Tạp chí Tin học và Điều khiển học, T.14, Số.2 (1998), (1-7)

ON THE COMPUTATIONAL ALGORITHM RELATED TO ANTIKEYS

VU DUC THI[1]

Astract. The keys and antikeys play important roles for the investigation of functional dependency in the relational datamodel. The main purpose of this paper is to prove that the time complexity of finding a set of antileys for a given relation scheme S is exponential in the number of attributes. Some another results connecting the functional dependency are given.

Key Word and phrase: Relation, relational datamodel, functionsl dependency, relation scheme, generating Armstrong relation, dependency inference, strong schemen, membership problem, closure, closed set, minimal generater, 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 $R = \{h_1.... h_n\}$ 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 $F_R = \{(A, B)$: 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 $A \rightarrow B$ 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 $A \rightarrow B$ holds in a relation R if $A \xrightarrow[R]{f} > B$. We also say that R statisfies the FD $A \rightarrow B$.

[1]Let U be a finite set, and denote P(U) its power set.

Let Y $\subseteq$ P(U) x 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) $(A, B) \subseteq Y, (C, D) \in Y \Rightarrow (AUC, BUD) \in Y$.

Clearly, F_R is an f-family over U.

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

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.

Clearly, in [1] if S = <U,F> is a relation scherme, 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, YY 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 $A \xrightarrow[R]{} U$ ($A \rightarrow U \in F^+$, (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.

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:

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

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

It is know [4] that if K ia an arnitrayru 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 $N = \{A \in I: A \# \cap \{A' \in I: A \subset A'\}\}$.

In [6] 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 that Rrepresents K iff $K_R = K$.

The following theorem is known in [8]

Definition 1.10. Let K be a relation, and K a Sperner system over U. We say that R presents K iff $K^{-1} = T_R$, where T_R is the maximal equality system of R.

2. RESULTS

In this section we investigate the connections betweem Armstrong relations, relation scheme, and Sperner systems from different aspects.

Definition 2.1. Let $S = \langle U, F = \{A_i \rightarrow B_i : i=1, \ldots, m\}\rangle$ be a relation scheme over U, and $X \subseteq U$. Denote $X^+ = \{a \in U : X \rightarrow \{a\} \in F^+\}$. Then X^+ is called the closure of X over S.

It is known that in [2] ther is an algorithms that computes X^+ from X and this algorithms has polynomial time complexity in $|U|$ and $|F|$.

Remark 2.2: If F is an f = family over U, we denote $H_F(A) = \{a \in U: A \rightarrow \{a\} \in F\}$. Where (A, B) or $A \rightarrow B$ denotes a fuctional dependency. Denote $Z(F) = \{A \in U: H_F(A) = A\}$. It is easy to see that $U, 0 \in Z(F)$ and $A, B \in Z(F)$ implies $A \cap B \in Z(F)$. Clearly, for a relation scheme $S = \langle U, F\rangle$ F^+ is an f-family over U.

Theorem 2.3. [4] Let F_1, F_2 be two f-families over U. Then $F_1 = F_2$ iff $Z(F_2) = Z(F_1)$, and $F_1 \subseteq F_2$ iff $Z(F_2) \subseteq Z(F_1)$.

Theorem 2.4. [4] Let K be a Sperner system, and $S = \langle U, F\rangle$ a relation scheme over U. Then $K_s = K$ iff $[U] \cup K^{-1} \subseteq Z(F^+) \subseteq \{U\} \cup G(K^{-1})$, where $G(K^{-1}) = \{A \in P(U): \exists B \in K^{-1}: A \subseteq B\}$.

According to [4], clearly.

Theorem 2.5. Let $K = \{K_1, \ldots, K_m\}$ be a Sperner system over U.

Denote $S = \langle U, F\rangle$ whith $F = \{K_1 \rightarrow U, \ldots, K_m \rightarrow U\}$.

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

We show the following problems that play important roles for the logical and structural is vestigation of the relational detamodel in pratice and design theory.

(1) Constructing Armstrong relation: Let $S = \langle U, F\rangle$ be a relation scheme. Construct a relation R for which $F^+ = F_R$.

(2) Constructing relation scheme: Let R be a relation. Construct a relation scheme $S = \langle U, F\rangle$ such that R is the Armstrong relation of S.

(3) FD-relation implication problem: Let R be a relation, and $S = \langle U, F\rangle$ a relation scheme. Decide whether $F_R \subseteq F^+$.

(4) FD-relation equivalence problem: Let $S = \langle U, F\rangle$ a relation scheme, R be a relation. Decide whether $F^+ = F_R$.

According to [2, 9, 13] two fist problems are inherently difficult. In [13] problem 3 is co-NP-complete. For the FD-equivalence problem we can costruct an algorithm to slove this problem in exponential time.

Fist we give an algorithm to slove the second problem. In [8] we proved the follwing theorem.

Theorem 2.6. Let $R = \{h_1, \ldots, h_m\}$ be a relation, and F an f-family over U.

$$H_F(A) = \begin{cases} \quad {}_{ij} \\ U \text{ otherwise} \end{cases}$$

Where $H_F(A) = \{a \in U: (A, \{a\}) \in F\}$ and E_R is the equality set of R.

In relation scheme S = <U, F>, a functinal dependency $A \to B \in F$ is calledredudant if ether A = B or there is $c \to B \in F$ such that $C \subseteq A$.

Theorem 2.7

Input: a relation $R = \{h_1, ..., h_m\}$ over U.

Output: a relation scheme S = <U, F> such that $F_R = F^+$.

Step 1. Find the equality set $F_R = \{E_{ij}: 1 \leq i \leq j \leq m\}$.

Step 2: Find the minimal generator N, where $N = \{A \in E_R: A \# \cap \{B \in E_R: A \subset B\}\}$.

Denote element of N by $A_1, ..., A_t$.

Step 3: For every $B \subseteq U$ if there is A_i such that $B \subseteq A_i$, we compute $C = \bigcap_{B \subseteq A_i} A_i$ and set $B \to C$. In the converse case we set $B \to U$.

Denote: 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 redudant functional dependency$\}$.

Cleary, according to Theorem 2.6, algorithm 2.7 find q relation scheme S such that a given relation R is an Armstrong relation of S.

The fllowing example shows that for a given relation R Algorithm 2.7 can be applied to construct a relation scheme S such that S is an Armstrong relation of S.

Example 2.8 R is a relation over R = {a, b, c, d}

	b	c	d
0	0	0	0
0	0	0	1
2	0	0	0
3	3	0	0
4	0	4	4
5	5	5	0

Clealry, $E_R = \{\{b, c, d\}, \{a, b, c\}, \{b, c\}, \{c, d\}, \{b\}, \{c\}, 0\}$

The minimal generetor $N = \{\{a, b, c\}, \{b, c, d\}, \{c, d\}, \{b\}, \{d\}\}$

It is easy to see that S = <U, F>, where U = {a, b, c, d}, $F = \{\{a, d\} \to U, \{a\} \to \{a, b, c\}, \{b, c\} \to \{b, c, d\}$.

It can be seen that the time complexity of algorithm 2.7 is exponential in the number of attributes. In [14] it is known that there is relation R containing O(n) rows such that a minimal relation scheme S of F containing $O(2^{n/2})$ FDs, where $n = |U|$. From this obsetvation and by Algorithm 2.7 the following proposition (in [13, 14]) is clear.

Proposition 2.9. The time comlexity of the second problem is exponential in the size os a given relation.

In [16] we give an algorithm which find a relation scheme $S = <U, F>$ from a given relation R such that $F_R = F^+$ and we show that in many cases trhe tme complexity of this algorithm is polynomial in the size of R.

Let $S = <U, F>$ be a relation scheme and $R = \{h_1, ..., h_m\}$ be a relation over U, we compute the minimal generator N_R of $Z(F_R)$ from E_R (in polynomial time in the minimal generator N_s of Z(S) (in exponential time in the number of element of U). By Theorem 2.3 we compare N_R with N_s, the two last problems are solved.

Let $S = <U, F>$ be a relation scheme over U, K_s is a set of all minimal keys of S. We call K_s^{-1} is a set of all antikeys of S. From S we construct $Z(S) = \{X^+: X \subseteq U\}$, and compute the minimal generator N_s of Z(S). We set

$$T_S = \{A \in N_s: \exists B \in N_S: A \subset B\}$$

It is known [1] that for given relation scheme S there is relation R such that R is an Armstrong relation of S. On the other hand, by Theorem 1.10, and Theorem 2.6 the following proposition is clear.

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

$$K_S^{-1} = T_S.$$

Theprem 2.11 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 al which finds a set of all antikeys of a given relation scheme in exponential time in the number of attributes.

(2) There exists a relation scheme $S = <U, F>$ such that the number of elements of K_S^{-1} is exponential in the number of attributes (in our example $|K_S^{-1}|$ is exponential not only in the number of attributes, but also in the number of element of F).

For (1): we construct a following algorithm algorithm.

Let $S = <U, F>$ be a relation scheme over U.

Step 1: For every $A \subseteq U$ we compute A^+, and set $Z(S) = \{A^+: A \subseteq U\}$

Step 2: We construct the minimal generator N_S of Z(S).

Step 3: We compute the set T_S from N_S.

According to Proposition 2.10 we have $T_S = K_S$.

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

For (2). Let us take a partition $U = X_1 \cup, ..., \cup X_m \cup W$, where $m = [n/3]$, and $|X_i| = 3\ (1 \le i \le m)$.

We set

$K = \{B: |B| = 2, B \subseteq X_i \text{ for some } i\}$ if $|W| = 0$,

$K = \{B: |B| = 2, B \subseteq X_i \text{ for some } i: 1 \le i \le m - 1 \text{ or } B \subseteq X_m \cup W\}$

$K = \{B: |B| = 2, B \subseteq X_i \text{ for some } i: 1 \le i \le m \text{ or } B = W\}$ if $|W| = 2$

It is easy to see that

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

$K^{-1} = \{A: |A \cap X_i| = 1, (1 \le i \le m - 1) \text{ and } |A \cap (X_m \cup W|\ \}$ if $|W| = 1$,

$K^{-1} = \{A: |A \cap X_i| = 1, (1 \le i \le m) \text{ and } |A \cap W| = 1\}$ if $|W| = 2$.

It is clear that $n - 1 \le |K| \le n + 2{,}3^{[n/4]} < |K^{-1}|$.

Thus, if denote the elements of K by $K_1, ..., K_t$, then we set $S = <U, F>$, where $F = \{K_1 \to U, ..., K_t \to U\}$. By Theorem 2.5 K^{-1} is the set of all antykeys of S. It is clear that for the arbitrary set of attributes we always can construct a relation scheme $S = <U, F>$ such that $|F| < |U| < 2$, but the number of element of F. The theorem is proved.

REFERENCES

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

[2] **Beeri C. Dowd M. Fagin R. Staman R.** *On the structure of Armstrong relations for Functional Dependencies.* J. ACM 31.1 (1984) pp. 30 - 46.

[3] **Beeri C. Bernstein P. A.** *Computainonal problems related to the design of norma form relational schemas.* ACM. Trans, on Database Syst, 4.1 (1979) pp. 30 - 59

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

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

[6] **Demetrovics J.** *Relacios adatmodell logikai es strukturalis vizsgalata MTA - SZTALI* Tanulmanyok, Budapest, 114 (1980) pp 1 - 97.

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

[8] **Demetrovics J. Thi V.** *D some results about functional dependencies.* Acta Cybernetical Hungary VIII/3 (1983) pp. 273 - 278.

[9] **Demetrovics J. Thi V. D** *Relations and minimal keys.* Acta Cybernica Hungary VIII/3 (1988) pp. 279 - 285.

[10] **Thi. V. D**. *Minimal keys and antikeys*. Acta Cybernetica Hungary Tom 7, Fasc. 4 (1986) pp. 361 - 371.

[11] **Thi V. D** *Strong dependencies and s-semilattices*. Acta Cybernica Hungary VIII/2 (1987) pp. 195 - 202.

[12] **Thi V. D** *Logical dependencies and irredundant relations*. Computes and Artificial Interlligence, vol.7. N. 2 (1988) pp 165 - 184.

[13] **Gottlob G. Libkin L**. *Invesgations on Armstrong relations, dependency inference, and exclused functional dependencies*. Acta Cybernica Hungary Tom, 9. Fase, 4 (1990) pp 385 - 402.

[14] **Mannila H. Raiha K. J.** *Algorithms for inferring functional dependencies. Proc. Of the thirteenth international conference of very large data bases*. Brighton, September 1987.

[15]. **Lucchesi C. L. Osborn S. L.** *Candidate keys for relations*. J. Comput, Syst. Scien. 17.2 (1987) pp 270 - 279.

[16] **Demetrovics J. Thi. V. D.** *On the algorithms for generating Armstrong relations and for inferring fuctional dependencies in the relational datamodel*. Computers Math, Applic, Great Britain, 26. 4 (1993) pp 43 - 55.

[17] **Demetrovics J. Thi. V. D.** *Some result about normal forms for functional dependency in the relation datamodel*. Discrete Aplied Mathematics 69 (1996) pp. 61 - 74.

[18]. **Thi V. D.** *Some result about hypergraph. J. Informatics and Çybernetics*. Hanoi Vietnam 13,2 (1997) pp. 8 - 15.