Annales Societatis Mathematicae Polonae
Series IV: Fundamenta Informaticae VII.3 (1984)

ON SOME SUBSET OF THE PARTITION SET

Jerzy W. Grzymała-Busse

The University of Kansas

Zdzisław Pawlak

Polish Academy of Sciences

Received September 9, 1983

AMS categories: 68C05

A b s t r a c t. This paper contains a simple algorithm for minimal partition of a set, which is the departure point to study attribute dependencies in information system (see [3], [6], [7], [9]). Theoretical properties of such partitions have been studied by Łoś (see [5]) and the proposed algorithm has been implemented by Stevens (see [8]). The implementation shows many practical advantages of the proposed method.

K e y w o r d s: partition, block, minimal nontrivial partition, minimal number of partitions.

1. Let n be some positive integer.

The set $\{1,2,\dots,n\}$ will be denoted by N.

A partition π on N is a set of mutually disjoint subsets of N whose union is N. These disjoint subsets of N will be called blocks of N. For $i,j \in N$ we will write $i \equiv j(\pi)$ if and only if both i and j are members of the same block of π. The parition on N such that all blocks of it contain one element will be denoted by 0.

If π_1 and π_2 are partitions on N, then the product $\pi_1 \cdot \pi_2$ is the partition on N such that $i \equiv j(\pi_1 \cdot \pi_2)$ if and only if $i \equiv j(\pi_1)$ and $i \equiv j(\pi_2)$, where $i,j \in N$. The sum $\pi_1+\pi_2$ is the partition N such that $i \equiv j(\pi_1+\pi_2)$ if and only if there esists a sequence in N

$$i = i_0, i_1, i_2, \dots, i_r = j$$

for which either $i_s \equiv i_{s+1}(\pi_1)$ or $i_s \equiv i_{s+1}(\pi_2)$, where $s = 0,1,\dots,r-1$.

For the rest of the paper we assume that Π is a set $\{\pi_1, \pi_2, \dots, \pi_m\}$ of partition on N.

2. The paper is devoted to the solution of the following problem:

<u>PROBLEM.</u> Find the minimal number k and the set Π_k of all subsets of k different elements of Π such that if $\{\pi_{i_1}, \pi_{i_2}, \dots, \pi_{i_k}\} \in \Pi_k$ then

$$\pi_{i_1} \cdot \pi_{i_2} \cdots \pi_{i_k} = \pi_1 \cdot \pi_2 \cdots \pi_m .$$

Let i and j be different members of N. Let π_{ij} be the partition on N such that $\{i,j\}$ is a block of π_{ij} and all blocks of π_{ij} except $\{i,j\}$ have one element of N.

In other words, the partition π_{ij} can be represented in the following form:

$$\left\{\overline{1};\ \overline{2};\ \ldots;\ \overline{i-1};\ \overline{i,j};\ \overline{i+1};\ \ldots;\ \overline{j-1};\ \overline{j+1};\ \ldots;\ \overline{n}\right\}.$$

Any partition on N of the above type, i.e. the partition on N with one block containing two elements of N and all remaining blocks containing one element of N, will be called minimal nontrivial partition on N.

Obviously, $\pi_{ij} = \pi_{ji}$.

Any partition $\pi_1 \in \Pi$ can be represented by a set P_1 of minimal nontrivial partitions on N in the following way: $j<j$ and $i \equiv j(\pi_1)$ if and only if $\pi_{ij} \in P_1$.

Property 1. $\sum_{\pi_{ij} \in P_1} \pi_{ij} = \pi_1$.

Proof. Let us denote

$$\pi = \sum_{\pi_{ij} \in P_1} \pi_{ij}.$$

We have to show that $\pi = \pi_1$. Obviously, if $i \equiv j(\pi_1)$, then $\pi_{ij} \in P_1$, whence $i \equiv j(\pi)$.

Suppose now that $i \equiv j(\pi)$. Then there exists a sequence in N

$$i = i_0, i_1, \ldots, i_r = j$$

with the property that for every $s=0,1,\ldots,r-1$ there exists a $\pi_{ij} \in P_1$ such that

$$i_s \equiv i_{s+1}(\pi_{ij}).$$

Considering the definition of P_1 and π_{ij}, we have

$$i_s \equiv i_{s+1}(\pi_1)$$

for $s = 0,1,\ldots,r-1$. Since $\equiv$ is obviously transitive, this implies that $i \equiv j(\pi_1)$, QED.

Property 2. Let π_1 and π_2 be partitions on N. Then

$$\pi_1 \cdot \pi_2 = \begin{cases} \sum_{\pi\pi_{ij} \in P_1 \cap P_2} \pi_{ij} & \text{if } P_1 \cap P_2 \neq \emptyset. \\ 0 & \text{if } P_1 \cap P_2 = \emptyset. \end{cases}$$

<u>Proof</u>. If $P_1 \cap P_2 = \emptyset$, then obviously $\pi_1 \cdot \pi_2 = 0$, for $i \equiv j(\pi_1 \cdot \pi_2)$ would imply that $\pi_{ij} \in P_1 \cap P_2$. Assume now that $P_1 \cap P_2 \neq \emptyset$ and denote

$$\pi = \sum_{\pi_{ij} \in P_1 \cap P_2} ij$$

We have to show that $\pi = \pi_1 \cdot \pi_2$.

Suppose first that $i \equiv j(\pi_1 \cdot \pi_2)$. Then $i \equiv j(\pi_1)$ and $i \equiv j(\pi_2)$, whence $\pi_{ij} \in P_1 \cap P_2$. Obviously, this implies that $i \equiv j(\pi)$.

Conversely, suppose that $i \equiv j(\pi)$. Then there exists a sequence in N

$$i = i_0, i_1, \ldots, i_r = j$$

with the property that for every $s \in \{0, 1, \ldots, r-1\}$ there exists a $\pi_{ij} \in P_1 \cap P_2$ such that

$$i_s \equiv i_{s+1}(\pi_{ij}).$$

Since $\pi_{ij} \in P_1 \cap P_2$, then by the definitions of P_1 and P_2 we have

$$i_s \equiv i_{s+1}(\pi_1) \quad \text{and} \quad i_s \equiv i_{s+1}(\pi_2)$$

for $s = 0, 1, \ldots, r-1$, whence $i \equiv j(\pi_1)$ and $i \equiv j(\pi_2)$. This yields $i \equiv j(\pi_1 \cdot \pi_2)$, QED.

<u>Proposition</u>. Let P be the set $\bigcap_{\pi_i \in \Pi} P_i$ and let S_i be the set $P_i \setminus P$, where $i \in \{1, 2, \ldots, m\}$. Then the number k is the minimal number such that there exist $S_{i_1}, S_{i_2}, \ldots, S_{i_k}$ with $S_{i_1} \cap S_{i_2} \cap \ldots \cap S_{i_k} = \emptyset$. The set Π_k is the set of all sets $S_{i_1}, S_{i_2}, \ldots, S_{i_k}$ with $S_{i_1} \cap S_{i_2} \cap \ldots \cap S_{i_k} = \emptyset$.

<u>Proof</u>. If $P = \emptyset$ the proof is obvious. Otherwise, thanks to Property 2 and the definition of P,

$$\pi_1 \cdot \pi_2 \cdots \pi_m = \sum_{\pi_{ij} \in P} \pi_{ij}.$$

Let $\pi_{i_1}, \pi_{i_2}, \ldots, \pi_{i_l} \in \Pi$. Then

$$\pi_{i_1} \cdot \pi_{i_2} \cdots \pi_{i_l} = \sum_{\pi_{ij} \in P_{i_1} \cap P_{i_2} \cap \cdots \cap P_{i_l}} \pi_{ij} \, .$$

But

$$P_{i_1} \cap P_{i_2} \cap \cdots \cap P_{i_l} = (P_{i_1} \setminus P) \cap (P_{i_2} \setminus P) \cap \cdots \cap (P_{i_l} \setminus P) \cup P$$

$$= S_{i_1} \cap S_{i_2} \cap \cdots \cap S_{i_l} \cup P,$$

and hence

$$\pi_{i_1} \cdot \pi_{i_2} \cdots \pi_{i_l} = \sum_{\pi_{ij} \in S_{i_1} \cap S_{i_2} \cap \cdots \cap S_{i_l}} \pi_{ij} + \sum_{\pi_{ij} \in P} \pi_{ij}$$

Thus

$$\pi_{i_1} \cdot \pi_{i_2} \cdots \pi_{i_l} = \pi_1 \cdot \pi_2 \cdots \pi_m$$

if and only if

$$S_{i_1} \cap S_{i_2} \cap \cdots \cap S_{i_l} = \emptyset .$$

3. From the above result it follows an algorithm for determining the number k and the set $\overline{\Pi}_k$.

<u>Algorithm</u>.

Step 1. Compute sets P_i, $i = 1,2,\ldots,m$.

Step 2. Compute the set P.

Step 3. Compute sets S_i, $i = 1,2,\ldots,m$.

Step 4. Initiate $k=1$.

Step 5. Compute all possible sets $S_{i_1} \cap S_{i_2} \cap \cdots \cap S_{i_k}$.

Step 6. Check if any of sets computed in Step 5 is equal to the empty set. If so, print the number k and all partitions $\pi_{i_1}, \pi_{i_2}, \ldots, \pi_{i_k}$ which correspond to $S_{i_1}, S_{i_2}, \ldots, S_{i_k}$ such that $S_{i_1} \cap S_{i_2} \cap \cdots \cap S_{i_k} = \emptyset$ and stop. Otherwise, do Step 7.

Step 7. Increment k by 1 and do Step 5.

Acknowledgment. Thanks are due to dr B. Konikowska for critical remarks.

References

[1] W.W. Armstrong, Dependency Structures of Data Base Relationships, Proc. IFIP Congress, Stockholm, Aug. 5-10, 1974.

[2] R. Fagin, Relational Data Base Decomposition and Propositional Logic, IBM Research Report, San Jose, July - August 1976.

[3] T. Imieliński, A note on minimization the set of attributes. Fundamenta Informaticae, Vol. No 1 (1983) pp. 45-51.

[4] M. Jeagermann, W. Marek, Dependencies of attributes in information Systems. ICS PAS Reports, No 428 (1981).

[5] J. Łoś, Characteristic Sets of a System of Equivalence Relations, Colloquium Mathematicum XLII, 291-293 (1979).

[6] E. Orłowska, On dependency of attributes in information Systems, ICS PAS Reports, No 425 (1981).

[7] Z. Pawlak, Information Systems; Theoretical Foundations, Information Systems. Vol. 6, No 3 (1981), pp. 205-218.

[8] M.E. Stevens, A program for sets partitions, Kansas University (unpublished manuscript) (1983).