Some results about functional dependencies*

J. DEMETROVICS and V. D. THI

Abstract

§ 1. Introduction

The relational datamodel was defined by E. F. Codd [2]. In this datamodel a relation is a table (matrix) in which each column corresponds to a distinct attribute and each row to a distinct record. Relations are used to describe connections among data items. The functional dependency is one of the main concepts in relational datamodel. The mathematical structure of functional dependencies was thouroughly investigated by W. W. Armstrong [1]. The equivalence of sets of minimal keys with Sperner-systems was proved [4]. It is known [1] that for a given family F of functional dependencies there is a relation representing F in the sense that the full family of functional dependencies of this relation is exactly F. Also it is shown [4] that for an arbitrarily given Sperner-system there exists a relarion R representing this Sperner-system so that this Sperner-system is exactly the set of all minimal keys of R. In this paper we give necessary and sufficient conditions for a relation to represent a given family of functional dependencies or a Sperner-system.

The closure operation is a useful and interesting instrument for investigating the structure of functional dependencies. In this paper we investigate the connection between closure operations and sets of minimal keys, too. Now we give some necessary definitions.

Let $\Omega=\{a_1, ..., a_n\}$ be a finite non-empty set of attributes. For each attribute a_i there is a non-empty set $D(a_i)$ of all possible values of that attribute. An arbitrary finite subset of the Cartesian product $D(a_1)\times ... \times D(a_n)$ is called a relation over Ω. It can be seen that a relation over Ω is a set of mappings $h: \Omega \to \bigcup_{a\in\Omega} D(a)$, where $h(a)\in D(a)$ for all a.

*This paper was supported by grants from the Hungarian Academy of Sciences OTKA Nr. 1066 and 1812.

The main purpose of this paper is to give necessary and sufficient conditions for a relation to represent an arbitrarily given family of functional dependencies or a closure operation or a Sperner-system. The connection between closure operations and sets of minimal keys is investigated too.

Definition 1.1. [2] Let $R=\{h_1, \ldots, h_m\}$ be a relation over the finite set of attributes Ω. Let $A, B \subseteq \Omega$. We say that B functionally depends on A in R (denoted as $A \xrightarrow[R]{} B$) iff $(\forall h_i, h_j \in R)((\forall a \in A)(h_i(a)=h_j(a)) \to (\forall b \in B)(h_i(b)=h_j(b)))$.

Let $F_R=\{(A, B): A \xrightarrow[R]{} B\}$. F_R is called the full family of functional dependencies of R.

Definition 1.2. [1] Let Ω be a finite set, and denote $P(\Omega)$ its power set. Let $F \subseteq P(\Omega) \times P(\Omega)$. We say that F is an f-family over Ω iff for all $A, B, C, D \subseteq \Omega$

(F1) $(A, A) \in F$;
(F2) $(A, B) \in F, (B, C) \in F \to (A, C) \in F$;
(F3) $(A, B) \in F,\ A \subseteq C,\ D \subseteq B \to (C, D) \in F$;
(F4) $(A, B) \in F,\ (C, D) \in F \to (A \cup C, B \cup D) \in F$.

By [1], F_R is an f-family over Ω. It is known [1] that if F is an f-family, then there is a relation R over Ω such that $F_R=F$.

Definition 1.3. The mapping $L: P(\Omega) \to P(\Omega)$ is called a closure operation over Ω iff for every $A, B \subseteq \Omega$:

(1) $A \subseteq L(A)$;

(2) $A \subseteq B \to L(A) \subseteq L(B)$;

(3) $L(L(A)) = L(A)$.

Remark 1.1. It is easy to see that if F is an f-family and for all $A \subseteq \Omega$, we set $L_F(A)=\{a \in \Omega: (A, \{a\}) \in F\}$ then L_F is a closure operation over Ω. Conversely, it is shown [1] that if L is a closure operation over Ω, then there is exactly one f-family such that $L_F=L$, where $F=\{(A, B): B \subseteq L(A)\}$. Thus, between closure operations and f-families over Ω there exists an one-to-one correspondence.

Definition 1.4. Let R be a relation, F an f-family and L a closure operation over Ω. We say that R represents F (L) iff $F_R=F$ $(L_{F_R}=L)$.

Definition 1.5. Let R be a relation, L a closure operation over Ω, and $K \subseteq \Omega$. We say that K is a key of R (of L) if $K \to \Omega$ $(L(K)=\Omega)$. K is a minimal key of R (of L) if K is a key of R (of L) and for any proper subset B of K, $B \not\xrightarrow[R]{} \Omega$ $(L(B) \neq \Omega)$. Denote $\mathscr{K}_R$ the set of all minimal keys of R and $\mathscr{K}_L$ that of L. Clearly, $K_i, K_j \in \mathscr{K}_R$ implies $K_i \not\subseteq K_j$. Systems of subsets of Ω satisfying this condition are Sperner-systems. Consequently, $\mathscr{K}_R$, $\mathscr{K}_L$ are Sperner-systems.

For a Sperner-system $\mathscr{K}$ we can define the set of antikeys of $\mathscr{K}$ (denoted by $\mathscr{K}^{-1}$) as follows:

$$\mathscr{K}^{-1} = \{B \subset \Omega: (K \in \mathscr{K}) \to (K \not\subseteq B) \text{ and } (B \subset C) \to (\exists K \in \mathscr{K})(K \subseteq C)\}.$$

It is easy to see that $\mathscr{K}^{-1}$ is also a Sperner-system. Clearly, the elements of $\mathscr{K}^{-1}$ do not contain the elements of $\mathscr{K}$ and they are maximal for this property.

Definition 1.6. Let $R=\{h_1, \ldots, h_m\}$ be a relation over Ω. For $1 \leq i < j \leq m$ denote E_{ij} the set $\{a \in \Omega: h_i(a)=h_j(a)\}$. We set $E_R=\{E_{ij}: 1 \leq i < j \leq m\}$. E_R is called the equality set of R.

§ 2. Results

Now we give a necessary and sufficient condition for a relation representing a given f-family. It is a precise characterization for relations represent f-families.

Theorem 2.1. Let $R=\{h_1, ..., h_m\}$ be a relation and F an f-family over Ω. Then R represents F iff for every $A\subseteq\Omega$

$$L_F(A)=\begin{cases}\bigcap\limits_{A\subseteq E_{ij}} E_{ij} & \text{if } \exists E_{ij}\in E_R\colon A\subseteq E_{ij},\\ \Omega & \text{otherwise,}\end{cases}$$

where $L_F(A)=\{a\in\Omega\colon (A, \{a\})\in F\}$ and E_R is the equality set of R.

Proof. It is easy to see that F_R is an f-family over Ω, first we prove that in an arbitrary relation R for all $A\subseteq\Omega$

$$L_{F_R}(A)=\begin{cases}\bigcap\limits_{A\subseteq E_{ij}} E_{ij} & \text{if } \exists E_{ij}\in E_R\colon A\subseteq E_{ij},\\ \Omega & \text{otherwise.}\end{cases}$$

We suppose that A is a set such that there is not an $E_{ij}\in E_R$ so that $A\subseteq E_{ij}$. Then for all h_i, $h_j\in R$ $\exists a\in A\colon h_i(a)\neq h_j(a)$. According to the definition of functional dependency $A\xrightarrow[R]{}\Omega$ holds. By the definition of the mapping L_{F_R} we obtain $L_{F_R}(A)=\Omega$. It is obvious that $L_{F_R}(\emptyset)=\bigcap\limits_{E_{ij}\in E_R} E_{ij}$ holds. If $A\neq\emptyset$ and there is an $E_{ij}\in E_R$ so that $A\subseteq E_{ij}$, then we set $V=\{E_{ij}\colon A\subseteq E_{ij},\ E_{ij}\in E_R\}$ and $E=\bigcap\limits_{E_{ij}\in V} E_{ij}$. Clearly, $A\subseteq E$. If $V=E_R$ holds, then $(A, E)\in F_R$ holds. If $V\neq E_R$ holds, then it can be seen that for all $E_{ij}\in V$ $(\forall a\in A)(h_i(a)=h_j(a))\rightarrow(\forall b\in E)(h_i(b)=h_j(b))$ and for all $E_{ij}\notin V$ $\exists a\in A\colon h_i(a)\neq h_j(a)$. Thus, $(A, E)\in F_R$ holds. By the definition of L_{F_R}, $E\subseteq L_{F_R}(A)$ holds. Clearly, by the definition of relation we have $E\subset\Omega$. From $A\subseteq E\subseteq L_{F_R}(A)$ and according to the definition of closure operation we obtain $(E, L_{F_R}(A))\in F_R$. Now we assume that c is an attribute such that $c\notin E$. Consequently, there is an $E_{ij}\in V$ so that $c\notin E_{ij}$. Thus, $\exists h_i, h_j\in R\colon \forall b\in E\colon h_i(b)=h_j(b)$ holds, but $h_i(c)\neq h_j(c)$. According to Definition 1.1, $(E\cup c)$ does not depend on E. Thus, for all attributes $c\notin E$ $(E, E\cup c)\notin F_R$ holds. By the definition of L_{F_R} we obtain $L_{F_R}(A)=\bigcap\limits_{E_{ij}\in V} E$. By Remark 1.1 it is easy to see that $F_R=F$ holds iff $L_{F_R}=L_F$ holds. The proof is complete. □

The following corollary is obvious.

Corollary 2.1. Let R be a relation and L a closure operation over Ω. Then R represents L iff for all $A\subseteq\Omega$

$$L(A)=\begin{cases}\bigcap\limits_{A\subseteq E_{ij}} E_{ij} & \text{if } \exists E_{ij}\in E_R\colon A\subseteq E_{ij},\\ \Omega & \text{otherwise.}\end{cases}\qquad\square$$

Definition 2.1. Let L be a closure operation over Ω. Let $Z(L)=\{A\subseteq\Omega\colon L(A)=A\}$, and $M(L)=\{A\subset\Omega\colon A\in Z(L), A\subset B\rightarrow L(B)=\Omega\}$ The elements of $Z(L)$ are called closed sets. $M(L)$ is the family of maximal closed sets (except Ω).

Clearly, $Z(L)$ is closed under intersection.

Definition 2.2. Let $N \subseteq P(\Omega)$. Denote N^+ the set $\{\bigcap N' : N' \subseteq N\}$. By convention $\bigcap \emptyset = \Omega$, i.e. N^+ always contains Ω. It can be seen that for all $E_{ij} \in E_R$ we have $E_{ij} \in Z(L_{F_R})$, i.e. $E_R^+ \subseteq Z(L_{F_R})$. By Theorem 2.1, $Z(L_{F_R}) \subseteq E_R^+$ holds. Clearly, if L_1, L_2 are two closure operations over Ω then $L_1 = L_2$ holds iff $Z(L_1) = Z(L_2)$. Consequently, the next corollary is clear.

Corollary 2.2. Let R be a relation and L a closure operation over Ω. Then R represents L iff $Z(L) = E_R^+$ holds. □

Definition 2.3. Let F be an f-family over Ω and $(A, B) \in F$. We say that (A, B) is a maximal right side dependency of F iff

$$\forall B' \quad (B \subseteq B'): (A, B') \in F \rightarrow B' = B.$$

Denote by $M(F)$ the set of all maximal right side dependencies of F. We say that B is a maximal side of F iff there is an A so that $(A, B) \in M(F)$. Denote $I(F)$ the set of all maximal sides of F.

It can be seen that $I(F) = Z(L_F)$. Consequently, the following corollary is obvious.

Corollary 2.3. Let F be an f-family and R a relation over Ω. Then R represents F iff $I(F) = E_R^+$. □

It is known ([1], [4]) that for an arbitrary non-empty Sperner-system $\mathcal{K}$ there is a relation R so that $\mathcal{K}_R = \mathcal{K}$.

Definition 2.4. Let R be a relation and $\mathcal{K}$ a Sperner-system over Ω. We say that R represents $\mathcal{K}$ iff $\mathcal{K}_R = \mathcal{K}$.

The next theorem is a useful precise characterization of relations which represent a given Sperner-system. First we define the following concept.

Definition 2.5. Let R be a relation over Ω, and E_R the equality set of R, i.e. $E_R = \{E_{ij} : 1 \leq i < j \leq m\}$, where $E_{ij} = \{a \in \Omega : h_i(a) = h_j(a)\}$. Let $T_R = \{A \subset \Omega : \exists E_{ij} \in E_R : E_{ij} = A$ and $\bar{\exists} E_{st} \in E_R : A \subset E_{st}\}$. Then T_R is called the maximal equality system of R.

Theorem 2.2. Let $\mathcal{K}$ be a non-empty Sperner-system and R a relation over Ω. Then R represents $\mathcal{K}$ iff $\mathcal{K}^{-1} = T_R$, where T_R is the maximal equality system of R.

Proof. As $\mathcal{K}$ is a non-empty Sperner-system, $\mathcal{K}^{-1}$ exists. On the other hand, $\mathcal{K}$ and $\mathcal{K}^{-1}$ are uniquely determined by each other, we obtain $\mathcal{K}_R = \mathcal{K}$ holds iff $\mathcal{K}_R^{-1} = \mathcal{K}^{-1}$ does. Consequently, we must prove that $\mathcal{K}_R^{-1} = T_R$.

It is obvious that F_R is an f-family over Ω. Now we suppose that A is an antikey of $\mathcal{K}_R$. Clearly, $A \neq \Omega$. If there is a B such that $A \subset B$ and $A \xrightarrow[R]{} B$ then by definition of antikeys we obtain $B \xrightarrow[R]{} \Omega$. Hence $A \xrightarrow[R]{} \Omega$ holds. This contradicts to $K \in \mathcal{K}_R : K \subseteq A$. So $A \in I(F_R)$ holds. If there is a B' so that $B' \neq \Omega$, $B' \in I(F_R)$, and $A \subset B'$, then B' is a key of R. This contradicts to $B' \neq \Omega$. Thus, $A \in I(F_R) \setminus \Omega$ and $\bar{\exists} B'$ $(B' \in I(F_R) \setminus \Omega)$: $A \subset B'$. On the other hand, $\Omega \notin T_R$ by definition of R. It is easy to see that $E_{ij} \in I(F_R)$. Hence $T_R \subseteq I(F_R)$ holds. If D is a set such that $\forall C \in T_R: D \nsubseteq C$, then from Definition 1.1, D is a key of R. Consequently, T_R is the set of maximal distinct elements of $I(F_R)$. So we obtain $A \in T_R$.

Conversely, we assume that $A \in T_R$. According to the definition of a relation and T, we have $A \underset{R}{\not\to} \Omega$, i.e. $\forall K \in \mathcal{K}_R: K \nsubseteq A$. On the other hand, by definition of T_R for all D $(A \subset D)$ $D \underset{R}{\to} \Omega$ holds. Consequently, by the definition of antikeys $A \in \mathcal{K}_R^{-1}$. The proof is complete. □

Now we investigate the connection between closure operations.

Lemma 2.1. [6] Let L be a closure operation over Ω, and $\mathcal{K}_L$ the set of minimal keys of L. Then $\mathcal{K}_L^{-1} = M(L)$. □

Definition 2.6. [3] Let Q be a set of all closure operations over Ω. An ordering over Q is defined as follows:

For $L, L' \in Q$ let $L \leqq L'$ iff for all $A \subseteq \Omega$, $L'(A) \subseteq L(A)$. It can be seen that Q is a partially ordered set for this ordering. If $L \leqq L'$ but $L \neq L'$ then the notation $L < L'$ is used.

Theorem 2.3. [3] Let L, L' be two closure operations over Ω. Then $L \leqq L'$ iff $Z(L) \subseteq Z(L')$. □

Based on Theorem 2.3 it is easy to see that $L < L'$ iff $Z(L) \subset Z(L')$.

Theorem 2.4. Let $\mathcal{K}$ be a non-empty Sperner-system over Ω, and $\mathcal{K}^{-1}$ the set of all antikeys of $\mathcal{K}$. Let

$$L(A) = \begin{cases} \bigcap\limits_{\substack{A \subseteq B \\ B \in \mathcal{K}^{-1}}} B & \text{if there is a } B \in \mathcal{K}^{-1}: A \subseteq B, \\ \Omega & \text{otherwise.} \end{cases}$$

Then L is a closure operation over Ω and $\mathcal{K}_L = \mathcal{K}$. If L' is an arbitrary closure operation over Ω such that $\mathcal{K} = \mathcal{K}_{L'}$, than $L \leqq L'$ holds.

Proof. Clearly, L is a closure operation over Ω. Also it is obvious that for all $B \in \mathcal{K}^{-1}$ we have $L(B) = B$, i.e. $B \in Z(L)$. On the other hand, $\mathcal{K}^{-1}$ being a Sperner-system over Ω we obtain $M(L) = \mathcal{K}^{-1}$. By Lemma 2.1 $\mathcal{K}^{-1} = \mathcal{K}_L^{-1}$. Since $\mathcal{K}$ and $\mathcal{K}^{-1}$ are uniquely determined by each other $\mathcal{K}_L = \mathcal{K}$.

Suppose that L' is an arbitrary closure operation so that $\mathcal{K} = \mathcal{K}_{L'}$, it can be seen that $Z(L) = (\mathcal{K}^{-1})^+$. By Lemma 2.1, $M(L') = \mathcal{K}^{-1} = \mathcal{K}_{L'}^{-1}$. Consequently, $M(L') = M(L) = \mathcal{K}^{-1}$. Hence $Z(L) \subseteq Z(L')$ holds and by Theorem 2.3 we obtain $L \leqq L'$. Clearly, L is the closure operation for which $\mathcal{K} = \mathcal{K}_L$ and for any closure operation L' such that $\mathcal{K} = \mathcal{K}_{L'}$, and $L \neq L'$ we obtain $L < L'$. The theorem is proved. □

Corollary 2.4. Let $\mathcal{K}$ be a non-empty Sperner-system over Ω. Denote by V the set of all closure operations over Ω the minimal keys of which are exactly the elements of $\mathcal{K}$. Then L as constructed in Theorem 2.4 is the unique minimal element of the partially ordered set V for the ordering defined. □

Remark 2.1. In [6] we constructed an algorithm which computes the set of all antikeys of an arbitrary Sperner-system. By Theorem 2.4 and this algorithm we can explicitly construct the closure operation L for which $\mathcal{K}=\mathcal{K}_L$ to an arbitrarily given Sperner-system $\mathcal{K}$. □

The next remark shows that conversely, the set of all minimal keys of a given closure operation can be found.

Remark 2.2. In [5] we construct an algorithm which determines the set H such that $H^{-1}=\mathcal{K}$ for a given Sperner-system $\mathcal{K}$. Thus, if $\mathcal{K}$ is a set of antikeys then H is a set of minimal keys. Consequently, from a given closure operation L we can construct the family $M(L)$. By Lemma 2.1 $M(L)=\mathcal{K}_L^{-1}$ holds. From $M(L)$ we can determine the set of all minimal keys of L by this algorithm. □

Резюме

Одно из главных понятий теории реляционных баз данных является понятие функциональной зависимости. Статья изучает реляции которые представляют данную фамилию функциональных зависимостей, операции замыкания и системы Спернера. А также изучается связь между операциями замикания и минимальными ключами.

COMPUTER AND AUTOMATION INSTITUTE
HUNGARIAN ACADEMY OF SCIENCES
VICTOR HUGO U. 18—22
1123 BUDAPEST
HUNGARY

References

[1] W. W. Armstrong, Dependency structures of data base relationships. Information Processing 74, North-Holland (1974) 580—583.
[2] E. F. Codd, Relational model of data for large shared data banks. Communications of ACM, 13 (1970), 377—384.
[3] G. Burosch, J. Demetrovics, G. O. H. Katona, The poset of closures as a model of changing databases. Order, 4 (1987), 127—142.
[4] J. Demetrovics, On the equivalence of candidate keys with Sperner-systems. Acta Cybernetica 4 (1979), 3, 247—252.
[5] J. Demetrovics, V. D. Thi, Relations and minimal keys. Acta Cybernetica, Szeged 8 (1988), 3. 279—285.
[6] V. D. Thi, Minimal keys and antikeys. Acta Cybernetica, 7 (1986), 4, 361—371.

(Received Dec. 4, 1986)