Tạp chí Tin học và Điều khiển học T. 11, S. 4 (1995) (46 - 52)

SOME PROBLEMS RELATED TO PRIMITIVE MAXIMAL DEPENDENCIES

VU ĐỨC THI

Institute of Information Technology

Summary. This paper gives some results about primitive maximal dependencies. Some computational problems related to primitive maximal dependencies and antikeys are investigated.

I. INTRODUCTION

A full family of functional dependencies (FDs) was introduced by E.F. Codd. The primitive maximal dependencies (PMDs) are introduced in [4]. It is known [6] that a full family of FDs can be uniquely determined by its primitive maximal dependencies (recall that a FD $A \to \{a\}$ is PMD if $a \notin A$ and $A' \subset A$. $A' \to \{a\}$ imply $A' = A$).

It is shown [9] that from a set of PMDs of a given relation schemes we can effectively construct an Armstrong relation of s. In this paper we prove the following problem is NP-complete:

Given a relation scheme s and two attributes a, b decide whether there exits a PMD $A \to \{a\}$ such that $b \in A$.

This paper gives an algorithm finding all PMDs for a given relation scheme. We show that in many cases the time complexity of this algorithm is polynomial.

It is known [12] that the problem of finding all antikeys (maximal nonkeys) of an arbitrary relation is solved by polynomial time algorithm. We prove that the time complexity of the problem to find a set of antikeys for relation scheme is exponential in the number of attributes.

Some necessary definitions and results that are used in next section are in [21].

Definition 1.1. Let $s =, R, F >$ be a relation scheme. $AFD: A \to \{a\} \in F^+$ is called the primitive maximal dependency of s if $a \notin A$ and for all $A' \subseteq A: A' \to \{a\} \in F^+$ implies $A = A'$.

It is known [6] that an arbitrary full family of FDs can be uniquely determined by its primitive maximal dependencies.

Denote $T_a = \{A: A \to \{a\}$ is a PMD of $s\}$. It can be seen that $\{a\}, R \in T_a$.

II. RESULTS

First we present some computational problems related to PMDs.

We introduce following problem.

Theorem 2.1. *The following problem is NP-complete:*

Given a relation scheme s and two attributes a, b, decide whether there exits a PMD $A \to \{a\}$ such that $b \in A$.

Proof. For b we nondeterministically choose a subset A of R such that $b \in A$. By an algorithm finding the closure of A (see [2]) and based on Definition 1.1 we decide whether $A \in T_a$. It is obvious that this algorimth is nondeterministic polynomial. Thus, our problem lies in NP.

Now we shall show that our problem is NP-hard. It is known [11] that the prime attribute problem for relation scheme is NP-complete. Now we prove that this problem is polynomially reducible to our problem.

Let $s' =< P, F' >$ be a relation scheme over P. Without loss of generality we assume that P is not a minimal key of s', i.e. if $A \in K_{s'}$ then $A \subset P$. By a polynomial time algorothm finding minimal key of relation scheme (see [11]) we can find a minimal key C of s' from P and F'. Now we construct the relation scheme $s =< R, F >$ as follows:

$R = P \cup a$, where $a \notin P$ and $F = F' \cup C \to \{a\}$.

It is obvious that s is constructed in polynomial time in the sizes of P and F'. Clearly, $C \in K_s$ holds. Based on construction of s and definition of minimal key we can see that if $A \in K_{s'}$ then $A \in K_s$. Conversely, if B is a minimal key of s, then by $C \to \{a\} \in F$ we have $a \in B$. On the other hand, by definition of minimal key $B \in K_{s'}$. Thus, $K_{s'} = K_s$ holds. By $C \in K_s$ and $a \notin R$, if $B \to \{a\}$ is a PMD of s, then $B \in K_s$. It can be seen that if $A \in K_{s'}$, then $a \to \{a\} \in F^+$. According to Definition 1.1 $A \to \{a\}$ is a PMD of s. Consequently, an attribute b is prime of s' if and only if there exits a PMD $A \to \{a\}$ of s such that $b \in A$. The theorem is proved.

We present an algorithm finding all PMDs for a given relation scheme.

Definition 2.2. Let $s =< R, F >$ be a relation scheme, $a \in R$. Set $K_a = \{A \subseteq R : A \to \{a\}, \nexists B : (B \to \{a\}) (B \subset A)\}$. K_a is called the family of minimal sets of the attribute a.

Clearly, $R \notin K_a$, $\{a\} \in K_a$ and K_a is a Sperner system over R. It is easy to see that $K_a - \{a\} = T_a$ (see Definition 1.1).

Algorithms 2.3. (Finding a minimal sets of the attribute a)

Input: Let $s =< R, F >$ be a relation scheme, $A = \{a_1, ...a_t\} \rightarrow \{a\}$.

Output: $A' \in K_a$.

Step 0: We set $L(0) = A$.

Step i + 1: Set

$$L(i+1) = \begin{cases} L(i) - a_{i+1}, & \text{if } L(i) - a_{i+1} \rightarrow \{a\} \\ L(i), & \text{otherwise.} \end{cases}$$

Then we set $A' = L(t)$.

Lemma 2.4. $L(t) \in K_a$.

Proof. By the induction it can be seen that $L(t) \rightarrow \{a\}$, and $L(t) \subseteq ... \subseteq L(0)$ (1). If $L(t) = a$, then by the definition of the minimal set of attribute a we obtain $L(t) \in K_a$. Now we suppose that there is a B such that $B \subset L(t)$ anb $B \neq 0$. Thus, there exits a_j such that $a_j \in B$, $a_j \in L(t)$. According to the constructiuon of algorithm we have $L(j-1) - a_j \rightarrow \{a\}$. It is obvious that by (1) we obtain $L(t) - a_j \subseteq L(j-1) - a_j$ (2). It is clear that $B \subseteq L(t) - a_j$. From (1), (2) we have $B \rightarrow \{a\}$. The lemma is proved.

Clearly, by the linear-time membership algorithm in [3] the time complexity of algorithm 2.3 is $O(|R|^2\|F|)$.

Lemma 2.5. *Let $s =< R, F >$ be a relation scheme, $a \in R$, K_a be a family of minimal sets of a, $L(K_a, \{a\} \in L$. Then $L \subset K_a$ if and only if there are C, $A \rightarrow B$ such that $C \in L$ and $A \rightarrow B \in F$ and $\forall E \in L \Rightarrow E \not\subseteq A \cup (C - B)$.*

Proof. $\rightarrow$ We assume that $L \subset K_a$. Consequently, there exits $D \in K_a - L$. By $\{a\} \in L$ and K_a is a Sperner system over R, we can construct a maximal set Q such that $D \subseteq Q \subseteq R$ and $L \cup Q$ is a Sperner system. From the definition of K_a we obtain $Q \rightarrow \{a\}$ (1) and $a \in Q$ (2). If $A \rightarrow B \in F$ implies ($A \subseteq Q$, $B \subseteq Q$) or $A \subseteq Q$ then $Q^+ = Q$. By (2) $Q \rightarrow \{a\}$. This conflicts with (1). Consequently, there is a FD $A \rightarrow B$ such that $A \subseteq Q$ and $B \not\subseteq Q$. From the construction of Q there is C such that $C \in L$, $A \subseteq Q$, $C - B \subseteq Q$. It is obvious that $A \cup (C - B) \subseteq Q$. Clearly, $E \not\subseteq A \cup (C - B)$ for all $E \in L$.

$\Leftarrow$ We assume that there are C, and $A \rightarrow B$ such that $C \in L$. $A \rightarrow B \in F$ and $E \not\subseteq A((C - B)$ for all $E \in L$ (3). By the definition of L we obtain $A \cup (C - B) \rightarrow \{a\}$. By $\{a\} \in L$ there is D such that $D \in K_a$, $a \notin D$ and $D \subseteq A \cup (C - B)$. By (3) $D \in K_a - L$. Our proof is complete.

Based on this lemma and algorithm 2.3 we construct the following algorithm by induction

Algorithm 2.6. (Finding a family of minimal sets of attributes a).

Input: Let $s =< R, F >$ be a relation scheme, $a \in R$.

Output: K_a.

Step 1: Set $L(1) = E_1 = \{a\}$.

Step i+1: If there are C and $A \to B$ such that $C \in L(i)$, $A \to B \in F$, $\forall E \in L(i) \Rightarrow E \notin A \cup (C - B)$, then by algorithm 2.3 construct an E_{i+1}, where $E_{i+1} \subseteq A \cup (C - B)$, $E_{i+1} \in K_a$. We set $L(i+1) = L(i) \cup E_{i+1}$. In the converse case we set $K_a = L(i)$.

By Lemma 2.5 there exits a natural number n such that $K_a = L(n)$.

It can be seen that the worst-case time complexity of algorithm is $O(|R||F||K_a|(|R| + |K_a|))$. Thus, the time complexity of this algorithm is polynomial in $|R|$, $|F|$, and $|K_a|$. Clearly, if the number of element of K_a for a relation scheme $s =< R, F >$ is polynomial in the size of s, then this algorithm is effective. Especially, when $|K_a|$ is small. It is obvious that if for each $A \to B \in F$ implies $a \in A$ or $a \in B$, then $K_a = \{a\}$.

Based on algorithm 2.6 we construct an algorithm finding a set of all PMDs from a given relation scheme, as follows:

Algorithm 2.7. (Finding all PMDs)

Input: Let $s =< R, F >$ be a relation scheme.

Output: $P = \{A \to \{a\} : A \to \{a\}$ is a PMD of s, $a \in R\}$.

Step 1: For each $a \in R$ by algorithm 2.6 compute K_a.

Step 2: Set $P = \{A \to \{a\} : A \in K_a - \{a\}, a \in R\}$.

It can be seen that $s' =< R, P >$ is a cover of s, i.e. $F^+ = P^+$. Clearly, P is a set of all PMDs of s and the worst-case time complexity of algorithm 2.7 is $O(|R||F||K_a|(|R| + |K_a|)))$.

It is obvious that the time complexity of algorithm 2.7 is polynomial in $|r|$, $|F|$, and $|K_a|$.

Remark 2.8. Let $s =< R, F >$ be a relation scheme. Set $s' =< R \cup \{a\}, F' >$, where $a \notin R$ and $F' = F \cup R \to \{a\}$.

It can be seen that $a \in K_s$ holds iff $A \to \{a\}$ is a PMD of s'. Consequently, for finding a set of all minimal keys of a given relation scheme s, we can compute K_a of s'. It is obvious that $K_s = K_a - \{a\}$. Thus, using algorithm 2.6 we obtain K_s.

Now we prove that the time complexity of finding a set of antikeys for relation scheme is exponential in the number of attributes.

Let $s = < R, F >$ be a relation scheme over R. From s we construct $Z(s)$ and compute the minimal generator N_s of $Z(s)$. We put

$$T_s = \{A \in N_s : \; B \not\exists N_s : A \subset B\}.$$

It is known [1] that for a given relation scheme s there is a relation r such that r is an Armstrong relation of s. On the other hand, by Corollary 1.2 and Theorem 1.3 in [21] the following proposition is clear.

Proposition 2.9. *Let $s = < R, F >$ be a relation scheme over R. Then*

$$K_s^{-1} = T_s.$$

It is shown [8] that the problem of finding all antikeys of a relation is soved by polynomial time algorithm. For a relation scheme we have the following theorem.

Theorem 2.10. The time complexity of finding a set of all antikeys of a given relation scheme is exponential in the number of attributes.

Proof. We have to prove that:

(1) There is an algorithm which finds a sets of all antikeys of a given relation scheme in exponential time in the number of attributes.

(2) There exits a relation scheme $s = < R, F >$ such that the number of elements of K_{s-1} is exponential in the number of attributes (in our example $|Ks - 1|$ is exponential not only in the number of attributes, but also in the number of elements of F).

For (1): We construct a following algorithm.

Let $s = < R, F >$ be a relation scheme over R.

Step 1: For every $A \subseteq R$ compute A^+, and set $Z(s) = \{A^+ : \; A \subseteq R\}$.

Step 2: Construct the minimal generator N_s of $Z(s)$.

Step 3: Compute the set T_s from N_s.

According to Proposition 2.9 we have $T_s = K_s$.

Clearly, the time complexity of this algorithm is exponential in $|R|$.

As to (2): Let $R = \{a_1, ..., a_{3m}\}$.

We take a partition $R = X_1 \cup ... \cup X_m$, where $|X_i| = 3$ $(1 \le i \le m)$.

Set

$$K = \{B : \; |B| = 2, \; B \subseteq X_i \text{ for some } i\}.$$

It is easy to see that

$$K^{-1} = \{A : \ |A| \cap X_i| = 1, \ \forall i\}.$$

It is clear that $|K| = 3m, \ |K - 1| = 3m$.

Thus, if denote the element of K by $K_1, ..., K_t$, then set $s =< R, F >$, where $F = \{K_1 \to R, ..., K_t \to R)$. By Theorem 1.5 in [21] K^{-1} is the set of all antikeys of s. Consequently, we can construct a relation scheme $s =< R, F >$ such that $|F| = |R| = n$, but the number of antikeys of s is $3n/3$. The Theorem is proved.

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 relation schemes*, ACM Trans. on Database Syst. 4,1 (1979) 30-59.
3. Beeri C., Dowd M., Fagin R., Staman R., *On the Stucture of Armstrong relations for Functional Dependencies*, J.ACM 31, 1 (1984) 30-46.
4. Bekessy A., Demetrovics J., *Contribution to the theory of data base relations*, Discrete Math. 27 (1979) pp,. 1-10.
5. Demetrovics J., *Logical and strutural investigation on relational datamodel (in Hungarian)*, MTA-SZTAKI Tanulmanyok, Budapest, 114 (1980) 1-97.
6. Demetrovics J., Libkin L., Muchnik I.B., *Functional dependencies and the semilattice of closed classes*, Proceedings of MFDBS 87, Lecture Notes in Computer Science (1987) 136-147.
7. Demetrovics J., Thi V.D., *Some results about functional dependencies*, Acta Cybernetica 8, 3 (1988) 273-278.
8. Demetrovics J., Thi V.D., *Relation and minimal keys*, Acta Cybemetica 8,3 (1988) 279-285.
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 , British Vol 26, No 4 (1993), 43-55.
10. Demetrovics J., Thi V.D., Armstrong Relation, Functional dependencies and Strong Dependencies. Comput and AI (to appear).
11. Lucchesi C.L. Osborn S.L., *Candidate keys for relations*, J. Comput. Syst. Scien. 17,2 (1987) 270-279.

12. Thi V.D., Demetrovics J., *On the time complexity of Algonithms Related to Boyce- Codd Normal Form. J.PLISKA,* Studia Mathematika, the Bulgarian Academy of Sciencies, Bulgaria, No. 12 (1993).

13. Thi V.D., Demetrovics J., *Some Computational Problems Related to the funtional Dependencies in the Relational Datamodel,* Acta Scientiarum Mathmeticarum 57 (1993) 627-638.

14. Thi V.D., Demetrovics J., *Some Problems concerning keys for Relation Schemes and Relations in the Relational Datamodel,* Information Processing Letters, Holland, 46, (1993) 179-184.

15. Thi V.D., Demetrovics J., *Some Problems concerning armstrong relations of dual schemes and relation schemes in the relational datamodel.,* Acta Cybemetica, Vol 11, 1-2 (1993) 35-47.

16. Thi V.D., Demetrovics J., *Normal Forms and Minimal Keys in the Relational Datamodel,* Acta Cybemetica, Vol 11, 3 (1994) 205-215.

17. Thi V.D. *Minimal Keys and Antikeys,* Acta Cybemetica 7, 4 (1986) 361-371.

18. Thi V.D. *Investigation on Combinatorial Characterzations Related to Functional Dependency in the Relational Datamodel,* MTA SZTAKI Tanulmanyok, Budapest, 191 (1986) 1-157.

19. Thi V.D. *On Antikeys in the Relational Datamodel (in Hungarian),* Alkamazott Matamatikai Lapok 12 (1986) 111-124.

20. Thi V.D., *Logical Dependencies and irrdundant relations,* Computers and Artifical Intelligent, 7 (1988) 165-184.

21. Thi V.D., *On the equivalent descriptions of family of functional dependencies in the relational datamodel,* Tạp chí Tin học và Điều khiển học V. 11, S. 4 (1995) 40-50.