotný, Z. Pawlak, *On a representation of rough sets by means of information systems*, Fundam. Inform., **6** (1983), 286–296.

[3] Z. Pawlak, *Information systems – theoretical foundations*, Inform. Syst., **6** (1981), 205–218.

[4] Z. Pawlak, *On rough dependency of attributes in information systems*, Bull. Pol. Ac.: Tech., **33** (1985), 551–559.

М. Новотны, З. Павляк, **Частичная зависимость признаков**

Для каждой пары Z, T множеств признаков информационной системы определяется степень зависимости T от Z. Изучаются основные свойства степеней зависимости.