Tạp chí Tin học và Điều khiển học, T.14, S.1 (1998), (1 - 8)

SOME PROBLEMS ON THE FUNCTIONAL DEPENDENCY RELATED TO ARMSTRONG RELATIONS IN THE RELATIONAL DATAMODEL[1]

VU DUC THI [(1)]

Astract. In the paper, we give some results to combinatorial algorithms for functional dependency (FD for short) connecting the construction of Armstrong relations, relation schemes, the FD-relation implication problem, and the FD-relation equivalence problem. These algorithms play important roles in logical and structural investigations of the relational data model. Now, they are know to have exponential complexity. However, in this paper we show that if relations and relations schemes satisfy certain additional properties, then above problems are solved in polynomial time.

Key Words and Phrases: relation, relational datamodel, functional dependency, relation scheme, generating Armstrong relation, dependency inference, FD-relation implication scheme problem, FD-relation equivalence problem, minimal Armstrong relation, closed set, minimal generator, key, minimal key, antikey.

1. INTRODUCTION

Now we start with some necessary definitions, and in the nex sections we formulate our results.

Definition 1.1 Let $R = \{h_1, \ldots h_n\}$ be a relation over U, and $A, B \subseteq U$.

Then we say that B functionally depends on A in R (denoted $A \xrightarrow[R]{f} >$) iff

$(\forall h_i, h_j \in A)\ (h_i(a) = h_j(a)) \Rightarrow (\forall b \in B)\ (h_i(b) = h_j(b))$

Let $F_R = \{(A, B)\colon A, B \subseteq U, A \xrightarrow[R]{f} > B\}$, F_R is called the full family of functional dependencies of R. Where we write (A, B) or A- B for $A \xrightarrow[R]{f} B$ when R, f are clear from the context.

Definition 1.2. A functional dependency over U is a statement of the form $A \rightarrow B$, where A,

$B \subseteq U$. The FD A - B holds in a relation R if $A \xrightarrow[R]{+} B$. We also say that R statisfies the FD A - B.

---

[1] Institute of Information Technology National Centre for Natural Sciences and Technology of Vietnam P.O.Box 626, Boho Hanoi 10000, Vietnam

Definition 1.3. Let U be a finite set, and denote P(U) its power set.

Let $Y \subseteq P(U) \times P(U)$. We say that Y is an f-family over U iff for all $A, B, C, D \subseteq U$

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

(2) $(A, B) \subseteq Y, (B, C) \in Y \Rightarrow (A, C) \in Y$,

(3) $(A, B) \subseteq Y, A \subseteq C, D \subseteq B \Rightarrow (C, D) \in Y$,

(4) $(A, B) \subseteq Y, (C, D) \in Y \Rightarrow (A \cup C, B \cup D) \in Y$.

Clearly, F_R as above is an f-family over U.

It is known (1) that if Y is an arbitrary f-family, then there is a relation R over U such that $F_R = Y$.

Definition 1.4. A relation scheme S is a pair <U, F>. Where U is a set of attribites, and F is a set of FDs over U. Let F^+ be a set of all FDs that can be derived from F by the rules in definition 1.3. Denote $A^+ = \{a: A \rightarrow \{a\} \in F^+\}$. A^+ is called the closure of A over S. It is clear that $A \rightarrow B \in F^+$ iff $B \subseteq A^+$.

Clearly, in (1) if S = <U, F> is a relation scheme, then there is a relation R over U such that $F_R = F^+$. Such a relation is called an Armstrong relation of S.

Definition 1.5. Let be a relation, S = <U, F> be a relation scheme, Y be can f-family over U, and $A \subseteq U$. Then A is a key of R (a key of S, a key of Y) if $A \overset{f}{\underset{R}{\rightarrow}} U$ $(A \rightarrow U \in F^+$, $(A, U) \in Y)$. A is a minial key of R(S, Y) if A is a key of R(S,Y), and any proper subset of A is not a key of R(S, Y). Denote K_R, (K_S, K_Y) the set of all minial keys of R(S, Y).

Clearly, K_R, K_S, K_Y are Sperner systems over U.

It is known (4) that if K is an arbitrary Sperner system then there is a relation R such that $K_R = K$.

Definition 1.6. Let K be a Sperner system over U. We define the set of antikeys of K, denote by K^{-1}, as follows:

$K^{-1} = \{A \subset U: (B \in K) \Rightarrow (B \not\subset A)$ and $(A \subset C) \Rightarrow (EB \in K)\ (B \subseteq C)\}$

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

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 (does't countain U). We also regard the comparision of two attributes to be the elementary step of algorithms. Thus, if we assume that subsets of U are represented as sorted lists of attribites, then a Boolean operation on two subsets of requires at most $|U|$ elementary steps.

Definition 1.7. Let $I \subseteq P(U)$, $U \in I$, and $A, B \in I \Rightarrow A \cap B \in I$. Let $M \subseteq P(U)$. Denote $M^+ = \{\cap M': M' \subseteq M\}$. We say that M is a generator of I iff $M^+ = I$. Note that $U \in M^+$ but not in M, since 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 (4) it is proved that N is the unique minimal generator of I. Thus, for any generator N of I we obtain $N \subseteq N'$.

Definition 1.8. Let R be a relation over U, and E_R the equality set of R, i.e $E_R = \{E_{ij}: 1 \leq i < j \leq |R|\}$, where $E_{ij} = \{a \in U: h_i(a) = h_j(a)\}$. Let $T_R = \{A \in P(U): \exists E_{ij} = A, \exists E_{pq}: A \subset E_{pq}\}$. Then T_R is called the maximal equality system of R.

Definition 1.9. Let R be a relation, and K a Sperner system over U. We say R represents K iff $K^{-1} = T_R$, where T_R is the maximal equanlity system of R.

Remark 1.11. Let us take partition $U = X_1U, ..., UX_m \cup W$, where m = [n/3], and $|X_i| = 3$ $(1 \leq i \leq m)$

We set

$H = \{A: |A \cap X_i| = 1, \forall i\}$ if $|W| = 0$

$H = \{A: |A \cap X_i| = 1, (1 \leq i \leq m - 1)$ and $|A \cap (X_m \cup W)| = 1\}$ if $|W| = 1$,

$H = \{A: |A \cap X_i| = 1, (1 \leq i \leq m)$ and $|A \cap W|$ if $|W| = 2$.

If set $K = H^{-1}$, i.e. H is a set of minimal keys of K, then we have

$K = \{C: = n - 3, C \cap X_4 = \varnothing$ for some i$\}$ if $|W| = 0$

$K = \{C: |C| = n - 3, c \cap X_i = \varnothing$ for some i $(1 \leq i \leq m - 1)$ or $|C| = n - 4, C \cap (X_m \cup W) = \varnothing\}$ if $|W| = 1$,

$K = \{C: |C| = n - 3; C \cap X_i = \varnothing$ for some i $(1 \leq i \leq m)$ or $|C| = n - 2, C \cap W = \varnothing\}$ if $|W| = 2$.

It is clear that $3^{(n/3)} < |H|, |K| \leq m + 1$.

Denote elements of K by C_1, ... C_i. Construct a relation $R = \{h_o, h_1, ... h_i\}$ as follows: For all $a \in U$ $h_0(a) = 0$, for i = 1, ..., t, $h_i(a) = 0$ if $a \in C_i$, in the converse case $h_i(a) = i$. Clearly, $|R| < |U|$ holds. According to Theorem 1.10 K is the set of antikeys of R and H is the set of minimal keys of R.

Thus, we always can construct a relation R in which the number of rows of R is less than |U|, but the number of elements of H is exponential in the number of R. On the other hand, it is known [7] that there is an exponential time algorithrn which finds a set of minimal keys of given relation. Consequently, the time complexity of finding a set of minimal keys of a given relation R is exponential in the size of R.

2. FUNCTIONAL DEPENDENCY

The following problems play important roles in the logical and structural investigation of the relational data model both in practive and design theory.

(1) Constructing Armstrong Relation:

Given a relation scheme $S = <U, F>$ construct a relation R such that $F_R = F^+$.

(2) Dependency Inference Problem:

Given a relation R over U, construct a relation scheme $S = <U, F>$ such that $F^+ = F_R$.

(3) FD- Relation Implication Problem:

Givena a relation R and a relation scheme $S = <U, F>$ over U, decide whether $F_R \subseteq F^+$.

(4) FD-Relation Equivalence Problem:

Given a relation R and a relation scheme $S = <U, F>$ over U, decide whether $R_R = F^+$.

By Remark 1.11 it is known that finding a set of all miniall keys of a given relation has exponential complexity. On the other hand, it is known (see [8]) that the problem of finding all minial keys of a given relation can be polynomially transformed to the dependency inference problem. Thus, the latter problem is also inherently difficult. The problem of constructing Armstrong relation is known to be inherently difficult. The FD - relation implication problem is co-NP-complete (see [8]). Finally, it is still unknown that the time complexity of the FD-relation equivalence problem is polynomial or not. However, it is easy to see that there is an exponential time algorithm to solve this problem. We give a special class of relations and relation schemes in which the complexities of the above problems are polynomial.

For an f-family F over U the following closure operation can be introduced on P(U) (c.f. [2]): $H_F(A) = \{a \in U: A \to \{a\} \in F\}$

Where the notation $A \to B$ is used as a synonym of (A, B). We denote the family of the closed sets with respect to the closure H_F by Z(F), i.e $Z(F) = \{A \subseteq U: H_F(A) = A\}$.

It is easy to see that $U, \varnothing \in Z(F)$ and $A, B \in Z(F) \Rightarrow A \cap B \in Z(F)$.

Theorem 2.1.[2] Let F_1, F_2 be two f-family over U. Then $F_1 = F_2$ iff $Z(F_1) = Z(F_2)$.

It is clear that if $S = <U, F>$ is a relation schem then F^+ is an f-family over U.

Theorem 2.2 [2] Let K be a Sperner-system and $S = <U, F>$ be a relation scheme over U. Then $K_S = K$ iff

$\{U\} \cup K^{-1} \subseteq A(F^+) \subseteq \{U\} \cup G(K^{-1})$,

where $G(K^{-1}) = \{A \subseteq U: \exists B \in K^{-1}: A \subseteq B\}$.

According to [2], clearly.

Theorem 2.3. Let $K = \{K_1, ..., K_i\}$ be a Sperner-system over U Consider the relation scheme $S = (U, F)$ with $F = \{K \to U, ..., K_i \to U\}$.

Then $K_S = K$, and $Z(F^+)$, and $Z(K^+) = G(K^{-1}_S) \cup \{U\}$.

Theorem 2.4. [2] Let K be Sperner system over U. We stay that K is saturated if for any $A \notin K$ $\{A\} \cup K$ is not a Sperner system. If K is a saturated Sperner system then $K = F_R$ uniquely determines F, where K_F is the set of all minimal keys of an f-family F. The following is an example of a saturated Sperner-system K such that K^{-1} is not saturared.

Example 2.5. Let $U = \{1, 2, 3, 4, 5, 6\}$ and

$K = \{(1, 2), (3, 4), (5, 6), (1, 3, 5), (1, 3, 6), (1, 4, 5), (1, 4, 6) (2, 3, 5), (2, 3, 6), (2, 4, 5), (2, 4, 6)\}$.

It is easy to see that K is a saturared sperner-system but K^{-1} is not saturared since

$K^{-1} = \{(1, 3), (1, 4), (1, 5), (1, 6), (2, 3), (2, 4), (2, 5), (2, 6), (3, 5), (3, 6), (4, 5) (4, 6)\}$, and

$K^{-1} \cup \{(1, 2)\}$ is a Sperner-system over U.

It is clear that there also exists a Sperner-system K such that K is not saturared but K^{-1} is.

Definition 2.6. A Sperner-system K over U is called embedded iff for every $A \in K$ there exists a. $B \in H$ such that $A \subset B$. K is called is called completely embedded iff K is embedded and for every $B \in H$ there is an $A \in K$ such that $A \subset B$, where H denotes the Sperner-system for which $H^{-1} = K$.

Proposition 2.7. [9] Let K be a Sperner-system over U. Then K is saturared iff K^{-1} is embedded.

Definition 2.8. A relation scheme $S = <U, F>$ with $F = \{K_1 \rightarrow U, ..., K_t \rightarrow U\}$ where $(K_1, ..., K_t)$ is a Sperner-system over U is called a k-relation scheme.

It can be seen that if $S = <U, F>$ is in Boyce-Codd normal form then using the algorithm for finding a minimal cover we can construct a k-relation scheme $S = <U, F'>$ in polymial time such that $F^+ = F'^+$, (see [8]). Conversely, it is clear that a k-relation scheme is in BCNF.

In [7] we constructed an algorithm to find K^{-1}_R in polynomial time for a given relation R.

Theorem 2.9. Let $S = <U, F>$ a k-relation scheme with given as input.

(1) Assuming K is saturared, an Armstrong relation for S, i.e. a relation R with $F_R = F^+$ can be constructed in polynomial time.

(2) Assuming K^{-1}_R is completely embedded, the dependency inference problem for R can be solved in polynomial time, i.e. a relation scheme $S = <U, F>$ with $F^+ = F_R$ can be constructed in polynomial time.

(3) Assuming either K is saturared or K^{-1}_R is completely embedded, the FD-relation implication problem and the FD-relation equivalence problem for S and R can be solved in polynomial time.

Proof: (1) Assume that K is saturared. Let us define the following families. $Q = \{K - \{a_i\}: a_i \in K, 1 \le i \le t\}$

and

$P = \{A \in Q: H_{F^+}(A) = A$ and $H_{F^+}(A \cup \{a\}) = U\ \forall a \in U - A\}$.

Assuming that $P = \{A_1, ..., A_m\}$, let us define the relation.

$R = \{h_0, h_1, ..., h_m\}$ as follows:

For all $a \in U$, $h_0(a) = 0$,

For $i = 1,... m$

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

It can be seen that Q, P, R can be constructed in polynomial time. Proposition 2.7 implies $P = K^{-1}$, where $K = \{K_1, ..., K_t\}$

Based on Theorem 1.10 and Theorem 2.4. $F^+ = F_R$ holds.

(2) By the algorithm in [7] the maximal equality system TT of R can be constructed in polynomial time. Theorem 1.10 implies $T_R = K^{-1}{}_R$. Let us derine $M = \{T \cup \{a\}: A \in U, T \in T_R\}$ and

$L = \{E \in M: H_{F_R}(E) = U, \forall a \in E, H_{F_R}(E - a) \neq U\}$

By Proposition 2.7, and the definitions of completely embedded Sperner-system we have $L = K_R$.

Assuming $L = \{E_1, ... E_s\}$, define $S = <U, F>$ with $F = \{E_1 \to U, ... E_s \to U\}$.

It can be seen that M, L, S can be constructed in polynomial time.

Theorems 2.3, 2.4, and Proposition 2.7 imply $F^+ = F_R$.

(3) If K is saturared, then by (1) by we can consrtruct an Armstrong relation. R for S. We can compare F_R. with F_R.

If $K^{-1}{}_R$ is completely embedded, then by (2) we can construct a relation scheme $S' = <U, F'>$ such that $F'^+ = F_R$. We have then to compare F and F'^+. It is easy to see that this can be done in polynomial time. The theorem is proved.

Is is known that the number of minial keys of a given relation can be an exponential function of the size of the relation. But in many cases there are only polynomially many minimal keys and they can found in polynomial. On the other hand, according to Theorem 2.4 it can be seem that the next corollary is an easy consequence of Theorem 2.9.

Corollary 2.10 Let $S = <U, F>$ be a relation scheme, R a relation over U. Then

If the set of all minial keys of S is saturared, then problem 1 can be solved in polynomial time.

If the set of all antikeys of R is completely embedded, then problem 2 has polynomial complexity.

If either the set of all minimal keys of S is saturated or the set of all antikeys of R is completely embedded, then problems 3,4 are solved in polynomial time.

Definition 2.11. Let K_1 and K_2 are Sperner-system over U. We set

$K = K_1 \cup K_2$ and $T_R = (A \in K: \not\in B \in K: A \subset B\}$

We say they the union $K = K_1 \cup K_2$ is pseudo-saturated if T_R is a saturared Sperner system over U.

Definition 2.12: (FD-relation key-equivalence problem)

Theorem 2.13. Let $S = <U, F>$ be a relation scheme and R a relation over U. $K_S(K^{-1}_R)$ is the set of all minial keys of S (the set of antikeys of R), and K_S is computer in polymial time in the size S.
Then if $K = K_S \cup K^{-1}_R$ is pseudo-saturated, then the FD-relation key-equivalence problem is solved in polynomial time in the sizes of S, and R.

Proof: It is known [7] that K^{-1}_R is computer in polynomial time in size of R. We assume that $K_S = \{A_1, ..., A_p\}$, and $K^{-1}_R = \{B_1,... B_q\}$.

If there is an A_i $(1 \le i \le p)$ such that $A_i \subseteq B_j$ $(1 \le j \le q)$ then $K_S \neq K_R$. Consequently, we can assume that $A_i \subseteq B_j$ holds for all i, j.

For each j = 1... q we computer $H_{F^+}(B_j)$ (it can be seem that $\forall D \subseteq U$, $H_{F^+} D$) is computed in polynomial time in the size of S)

and set $M = \{B_j \cup \{a\}: a \in U - B_j\} = \{M_1, ..., M_t\}$. Clearly, M is constructed in polinomial time. If $(B_j) \neq U$, and $\forall l = 1, ..., t$, $H_{F^+}(M_l) = U$ hold then $B_j \in K^{-1}_S$ holds, otherwise we obtain $B_j \notin K^{-1}_S$.

For each i = 1, ..., p we set $N = \{A_i - \{a\}: a \in A_i\} = \{N_1, ..., N_s\}$. It is easy to see that we compute N in polynomial time. If there is a N_n $(1 \le n \le s)$ such that $N_n \not\in B_j$ $\forall j = 1, ... q$ then $A_i \not\in K_R$ holds. In the converse case we have $A_i \in K_R$. Clearly, if there exists $A_i \in K_R$ then $K_S \neq K_R$. We assume that $\forall i = 1, ..., p$ we have $A_i \in K_R$. We set $Q = \{A_i - \{a\}: a \in A_i, i = 1,... , p)$.

$P = \{A \in Q: H_{F^+}(A \cup \{a\}) = U \ \forall a \in U - A\}$

$J = \{A_j \cup \{a\}: a \in U - B_j, j = 1, ..., q\}$

and

$I = \{B \in J: H_{F_R}(B) = U$ and $H_{F_R}(B - a) \neq U \ \forall a \in B\}$

Based on the definition of K_S and the definition of K^{-1}_R we can see that either there is an $A \in P$ such that $A \notin K^{-1}_R$ or there exists a $B \in I$ but $B \notin K_S$ then $K_S \notin K_R$ holds.

Clearly, P, I, are constructed in polynomial time in the sizes of S, R, K_S, K^{-1}_R. Finally, we see that if for all i = 1, ..., p and j = 1, ... q $A_i \in K_R$m $B_j \in K_S$, $P \subseteq K^{-1}_R$ and $I \subseteq K_S$ hold then based on $K^{-1}_R \cup K_S$ is pseudo-saturated and by the definition of set of minimal keys and the definition of set of antikeys we obtain $K_R = K_S$. The proof is complete.

It can be seem that in the BCNF class of relation R and relation schemes S = <U, F> we have $F_R = F^+$ iff $K_R = K_S$ hold.

Consequently, the following proposition is clear.

Proposition 2.14. Let S = <U, F> be a relation scheme in BCNF and R a relation over U in BCNF. Then if $K_S \cup K^{-1}_R$ is pseudo-saturated then the FD-relation equivalence problem is solved in polynomial time in the size of S and R.

REFENCES

[1] Armstrong W. W *Dependency structurres of Database Relationships, Information Processing 74, Holand publ, Co.* (1974) pp. 580-583

[2] Burosch G. Demetrovics J., *Katona G. O. H The poset of closures as a model of changing database. Order* 4 (1987) pp, 127 - 142.

[3] Czedli G. *On the dependencies in the relation model of data, EIK* 17(1981) 2/3 pp, 103 - 112.

[4] Demetrovics J, *Relacios adatmodell logikai es structuralis vizsgalata MTA- SZTALI Tanulmanyok, Bodapest*, 114 (1980) pp, 1-97.

[5] Demetrovics J., *Gyepesi G. On the functional dependency and some generalizations of it. Acta Cybernetica Hungary* V/3 (1983) pp, 295 - 305.

[6] Demetrovics J, *Thi V. D some results about functional dependencies. Acta Cybernetica Hungary* VIII/3 (1988) pp, 273 - 278.

[7] Dentrovics J, *Thi V. D Delations and minimal keys. Acta Cybernica Hungary* VIII/3 (1998) pp 279-285.

[8] Gottlob G., Liblin L. *Investigations on Armstrong relations, dependency inference, and exclused functional dependencies. Acta Cybernica Hungary Tom, 7 Fasc.* 4 (1990) pp 385-402.

[9] *Thi V. D minial keys and Antikeys, Acta Cebernica Hungary Tom*, Fasc 4 (1986) pp, 361 - 371.

[10] *Thi V. D Strong dependencies and s-semilattices. Acta Cybernica Hungary* VIII. 2 (1987) pp, 175 - 202.

[11] *Thi V. D Strong dependencies and irredundant relations. Computers and Artificial Interligence 7* (1988) pp 165 - 184.

[12] Demotrovics J, *Thi V. D. Some results about normal forms functional dependency in the relational datamodel. Descrete applied mathematics 69* (1996) pp, 61 - 74.

[13] *Thi V. D. On the nonkeys, J. Informatics and cybernetics. Hanoi VietNam* 13.1 (1997) pp 11-15.