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

COMPUTER SCIENCE

Dependency of Attributes in Information Systems

by

Zdzisław PAWLAK and Cecylia RAUSZER

Presented by Z. PAWLAK on November 13, 1984

Summary. The concept of the relation of dependency between attributes as well as between sets of attributes is introduced as a tool for describing properties of the set of all attributes of an information systems. The algebra of dependency is also studied.

1. Dependency of sets of attributes. The concept of dependency has been extensively studied by many authors and used for the database design. The relation of dependency introduced in the paper is defined in a different way than it has been done for example in [1, 3, 4, 5]. The definition of the relation of dependency followed from the application of information systems to medical data analysis [8].

The basic notion of this paper, an information system, was introduced by Pawlak [7] and then was widely examined by Marek and Pawlak [6] and other authors.

The concept of information systems is not univocaly understood by different authors. Thus it seems justified to start our considerations from the definition of information systems.

By an *information system* we mean an ordered quadruple

$$\mathcal{S} = (U, Q, V, \varrho)$$

where

U is a set called the *universe* of $\mathcal{S}$; elements of U are called *objects* of $\mathcal{S}$,

Q is a set of *attributes*,

$V = \bigcup_{q \in Q} V_q$ is a set of *values of attributes*; V_q will be called the *domain* of q,

$\varrho: U \times Q \to V$ is a *description function* such that $\varrho(x, q) \in V_q$ for every $q \in Q$ and $x \in U$.

For given information system $\mathcal{S}$ we introduce function $\varrho_x: Q \to V$ such that $\varrho_x(q) = \varrho(x, q)$, where $x \in U$, $q \in Q$ and ϱ is defined above. The function ϱ_x will be called *description* of x in $\mathcal{S}$.

We say that objects $x, y \in U$ are *indiscernible* with respect to $q \in Q$ in $\mathcal{S}$, in symbols $xq^* y$, if $\varrho_x(q) = \varrho_y(q)$. In other words $xq^* y$ iff x and y have the same description in $\mathcal{S}$.

Notice that for each $q \in Q$ the relation q^* is an equivalence relation over U.

We say that objects $x, y \in U$ are *indiscernible* with respect to $P \subseteq Q$, in symbols $xP^* y$, if $P^* = \bigcap_{q \in P} q^*$.

In other words $xP^* y$ means that the objects x, y are indiscernible with respect to each attribute q from P.

It is obvious that the relation P^* is an equivalence relation over U.

1.1 *For any sets of attributes* $P, R \subseteq Q$

$$P^* \cap R^* = (P \cup R)^*. \quad \blacksquare$$

We will write $P' \subset P$ to denote that P' is a proper subset of P.

A subset $P \subseteq Q$ is said to be *independent* in the information system $\mathcal{S}$ if for every $P' \subset P$, $P^* \subset P'^*$.

A subset $P \subseteq Q$ is said to be *dependent* in the information system $\mathcal{S}$ if there exists a P' such that $P' \subset P$ and $P'^* \subseteq P^*$.

In the paper the character $\mathcal{S}$ will be always used to denote an information system.

Let P be any set of attributes in $\mathcal{S}$. Notice

1.2. *P is dependent in $\mathcal{S}$ iff there is a $P' \subset P$ such that $P^* = P'^*$.* $\blacksquare$

1.3. *P is independent in $\mathcal{S}$ iff P is not dependent in $\mathcal{S}$.* $\blacksquare$

A set $P' \subseteq P$ is said to be *superfluous* in P if $(P - P')^* = P$.

EXAMPLE 1. For the purpose of illustration, let us consider one simple information system $\mathcal{S} = (U, Q, V, \varrho)$ where $U = \{x_1, x_2, x_3, x_4, x_5\}$, $Q = \{q_1, q_2, q_3, q_4\}$, $V_{q_1} = V_{q_2} = V_{q_3} = \{0, 1\}$, $V_{q_4} = \{0, 1, 2\}$ and function ϱ is given by the table

$U \backslash Q$	q_1	q_2	q_3	q_4
x_1	0	0	0	0
x_2	0	1	0	2
x_3	1	1	0	1
x_4	1	1	0	1
x_5	0	1	1	2

Observe that Q is dependent in $\mathcal{S}$, namely $Q^* = \{q_3, q_4\}^* = \{q_1, q_2, q_3\}^*$. The set $\{q_1, q_2, q_3\}$ is independent in $\mathcal{S}$, $\{q_1, q_2\}$ is superfluous in Q.

1.4. *If $P \subseteq Q$ and P is dependent in $\mathcal{S}$, then there is a $P' \subset P$ such that P' is superfluous in Q.*

Proof. If $P \subseteq Q$ and P is dependent in $\mathcal{S}$, then for some $P' \subset P$ $P'^* = P^*$. Thus $P - P'$ is superfluous in Q. ■

Let us note that the converse theorem does not hold (cf. the example 1). However

1.5. *P is dependent in $\mathcal{S}$ iff there is a $P' \subset P$ such that P' is superfluous in P.* ■

1.6. *If P is independent in $\mathcal{S}$ then every subset of P is also independent in $\mathcal{S}$.*

Proof. Suppose that there is a dependent (in $\mathcal{S}$) subset P' of P. Thus for some $P'' \subset P'$ $P''^* = P'^*$. Notice that $P' - P''$ is superfluous in P, i.e. $(P - (P' - P''))^* = P^*$. According to 1.5 P is dependent in $\mathcal{S}$.

From the last theorem follows

1.7. *If P' is a dependent set of attributes in $\mathcal{S}$, then every set P such that $P' \subset P$ is dependent in $\mathcal{S}$.* ■

1.8. *Let $P \subseteq Q$ be independent in $\mathcal{S}$. Then for $P_1, P_2 \subset P$*

$$P_1^* \subseteq P_2^* \quad \textit{implies} \quad P_2 \subseteq P_1.$$

Proof. Suppose that $P_1^* \subset P_2^*$. Then $P_1^* = (P_1 \cup P_2)^*$. If $P_1 \neq P_1 \cup P_2$ then $P_1 \cup P_2$ is dependent in $\mathcal{S}$, a contradiction. ■

Let $P = \{p_1, \ldots, p_n\}$ and let us denote by P_i the set $P - \{p_i\}$, $i = 1, \ldots, n$.

1.9. *P is independent in $\mathcal{S}$ iff for every $i = 1, \ldots, n$, $P^* \subset P_i^*$.*

Proof. ($\Rightarrow$) is obvious. ($\Leftarrow$). The case $n = 2$ is trivial.

Let $P = \{p_1, \ldots, p_k, p_{k+1}\}$ and for $n = k$ the lemma hold.

Suppose that for every $i = 1, \ldots, k+1$ $P^* \subset P_i^*$ and P is dependent in $\mathcal{S}$. Then there is a $P' \subset P$ such that $P^* = P'^*$. Notice that $\bar{\bar{P}}' \leqslant k$. Thus for some $i_0 \leqslant k$ $P_{i_0}^* = P'^*$, a contradiction. ■

2. Dependency of attributes. Let $\mathcal{S} = (U, Q, V, \varrho)$ be an information system and let $p, q \in Q$.

An attribute p is said to be *dependent* in $\mathcal{S}$ on an attribute q, if $q^* \subset p^*$. The dependency of two attributes p, q will be denoted by $p \to q$.

Intuitive interpretation of the relation $\to$ is as follows:

Holding of the condition $p \to q$ means that if objects x, y are indiscernible with respect to attribute p then the objects x, y are also indiscernible with respect to q.

A pair (p, q) of attributes is called *dependent* in $\mathscr{S}$ if $p \to q$ or $q \to p$.

A pair (p, q) of attributes is said to be *independent* in $\mathscr{S}$ if neither $p \to q$ nor $q \to p$.

2.1. *If a set of attributes P is independent in $\mathscr{S}$, then all its attributes are pairwise independent in $\mathscr{S}$.*

Proof. Suppose that there are $p, q \in P$ such that (p, q) is dependent in $\mathscr{S}$, i.e. either $p^* \subset q^*$ or $q^* \subset p^*$. Then the set $\{p, q\} \subset P$ is dependent in $\mathscr{S}$ that, by 1.6, contradicts our assumption. ■

Notice that the converse theorem to 2.1 is not true. Indeed, let us consider the following

EXAMPLE 2. Let $\mathscr{S}$ be the information system given by the table

U \ Q	q_1	q_2	q_3
x_1	1	2	0
x_2	1	0	1
x_3	0	2	1
x_4	0	1	1
x_5	2	1	1

For every i, j, $i \neq j$ the pair (q_i, q_j) of attributes is independent. However the set Q is dependent in $\mathscr{S}$. Namely $Q = \{(x_i, x_i): i = 1, \ldots, 5\} = \{q_1, q_2\}$.

2.2. *If there is a pair (p, q) of attributes dependent in $\mathscr{S}$, then every set P of attributes such that $p, q \in P$ is dependent in $\mathscr{S}$.* ■

3. The algebra of dependency. Let $\mathscr{S} = (U, Q, V, \varrho)$ be an information system. Recall that for any set of attributes P, $P^* = \bigcap_{q \in P} q^*$, where q^* is an equivalence relation over U, there is also an equivalence relation over U.

Let us denote by A the set of all equivalence relation over U that are determined by the information system $\mathscr{S}$, i.e.

$$A' = \{p^* : p \in P(Q)\}. \tag{1}$$

Let $A = A'/=$. For simplicity elements of A will be denoted in the same way as elements of A'.

Consider A as an ordered set by the set theoretical inclusion $\subseteq$.

Notice that the relation Q is the least element in $(A, \subseteq)$ and the relation $\emptyset^* = U \times U$ is the greatest element in $(A, \subseteq)$.

Let $\cap$ be a binary operation in A defined as follows:

For $p, q \in A$

$$p^* \cap q^* = (p \dot{\cup} q)^*, \tag{2}$$

where the operation $\dot{\cup}$ on the right hand side of (2) means the set theoretical union (cf. 1.1).

Recall that $A \subset B$ is to be meant that A is a proper subset of B.

3.1. *Let $p \subseteq Q$. If there exists a subset p_1 of Q such that $p_1 \subseteq Q - p$ and $p^* \cap p_1^* = Q^*$, then Q is dependent in $\mathscr{S}$.*

Proof. Immediate from (2). ■

The above remark suggests the following definition of an unary operation $\neg$ in $(A, \subseteq)$, called complement. Namely for any $p^* \in A$ if $\sup \{q^* \in A: p^* \cap q^* = Q^*\}$ exists in $(A, \subseteq)$, then

$$\neg p^* = \sup \{q^* \in A: p^* \cap q^* = Q^*\}. \tag{3}$$

3.2. *If Q is independent in $\mathscr{S}$, then for every $p^* \in A^*$, $\neg p^* = (-p)^*$, where $-p$ means the set theoretical complement, i.e. $-p = Q - p$.*

Proof. Suppose that Q is independent. We show that for every $p^* \in A, (-p)^* = \sup \{q^*: p^* \cap q^* = Q^*\}$.

Notice that for any $p^* \in A$, $p^* \cap -p^* = Q^*$. Let for some $p_0^*, q^* \in A$ $p_0^* \cap q^* = Q^*$ and $(-p_0)^* \subseteq q^*$. Then $p_0 \dot{\cup} q$ must be Q, as otherwise Q is dependent in $\mathscr{S}$. Thus according to 1.8 $q \subseteq -p_0$. Hence $(-p_0)^* = q$. ■

As a corollary we obtain

3.3. *If for some $p \subset Q$, $\neg p^* \neq (Q-p)^*$, then Q is dependent in $\mathscr{S}$.* ■

3.4. *If for some $p \subset Q$, $p^* = \neg p^*$, then Q is dependent in $\mathscr{S}$.* ■

Let $A^\neg$ be the set of all p^* such that $\neg p^*$ exists in $(A, \subseteq)$. Consider $A^\neg$ as an abstract algebra

$$\mathfrak{A} = (A^\neg, \cap, \neg, Q^*, U \times U) \tag{4}$$

where $\cap$ and $\neg$ are defined by (2) and (3), respectively.

3.5. *In each algebra of the form* (4) *the following conditions are satisfied*

1) $p^* \subseteq q^*$ *if and only if* $p^* \cap q^* = p^*$

2) $Q^* \subseteq p^*$, $p^* \subseteq U \times U$

3) $p^* \cap q^* \subseteq (p \cap q)^*$, *where the sign $\cap$ on the right hand side of the inclusion means the set theoretical meet*

4) $(p \cap q)^* \cap q^* = q^*$

5) $\neg Q^* = U \times U$, $\neg U \times U = Q^*$

6) $p^* \cap \neg p^* = Q^*$

7) *If* $p^* \subseteq q^*$ *then* $\neg q^* \subseteq \neg p^*$

8) $p^* \subseteq \neg\neg p^*$

9) $\neg p^* = \neg\neg\neg p^*$

10) $\neg p^* \cap \neg q^* \subseteq \neg (p^* \cap q^*)$

11) $\neg (p \cap q)^* \subseteq \neg p \cap \neg q$

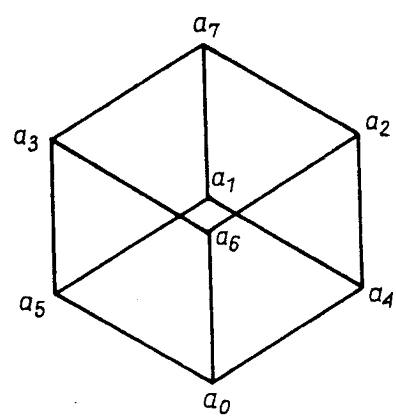
12) $(-p)^* \subseteq \neg p^*$.

Proof. By an easy verification. ■

In the sequel the algebra of the form (4) will be called an *algebra of dependencies* of the information system $\mathcal{S}$ or shortly an algebra of dependencies and it will be denoted by $\mathfrak{A}_{\mathcal{S}}$.

3.6. *In any algebra of dependencies the converse relations to* 3.5 (10) *and* (11) *do not hold.*

Proof. Let us consider the algebra $\mathfrak{A}_{\mathcal{S}}$ of the information system $\mathcal{S}$ given in Example 1. The algebra $\mathfrak{A}_{\mathcal{S}}$ can be presented by the following diagram



where

$$a_0 = \{q_3, q_4\}^* = \{q_1, q_2, q_3\}^* = \{q_1, q_3, q_4\}^* = \{q_2, q_3, q_4\}^* = Q^*,$$

$$a_1 = \{q_1\}^*, \quad a_2 = \{q_2\}^*, \quad a_3 = \{q_3\}^*, \quad a_4 = \{q_4\}^* = \{q_1, q_2\}^* =$$
$$= \{q_1, q_4\}^* = \{q_2, q_4\}^* = \{q_1, q_2, q_4\}^*, \quad a_5 = \{q_1, q_3\}^*, \quad a_6 = \{q_2, q_3\}^*$$
$$a_7 = U \times U = \{x_1, x_2, x_3, x_4, x_5\}^2.$$

Let us observe that $\geqslant(a_5 \cap a_4) = \geqslant a_0 = \geqslant Q^* = U \times U$. However $\geqslant a_5 = a_2$, $\geqslant a_4 = a_3$ and $a_2 \cap a_3 = a_6 \subset U \times U$. It proves that $\geqslant(p^* \cap q^*)$ does not imply $\geqslant p^* \cap \geqslant q^*$. Similarly, $\geqslant a_4 = a_3$, $\geqslant(\{q_1, q_2\} \cap \{q_1, q_4\})^* = \{q_1\}^* = \geqslant a_1 = a_0$. Thus the converse to 3.5 (11) does not hold.

It follows immediately from 3.4 that

3.7. *If q is superfluous in p, then* $\neg p^* \subset \neg(-q)^*$. ■

3.8. *If Q is independent, then $(p \cap q)^*$ is the smallest set in $\mathfrak{A}_{\mathscr{S}}$ containing p and q.*

Proof. It follows immediately from 1.8. ■

As a consequence we obtain

3.9. *If Q is independent then $(p \cap q)^*$ is the transitive/ closure in $\mathfrak{A}_{\mathscr{S}}$ of $p^* \cup q^*$.* ■

Properties of the algebra of the information system $\mathscr{S}$ are closely related to the properties of the set of all attributes Q in S. Namely

3.10. *The set Q of all attributes of an information system $\mathscr{S}$ is independent in $\mathscr{S}$ if and only if the algebra of dependencies $\mathfrak{A}_{\mathscr{S}}$ of $\mathscr{S}$ is a Boolean algebra such that $\bar{\bar{A}}^{\neg} = 2^n$ where $n = \bar{\bar{Q}}$.*

Proof. Suppose that Q is independent. It follows from 3.2 that $A^{\neg} = A$. According to 3.8 A is closed with respect to the set-theoretical union $\cup$. To show that $\mathfrak{A}_{\mathscr{S}}$ is a Boolean algebra we need only to prove that

$$\neg p^* \cup p^* = U \times U.$$

On account of 3.2 and 3.8 we infer

$$\neg p^* \cup p^* = (-p)^* \cup p^* = (-p \cap p)^* = Q^* = U \times U.$$

Suppose now that Q is dependent. Then there is a proper subset p of Q such that $p^* = Q^*$ and therefore $\bar{\bar{A}}^{\neg} < 2^n$. ■

Immediately from 3.10 we infer

3.11. *If the cardinality of $A^{\neg}$ is less than 2^n, where n is the cardinality of Q, then Q is dependent in $\mathscr{S}$.* ■

4. "Union" of the relations of dependency. In this section we shall introduce and discuss a new binary operation $+$ in the set of an equivalence relations of $\mathscr{S}$. This operation will be used, in a separate paper to show some properties of the decomposition structure examined in [2].

Define the binary operation $+$ as follows:

$$p^*+q^* = \bigcup z \{(x, y)\in U \times U: (x, z)\in q^* \,\&\, (z, y)\in p^*\}.$$

Notice that in general p^*+q^* need not be an equivalence relation i.e. in general, $p^*+q^* \notin A$. We can only say that p^*+q^* is a reflexive relation.

4.1. *For any* $p, q \subset Q$

(a) $p^* \subseteq p^*+q^*$, $q \subseteq p^*+q^*$

(b) $p^*+q^* \subseteq (p\cap q)^*$.

Proof. (a) is obvious. (b) is a consequence that $p^* \subseteq (p\cap q)^*$ as well as $q^* \subset (p\cap q)^*$. ■

Now we will impose on p^* and q^* some conditions so that p^*+q^* becomes an equivalence relation in $\mathscr{S}$.

4.2. *If* $p^* \subseteq q^*$ *then* $p^*+q^* \in A$, *i.e.* p^*+q^* *is an equivalence relation in* $\mathscr{S}$. *Moreover* $p^*+q^* = q^*$.

Proof. According to 4.1 (a) we need only to show that

$$\bigcup z \{(x, y)\in U \times U: (x, z)\in q^* \,\&\, (z, y)\in p^*\} \subseteq q^*.$$

But it follows immediately by assumption. ■

Notice without proof that

4.3. *If* $p, q \subset Q$ *then* $p^*+(p^*\cap q^*) = p^*$. ■

4.4. *Let* $p, q \subset Q$. *If* q *is superfluous in* p, *then* $p^*+(-q)^*$ *is an equivalence relation in* $\mathscr{S}$. *Moreover* $p^*+(-q)^* = p^*$.

Proof. On account of 4.1 (a) we need to show that $p^*+(-q)^* \subset p^*$. Let $(x, y)\in p^*+(-q)^*$. Then for some z, $(x, z)\in(-q)^*$ and $(z, y)\in p^*$. Thus $(x, y)\in(p\cap -q)^*$. By the assumption q is superfluous in p, i.e. $(p\cap -q)^* = p^*$. ■

4.5. *If* p^*+q^* *is an equivalence relation in* $\mathscr{S}$, *then* p^*+q^* *is the smallest element in the ordered set* $(A, \subseteq)$ *containing* p^* *and* q^*.

Proof. According to 4.1 (a) p^* and q^* are contained in p^*+q^*. Let assume that for some $r^*\in A$, $p^* \subseteq r^*$ and $q^* \subseteq r^*$. Let $\langle x, y\rangle \in p^*+q^*$. Then for some z^*, $\langle x, z\rangle \in q^*$ and $\langle z, y\rangle \in p^*$. Thus by transitivity of r^*, $\langle x, y\rangle \in r^*$, i.e. $p^*+q^* \subseteq r^*$. ■

It turns out that the binary operation + defined in this section may be used to describe properties of multivalued dependencies [9]. Namely we can prove

4.6. *Multivalued dependency* $p^* \to q^*$ *holds in an information system* $\mathscr{S}$ *if and only if* $p^*\cap q^*+p^*\cap(-q)^*$ *is an equivalence relation in* $\mathscr{S}$, *i.e.*

if and only if $p^* \cap q^* + p^* \cap (-q)^*$ *is an element of the algebra of dependencies of* $\mathcal{S}$. ■

Connections between multivalued dependencies and the operation + are examined in a separate paper.

INSTITUTE OF COMPUTER SCIENCE, POLISH ACADEMY OF SCIENCE, P.O. BOX 22, 00-901 WARSZAWA
(INSTYTUT PODSTAW INFORMATYKI, POLSKA AKADEMIA NAUK)
INSTITUTE OF MATHEMATICS, WARSAW UNIVERSITY, PKIN, 00-901 WARSZAWA
(INSTYTUT MATEMATYKI, UNIWERSYTET WARSZAWSKI)

REFERENCES

[1] W. W. Armstrong, *Dependency structures of database relationships*, Proc. IFIP 74, North-Holland Publ. Co., Amsterdam, 1974, 580–583.

[2] W. W. Armstrong, C. Delobel, *Decompositions and functional dependencies in relations*, ACM Trans. Database Syst., **5** (1980), 404–430.

[3] E. F. Codd, *Further normalization of the database relational model*, Data Base Systems, Courant Computer Science Symposia 6, Prentice Hall, Englewood Cliffs, N. J., 1971, 33–64.

[4] C. J. Date, *An Introduction to Database Systems*, 2nd ed., Addison-Wesley, Reading, Mass. 1977.

[5] R. Fagin, *Horn clauses and database dependencies*, J. ACM, **23** (1982), 952–985.

[6] W. Marek, Z. Pawlak, *On information storage and retrieval systems*, Theoretical Computer Science, **1** (1976), 331–354.

[7] Z. Pawlak, *Mathematical foundations of information retrieval*, CC PAS Reports, No. 101, Warszawa (1973).

[8] Z. Pawlak, *Rough classification*, CC PAS Reports, No. 506, Warszawa 1983.

[9] J. D. Ullman, *Principles of database systems*, Computer Science Press, 1980.

З. Павляк, Ц. Паушер, **Зависимость характерных признаков в информационных системах**

Целью настоящей работы является изучение свойств отношений зависимости между характерными признаками, а также множествами характерных прознаков в информационной системе.

Вводятся понятия алгебры отношений зависимости и показываются некоторые свойства данной алгебры.