Tạp chí Tin học và Điều khiển học, T.17, S.3 (2001), 48-52

ON THE MINIMAL FAMILY

VU DUC THI

Abstract. Equivalent descriptions of family of functional dependencies (FDs) play important role in the design and implementation of the relational datamodel. In this paper, we introduce the new concept of minimal family. We prove that these families are equivalent descriptions of family of FDs.

Tóm tắt. Những sự mô tả tương đương của họ các phụ thuộc hàm có vai trò quan trọng trong việc thiết kế và việc thực hiện mô hình dữ liệu quan hệ. Trong bài này, chúng tôi trình bày khái niệm mới về họ cực tiểu. Chúng tôi chứng minh rằng những họ này là những mô tả tương đương của họ các phụ thuộc hàm.

1. INTRODUCTION

It is known [1, 4-8, 14, 17] that closure operations, meet-semilattices, families of members which are not intersections of two other members give the equivalent descriptions of FDs, i.e. they and family of FDs determine each other uniquely. These equivalent descriptions were successfully applied to find many desirable properties of functional dependency. Equivalent descriptions of family of FDs have been widely studied in the literature. In this paper, we investigate the minimal family. We show that it is equivalent description of family of FDs.

Let us give some necessary definitions and results that are used in next section. The concepts given in this section can be found in [1, 2, 4, 6, 7, 8, 17].

Let $R = \{a_1, \ldots, a_n\}$ be a nonempty finite set of atributes. A functional dependency (FD) 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) $(A \to B \in F,\ A \subseteq C,\ D \subseteq B) \Longrightarrow (C \to D \in F)$,

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

Let r be a relation, $s = \langle R, F\rangle$ be a relation scheme. Then A is a key of r (a key of s) if $A \rightarrow R \in F_r$ ($A \rightarrow R \in F^+$). 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. $A, B \in K_r$ 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 $A = A$ 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 set s 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 $T_r = \{A \in P(R) : \exists E_{ij} = A,\ \nexists E_{pq} : A \subset E_{pq}\}$. 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.*

Let $s = \langle R, F\rangle$ be a relation scheme over R, K_s is a set of all minimal keys of s. Denote by K_s^{-1} the set of all antikeys of s. From Theorem 1.1 we obtain the following corollary.

Corollary 1.2. *Let $s = \langle R, F\rangle$ be a relation scheme and r a relation over R. We say that r represents s if $K_r = K_s$. Then r represents s iff $K_s^{-1} = T_r$, where T_r is the maximal equality system of r.*

In [6] we proved the following theorem.

Theorem 1.3. *Let $r = \{h_1, ..., h_m\}$ be a relation, and F an f-family over R. Then $F_r = F$ iff for every $A \subseteq R$*

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

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

Theorem 1.4. [3] *Let $K = \{K_1, ..., K_m\}$ be a Sperner system over R. Set $s = \langle R, F \rangle$ with $F = \{K_1 \to R, ..., K_m \to R\}$. Then $K_s = K$.*

2. MINIMAL FAMILY

In this section we introduce the new concept of minimal family. We show that this family and family of FDs determine each other uniquely.

Now we introduce the following concept.

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 $\nexists B' \in R(Y) : C \subset B' \subset B$, there is an $A \in L(B) : A \subseteq C$, where $L(B) = \{A : (A, B) \in Y\}$.

Remark.

- $R \in R(Y)$.
- From $A \subset B'$ implies $B \subset B'$ there is no a $B' \in R(Y)$ such that $A \subset B' \subset B$ and $A = A'$ implies $B = B'$.
- 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-tern 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 2.2. *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 C, 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 $(A', E) \in Y$ 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. The theorem is proved.

It is known [1] that there is an one-to-one correspondence between families of FDs and meet-semilattices and by Theorem 2.2 we obtain the following.

Proposition 2.3. *There is an one-to-one correspondence between minimal families and families of FDs.*

REFERENCES

[1] Armstrong W. W., *Dependency Structures of Database Relationships*, Information Processing 74, Holland Publ. Co., **74** (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** (1) (1979) 30–59.

[3] Beeri C., Dowd M., Fagin R., Staman R., On the structure of Armstrong relations for functional dependencies, *J. ACM* **31** (1) (1984) 30–46.

[4] Bekessy A., Demetrovics J., Contribution to the theory of database relations, *Discrete Math.* **27** (1979) 1–10.

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

[6] Demetrovics J., Thi V. D., Some results about functional dependencies, *Acta Cybernetica* **8** (3) (1988) 273–278.

[7] Demetrovics J., Thi V. D., Relations and minimal keys, *Acta Cybernetica* **8** (3) (1988) 279–285.

[8] Demetrovics J., Thi V. D., On keys in the relational datamodel, *Inform. Process. Cybern. EIK* **24** (10) (1988) 515–519.

[9] Demetrovics J., Thi V. D., On algorithm for generating Armstrong relations and inferring functional dependencies in the relational datamodel, *Computers and Mathematics with Applications* **26** (4) (1993) 43–55.

[10] Demetrovics J., Thi V. D., Armtrong relation, functional dependencies and strong dependencies, *Comput. and AI* **14** (3) (1995) 279–298.

[11] Demetrovics J., Thi V. D., Some computational problems related to Byyce–Codd normal form, *Annales Univ. Sci. Budapest, Sect. Comp.* **19** (2000) 119–132.

[12] Mannila H., Raiha K. J., Design by example: an application of Armstrong relations, *J. Comput. Syst. Scien.* **33** (1986) 126–141.

[13] Osborn S. L., Testing for existence of a covering Boyce–Codd normal form, *Infor. Proc. Lett.* **8** (1) (1979) 11–14.

[14] Thi V. D. Investigations on combinatorial characterizations related to functional dependencies in the relational datamodel, *MTA-SZTAKI Tanulmanyok, Budapest* **191** (1986) 1–157, Ph.D. Dissertation (Hungarian).

[15] Thi V. D. Minimal keys and antikeys, *Acta Cybernetica* **7** (4) (1986) 361–371.

[16] Thi V. D. On the antikeys in the relational datamodel, *Alkalmazott Matematikai Lapok* **12** (1986) 111–124 (Hungarian).

[17] Thi V. D., Logical dependencies and irredundant relations, *Computers and Artificial Intelligence* **7** (2) (1988) 165–184.

[18] Yu C. T., Johnson D. T., On the complexity of finding the set of candidate keys for a given set of functional dependencies, *IPL* **5** (4) (1976) 100–101.

Received May 26, 2000

Institute of Information Technology, NCST of Vietnam