Acta Cybernetica, Vol. 11, No. 4, Szeged, 1994

Some Remarks on Functional Dependencies in Relational Datamodels*

Vu Duc Thi †   Le Thi Thanh †

Abstract

The concept of minimal family is introduced. We prove that this family and family of functional dependencies (*FDs*) determine each other uniquely. A characterization of this family is presented.

We show that there is no polynomial time algorithm finding a minimal family from a given relation scheme. We prove that the time complexity of finding a minimal family from a given relation is exponential in the number of attributes.

Key Words and Phrases: relation, relational datamodel, functional dependency, relation scheme, closure, closed set, minimal generator, key, minimal key, antikey.

1 Introduction

The functional dependency introduced by E.F.Codd is one of important semantic constraints in the relational datamodel.

The family of FDs has been widely studied in the literature. In this paper we give a family of sets and show that it is determined uniquelyby family of FDs. This paper presents some results about computational problems related to this family.

Let us give some necessary definitions and results used in what follows.

Let $R = \{a_1, \ldots, a_n\}$ be a nonempty finite set of attributes. A functional dependency is a statement of the form $A \to B$, where $A, B \subseteq R$. The FD $A \to B$ holds in a relation $r = \{h_1, \ldots, h_m\}$ over R if $\forall h_i, h_j \in r$ we have $h_i(a) = h_j(a)$ for all $a \in A$ implies $h_i(b) = h_j(b)$ for all $b \in B$. We also say that r satisfies the FD $A \to B$.

Let F_r be a family of all FDs that hold in r. Then $F = F_r$ satisfies

(1) $A \to A \in F$,

(2) $(A \to B \in F,\ B \to C \in F) \Longrightarrow (A \to C \in F)$,

(3) $(A \to B \in F,\ A \subseteq C,\ D \subseteq B) \Longrightarrow (C \to D \in F)$,

*Research supported by Hungarian Foundation for Scientific Research Grant 2575.

†Computer and Automation Institute Hungarian Academy of Sciences, H-1111 Budapest, Lágymányosi u. 11. Hungary.

(4) $(A \to B \in F, C \to D \in F) \Longrightarrow (A \cup C \to B \cup D \in F)$.

A family of FDs satisfying (1)-(4) is called an f-family (sometimes it is called the full family) over R.

Clearly, F_r is an f-family over R. It is known [1] that if F is an arbitrary f-family, then there is a relation r over R such that $F_r = F$.

Given a family F of FDs, there exists a unique minimal f-family F^+ that contains F. It can be seen that F^+ contains all FDs which can be derived from F by the rules (1)-(4).

A relation scheme s is a pair $< R, F >$, where R is a set of attributes, and F is a set of FDs over R. Denote $A^+ = \{a: A \to \{a\} \in F^+\}$. A^+ is called the closure of A over s. It is clear that $A \to B \in F^+$ iff $B \subseteq A^+$.

Clearly, if $s =< R, F >$ is a relation scheme, then there is a relation r over R such that $F_r = F^+$ (see, [1]). Such a relation is called an Armstrong relation of s.

Let R be a nonempty finite set of attributes and $P(R)$ its power set. The mapping $H : P(R) \to P(R)$ is called a closure operation over R if for all $A, B \in P(R)$, the following conditions are satisfied :

(1) $A \subseteq H(A)$,

(2) $A \subseteq B$ implies $H(A) \subseteq H(B)$,

(3) $H(H(A)) = H(A)$.

Let $s =< R, F >$ be a relation scheme. Set $H_s(A) = \{a : A \to \{a\} \in F^+\}$, we can see that H_s is a closure operation over R.

Let r be a relation, $s =< R, F >$ be a relation scheme. Then A is a key of r (a key of s) if $A \to R \in F_r$ ($A \to R \in F^+$). A is a minimal key of $r(s)$ if A is a key of $r(s)$ and any proper subset of A is not a key of $r(s)$.

Denote $K_r(K_s)$ the set of all minimal keys of $r(s)$.

Clearly, K_r, K_s are Sperner systems over R, i.e. $A, B \in K_r(K_s)$ implies $A \not\subseteq B$.

Let K be a Sperner system over R. We define the set of antikeys of K, denoted by K^{-1}, as follows:

$$K^{-1} = \{A \subset R : (B \in K) \Longrightarrow (B \not\subseteq A) \text{ and } (A \subset C) \Longrightarrow (\exists B \in K)(B \subseteq C)\}.$$

It is easy to see that K^{-1} is also a Sperner system over R.

It is known [5] that if K is an arbitrary Sperner system over R, then there is a relation scheme s such that $K_s = K$.

In this paper we always assume that if a Sperner system plays the role of the set of minimal keys (antikeys), then this Sperner system is not empty (doesn't contain R). We consider the comparison of two attributes as an elementary step of algorithms. Thus, if we assume that subsets of R are represented as sorted lists of attributes, then a Boolean operation on two subsets of R requires at most $|R|$ elementary steps.

Let $L \subseteq P(R)$. L is called a meet-irreducible family over R (sometimes it is called a family of members which are not intersections of two other members) if $\forall A, B, C \in L$, then $A = B \cap C$ implies $A = B$ or $A = C$.

Let $I \subseteq P(R)$, $R \in I$, and $A, B \in I \Longrightarrow A \cap B \in I$. I is called a meet-semilattice over R. Let $M \subseteq P(R)$. Denote $M^+ = \{\cap M' : M' \subseteq M\}$. We say that M is a generator of I if $M^+ = I$. Note that $R \in M^+$ but not in M, by convention it is the intersection of the empty collection of sets.

Denote $N = \{A \in I : A \neq \cap\{A' \in I : A \subset A'\}\}$.

In [5] it is proved that N is the unique minimal generator of I.

It can be seen that N is a family of members which are not intersections of two other members.

Let H be a closure operation over R. Denote $Z(H) = \{A : H(A) = A\}$ and $N(H) = \{A \in Z(H) : A \neq \cap\{A' \in Z(H) : A \subset A'\}\}$. $Z(H)$ is called the family of closed sets of H. We say that $N(H)$ is the minimal generator of H.

It is shown [5] that if L is a meet-irreducible family then L is the minimal generator of some closure operation over R. It is known [1] that there is an one-to-one correspondence between these families and f-families.

Let r be a relation over R. Denote $E_r = \{E_{ij} : 1 \leq i < j \leq |r|\}$, where $E_{ij} = \{a \in R : h_i(a) = h_j(a)\}$. Then E_r is called the equality set of r.

Let $T_r = \{A \in P(R) : \exists E_{ij} = A, \not\exists E_{pq} : A \subset E_{pq}\}$. We say that T_r is the maximal equality system of r.

Let r be a relation and K a Sperner system over R. We say that r represents K if $K_r = K$.

The following theorem is known ([7])

Theorem 1.1 *Let K be a non-empty Sperner system and r a relation over R. Then r represents K iff $K^{-1} = T_r$, where T_r is the maximal equality system of r.*

In [6] we proved the following theorem.

Theorem 1.2 *Let $r = \{h_1, \ldots, h_m\}$ be a relation, and F an f-family over R. Then $F_r = F$ iff for every $A \subseteq R$*

$$H_F(A) = \begin{cases} \bigcap\limits_{A \subseteq E_{ij},} E_{ij} & \text{if } \exists E_{ij} \in E_r : A \subseteq E_{ij} \\ R & \text{otherwise,} \end{cases}$$

where $H_F(A) = \{a \in R : A \to \{a\} \in F\}$ and E_r is the equality set of r.

2 Results

In this section we introduce the concept of minimal family. We show that this family and family of FDs determine each other uniquely. We give some desirable properties of this family. We present some results about the relationship between this family, meet-semiattice and family of FDs.

Definition 2.1 *Let $Y \subseteq P(R) \times P(R)$. We say that Y is a minimal family over R if the following conditions are satisfied :*

(1) $\forall (A, B), (A', B') \in Y : A \subset B \subseteq R$, $A \subset A'$ implies $B \subset B'$, $A \subset B'$ implies $B \subseteq B'$.

(2) Put $R(Y) = \{B : (A, B) \in Y\}$. For each $B \in R(Y)$ and C such that $C \subset B$ and $\not\exists B' \in R(Y) : C \subset B' \subset B$, there is an $A \in L(B) : A \subseteq C$, where $L(B) = \{A : (A, B) \in Y\}$.

Remark 2.2 *(1.)* $R \in R(Y)$.

(2.) From $A \subset B'$ implies $B \subseteq B'$ there is no a $B' \in R(Y)$ such that $A \subset B' \subset B$ and $A = A'$ implies $B = B'$.

(3.) Because $A \subset A'$ implies $B \subset B'$ and $A = A'$ implies $B = B'$, we can be see that $L(B)$ is a Sperner system over R and by (2) $L(B) \neq \emptyset$.

Let I be a meet-semilattice over R.

Put $M^*(I) = \{(A, B) : \exists C \in I : A \subset C,\ A \neq \cap\{C : C \in I,\ A \subset C\}, B = \cap\{C : C \in I,\ A \subset C\}\}$.

Set $M(I) = \{(A, B) \in M^*(I) : \nexists(A', B) \in M^*(I) : A' \subset A\}$.

Note that if $C \in I$, then C is an one-term intersection. It is possible that $A = \emptyset$.

It can be seen that for any meet-semilattice I there is exactly one family $M(I)$.

Theorem 2.3 *Let I be a meet-semilattice over R. Then $M(I)$ is a minimal family over R.*

Conversely, if Y is a minimal family over R, then there is exactly one meet-semilattice I so that $M(I) = Y$, where $I = \{C \subseteq R : \forall(A, B) \in Y : A \subseteq C$ implies $B \subseteq C\}$.

Proof: Assume that I is a meet-semilattice over R. We have to show that $M(I)$ is a minimal family over R. It is obvious that $A \subset B \subseteq R$.

From $B' = \cap\{D : D \in I,\ A' \subset D\}\}$, we have $B' \subseteq D$. If $A \subset B'$, then $A \subset D$ and by $B = \cap\{C : C \in I : A \subset C\}$ we obtain $B \subseteq B'$. By $\nexists(A', B) \in M^*(I) : A' \subset A$ and from $A' \subset A \subset B$ implies $B' \subseteq B$ we can see that if $A' \subset A$ then $B' \subset B$. Thus, we obtain (1). Clearly, $L_I(B) = \{A : (A, B) \in M(I)\}$ is a Sperner system over R.

If there is a $B \in R(M(I))$ and D satisfying $D \subset B$ and $\forall B' \in R(M(I)) : D \subset B'$, $B' \subseteq B$ imply $B = B'$, then for all $A \in L_I(B) : A \not\subseteq D(*)$.

It can be seen that $D \neq \cap\{C : C \in I, D \subset C\}$ and $B = \cap\{C : C \in I, D \subset C\}$.

If $L_I(B) \cup D$ is a Sperner system over R, then by definition of $M(I)$ we have $D \in L_I(B)$. From (*) this is a contradiction.

If there exists an $A \in L_I(B) : D \subset A$, then this conflicts with the definition of $M(I)$. Thus, we have (2) in Definition 2.1. Consequently, $M(I)$ is a minimal family over R.

Conversely, Y is a minimal family over R. Clearly, I is a meet-semilattice over R. It is obvious that $(A, B) \in Y$ implies $A \notin I$.

Now we have to prove that $M(I) = Y$. Assume that $(A, B) \in Y$. By (1) in Definition 2.1 $\forall(A', B') \in Y : A' \subset B$ implies $B' \subseteq B$. From this and definition of I we obtain $B \in I$.

According to definition of I there is no $C \in I$ such that $A \subset C \subset B$. On the other hand, $A \subset B$ and B is an intersection of Cs, where $C \in I$, $A \subset C$. Thus, $B = \cap\{C : C \in I, A \subset C\}$ and $A \neq \cap\{C : C \in I, A \subset C\}$. Hence, $(A, B) \in M^*(I)$ holds.

Clearly, if $A = \emptyset$ then $(A, B) \in M(I)$. Assume that $A \neq \emptyset$ and $(A', B) \in M^*(I)$. It is obvious that by the definition of $M^*(I)$ $A' \subset B$ and $\nexists B' : A' \subset B' \subset B$. By (2) in Definition 2.1 there is an $A'' \in L(B) : A'' \subseteq A'$. Because $L(B)$ is a Sperner system over R and $A \in L(B)$ we have $A' \not\subset A$. Thus, $(A, B) \in M(N)$ holds.

Suppose that $A \subset R$ and $A \notin I$. Based on the above proof, $B \in R(Y)$ implies $B \in I$. Clearly, $R \in R(Y)$. Consequently, for A there is a $B \in R(Y)$ such that $A \subset B$ (**). We choose a set B so that $|B|$ is minimal for (**), i.e. $\nexists B' \in R(Y) : A \subset B' \subset B$. According to (2) in Definition 2.1 there exists an $A' \in L(B) : A' \subseteq A$. If there is $C \in I : A \subset C \subset B$, then $A' \subset C \subset B$. This conflicts with the definition of I. Consequently, for all $C \in I$ and $C \neq B$, $A \subset C$ implies $B \subset C$. From this and according to the definition of $M^*(I)$ $(A, B) \in M^*(I)$ implies $B \in R(Y)$.

Assume that $(A, B) \in M(I)$. By the above proof, $B \in R(Y)$ holds. We consider the set $L(B) = \{A' : (A', B) \in Y\}$. According to definition $M(I)$ we have $A \subset B$ and $\nexists B' \in R(Y) : A \subset B' \subset B$. By (2) in Definition 2.1 there is an $A' \in L(B)$ such that $A' \subseteq A$. If $A' \subset A$, then according to the above proof $(A', B) \in Y$ implies $(A', B) \in M(N)$. $A' \subset A$ contradicts the definition of $M(N)$. Thus, $A' = A$ holds. Consequently, we obtain $(A, B) \in Y$.

Suppose that there is a meet-semilattice I' such that $M(I') = Y$. We have to show that $I = I'$. By definition of $M(I')$ $E \in I'$ implies $E \in I$. Thus, $I' \subseteq I$ holds. Suppose that there is a $D \in I$ and $D \notin I'$. According to the definition of meet-semilattice $R \in I'$. Put $D" = \cap\{E \in I' : D \subset E\}$. By $D \notin I'$ we have $D \subset D"$. According to $M^*(I')$ $(D, D") \in M^*(I')$. From definition of $M(I')$ there is a $D' : D' \subseteq D$ and $(D', D") \in M(I')$. Thus, $D' \subseteq D \subset D"$ holds. This conflicts with the fact that $D \in I$. Hence, $I = I'$ holds. □

It is known [1] that there is an one-to-one correspondence between families of FDs and meet-semilattices and by Theorem 2.3 we obtain the following.

Proposition 2.4 *There is an one-to-one correspondence between minimal families and families of FDs.*

Because there are one-to-one correspondences between meet-irredundant families, closure operations and families of FDs, we also have the following.

Proposition 2.5 *There are one-to-one correspondences between minimal families, meet-irredundant families and closure operations.*

Remark 2.6 *Let $s =< R, F >$ be a relation scheme over R. A functional dependency $A \rightarrow B \in F^+$ is called basic of s if*

(1) $A \subset B$,

(2) $\nexists A' : A' \subset A$ *and* $A' \rightarrow B \in F^+$,

(3) $\nexists B' : B \subset B'$ *and* $A \rightarrow B' \in F^+$,

Denote by $B(s)$ the set of all basic FDs of s.

If a relation scheme is changed to a relation we have a basic functional dependency of r. Denote the set of all basic FDs of r by $B(r)$.

It can be seen that the set $\{A \rightarrow R : A \in K_s\}$ is a subset of $B(s)$.

Remark 2.7 *Let $s =< R, F >$ be a relation scheme over R. Put $Z(s) = \{A : A^+ = A\}$. $Z(s)$ is a meet-semilattice over R. $M(Z(s))$ is called the minimal family of s. According to definitions of $M(I)$ and $B(s)$ we can see that $M(Z(s)) = \{(A, B) : A \rightarrow B \in B(s)\}$.*

It is known [17] that there is no a polynomial time algorithm finding a set of all minimal keys of a given relation scheme. From this and by Remark 2.6 we have the following corollary.

Corollary 2.8 *Let $s =< R, F >$ be a relation scheme over R. There is no a polynomial time algorithm to find the minimal family of s.*

Definition 2.9 *Let R be a relation over R and F_r a family of all FDs that hold in r. Put $A_r^+ = \{a : A \to \{a\} \in F_r\}$. Set $Z_r = \{A : A = A_r^+\}$. Then $M(Z_r)$ is called the minimal family of r.*

It is easy to see that the set $\{A \to R : A \in K_r\}$ is a subset of $B(r)$.

We construct a following exponential time algorithm finding a minimal family of a given relation.

In relation scheme $s = < R, F >$, a functional dependency $A \to B \in F$ is called redundant if either $A = B$ or there is $C \to B \in F$ such that $C \subseteq A$.

Algorithm 2.10 *(Finding a minimal family of r)*

(Input:) *a relation $r = \{h_1, \dots, h_m\}$ over R.*

(Output:) *a minimal family of r.*

(Step 1:) *Find the equality set $E_r = \{E_{ij} : 1 \leq i < j \leq m\}$.*

(Step 2:) *Find the minimal generator N, where $N = \{A \in E_r : A \neq \cap\{B \in E_r : A \subset B\}\}$. Denote elements of N by $A_1, \dots, A_t$.*

(Step 3: *For every $B \subseteq R$ if there is an A_i $(1 \leq i \leq t)$ such that $B \subseteq A_i$, then compute $C = \bigcap_{B \subseteq A_i} A_i$ and set $B \to C$. In the converse case set $B \to R$. Denote by T the set of all such functional dependencies*

(Step 4:) *Set $F = T - Q$, where $Q = \{X \to Y \in T : X \to Y$ is a redundant functional dependency $\}$.*

(Step 5:) *Put $M(Z_r) = \{(B, C) : B \to C \in F\}$.*

According to Theorem 1.2 and definition of $M(Z_r)$, Algorithm 2.10 finds a minimal family of r.

It can be seen that the time complexity of Algorithm 2.10 is exponential in the number of attributes.

Let $s = < R, F >$ be a relation scheme over R. We say that s is in Boyce-Codd normal form (BCNF) if $A \to \{a\} \notin F^+$ for $A^+ \neq R$, $a \notin A$.

If a relation scheme is changed to a relation we have the definition of BCNF for relation.

Proposition 2.11 *Given a BCNF relation r over R. The time complexity of finding a minimal family of r is exponential in the number of elements of R.*

Proof: From a given BCNF relation r we use Algorithm 2.10 to construct the minimal family of r. By definition of BCNF, we obtain

$M(Z_r) = \{(B, C) : B \to C \in F\} = \{(B, R) : B \in K_r\}$.

Let us take a partition $R = X_1 \cup \dots \cup X_m \cup W$, where $|R| = n, m = [n/3]$, and $|X_i| = 3$ $(1 \leq i \leq m)$.

Set $M = (K^{-1})^{-1}$, i.e. K^{-1} is a set of minimal keys of M, we have

$M = \{C\colon |C| = n - 3, C \cap X_i = \emptyset$ for some $i\}$ if $|W| = 0$,

$M = \{C\colon |C| = n - 3, C \cap X_i = \emptyset$ for some i $(1 \leq i \leq m - 1)$ or $|C| = n - 4, C \cap (X_m \cup W) = \emptyset\}$ if $|W| = 1$,

$M = \{C\colon |C| = n-3, C \cap X_i = \emptyset$ for some i $(1 \leq i \leq m)$ or $|C| = n-2, C \cap W = \emptyset\}$ if $|W| = 2$.

It is clear that $3^{[n/4]} < |K^{-1}|, |M| \leq m+1$.
Denote elements of M by $C_1, \ldots, C_t$.
Construct a relation $r = \{h_0, h_1, \ldots, h_t\}$ as follows:
For all $a \in R$ $h_0(a) = 0$, for $i = 1, \ldots, t$

$$h_i(a) = \begin{cases} 0 & \text{if } a \in C_i \\ i & \text{otherwise.} \end{cases}$$

Clearly, $|r| < |R|$ holds. According to Theorem 1.1 M is the set of antikeys of r and K^{-1} is the set of minimal keys of r. From definition of BCNF, we can see that $M(Z_r) = \{(B, R) : B \in K^{-1}\}$.

Thus, we can construct a relation r in which the number of rows of r is less than $|R|$, but the number of elements of $M(Z_r)$ is exponential in the number of attributes. □

Since the class of BCNF relations is a special subfamily of the family of relations over R, the next corollary is obvious.

Corollary 2.12 *The time complexity of finding a minimal family of a given relation r is exponential in the number of attributes.*

References

[1] Armstrong W.W., Dependency Structures of Database Relationships, *Information Processing 74, Holland Publ. Co.* (1974), 580-583.

[2] Beeri C., Bernstein P.A., Computational problems related to the design of normal form relational schemas, *ACM Trans. on Database Syst.* **4** (1979), 30-59.

[3] Beeri C., Dowd M., Fagin R., Staman R., On the Structure of Armstrong Relations for Functional Dependencies, *J. ACM.* **31** (1984), 30-46.

[4] Békéssy A., Demetrovics J., Contribution to the theory of data base relations, *Discrete Math.* **27** (1979), 1-10.

[5] Demetrovics J., Logical and structural investigation on relational datamodel *MTA-SZTAKI Tanulmányok, Budapest, Hungarian.* **114** (1980), 1-97.

[6] Demetrovics J., Thi V.D., Some results about functional dependencies, *Acta Cybernetica.* **8** (1988), 273-278.

[7] Demetrovics J., Thi V.D., Relations and minimal keys, *Acta Cybernetica.* **8** (1988), 279-285.

[8] Demetrovics J., Thi V.D., On Keys in the Relational Datamodel, *J. Inform. Process. Cybernet.* **24** (1988), 515-519.

[9] Demetrovics J., Thi V.D., On algorithms for generating Armstrong relations and inferring functional dependencies in the relational datamodel, Computers and Mathematics with Applications. To appear.

[10] Demetrovics J., Thi V.D., Armstrong Relation, Functional Dependencies and Strong Dependencies, *Comput. and AI.* To appear.

[11] Mannila H., Raiha K.J., Design by Example: An Application of Armstrong relations, *J. Comput. Syst. Scien.* **33** (1986), 126-141.

[12] Osborn S.L., Testing for existence of a covering Boyce-Codd normal form, *Infor. Proc. Lett.* **8** (1979), 11-14.

[13] Thi V.D., Investigation on Combinatorial Characterisations Related to Functional Dependency in the Relational Datamodel, *MTA-SZTAKI Tanulmányok, Budapest, Hungarian. (Ph.D. Dissertation).* **191** (1986), 1-157.

[14] Thi V.D., Minimal keys and Antikeys, *Acta Cybernetica.* **7** (1986), 361-371.

[15] Thi V.D., On the Antikeys in the Relational datamodel, *Alkalmazott Matematikai Lapok.* (in Hungarian) **12** (1986), 111-124.

[16] Thi V.D., Logical Dependencies and Irredundant Relations, *Computers and Artificial Intelligence.* **7** (1988), 165-184.

[17] Yu C.T., Johnson D.T., On the complexity of finding the set of candidate keys for a given set of functional dependencies, *Inform. Proc. Letters.* **5** (1976), 100-101.

Received July 29, 1993