ON THE NONKEYS

VU DUC THI

Abstract. In this paper we give some results about nonkeys. We show that for relation scheme the problem decide whether there is a nonkey having cardinality greater than or equal to a given integer m is NP-complete. However, for relation this problem can be solved by a polynomial time algorithm

Key words and phrases:
Relational datamodel, relation, relation scheme, functional dependency, closure, key, NP-complete.

1. INTRODUCTION

It is known [3, 5-8, 13, 15-17] that nonkeys and maximal nonkeys play an important role for extremal problem in the relational datamodel as well as for many other problems. On the other hand, the NP-complete problems are essential problems in the algorithm theory. This paper gives some computational results related to nonkeys.

Let us give some necessary definitions that are used in the next section. The concepts give in this section can be found in [1-4, 7, 9, 10, 15-17].

Definition 1. Let $R = \{a_1, ..., a_n\}$ be a nonempty finite set of attributes, $r = \{h_1, ..., h_m\}$ be a relation over R, and $A, B \subseteq R$.
Then we say that B functionally depends on A in r (denoted $A \xrightarrow{f}{}_r B$) iff

$$(\forall h_i, h_j \in r)\left((\forall a \in A)(h_i(a) = h_j(a)) \Rightarrow (\forall b \in B)(h_i(b) = h_j(b))\right).$$

Let $F_r = \{(A, B) : A, B \subseteq R, A \xrightarrow{f}{}_r B)$. F_r is called the full family of functional dependencies of r. Where we write (A, B) or $A \rightarrow B$ for $A \xrightarrow{f}{}_r B$ when r, f are clear from the context.

Definition 2. A functional dependency over R is a statement of the form $A \rightarrow B$, where $A, B \subseteq R$. The FD $A \rightarrow B$ holds in a relation r if $A \xrightarrow{f}{}_r B$. We also say that r satisfiles the FD $A \rightarrow B$.

Clearly, F_r is a set of all FDs that hold in r.

Definition 3. Let R be a nonempty finite set, and denote $P(R)$ its power set. Let $Y \subseteq P(R) \times P(R)$. We say that Y is an f-family over R iff for all A, B, C, $D \subseteq R$

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

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

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

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

Clearly, F_r is an f-family over R.

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

Definition 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. Let F^+ be a set of all FDs that can be derived from F by the rules in Definition 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, if $s = \langle R, F \rangle$ be 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.

It is obvious that all FDs of s hold in r.

Definition 5. Let r be a relation, $s = \langle R, F \rangle$ be a relation scheme, Y be an f-family over R and $A \subseteq R$. Then A is a key of r (a key of s, a key of Y) if $A \underset{r}{\overset{f}{>}} R$ ($A \rightarrow R \in F^+$, $(A, R) \in Y$). A is a minimal key of $r(s, Y)$ if A is a key of $r(s, Y)$, and any proper of A is not a key of $r(s, Y)$. Denote K_r, (K_s, K_y) the set of all minimal keys of $r(s, Y)$.

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

Definition 6. Let K be a Sperner system over R. We define the set of antikeys of K, denoted by K^{-1}, as follows:

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

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

It is known [4] 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).

Definition 7. Let $I \subseteq P(R)$, $R \in I$, and $A, B \in I \Rightarrow A \cap B \in I$. Let $M \subseteq P(R)$. Denote $M^+ = \{\cap M' : M' \subseteq M\}$. We say that M is a generator of I iff $M^+ = I$. Note that $R \in M^+$ 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 8. Let r be a relation over R, and E_r the equality set of r, i.e. $E_r = \{E_{ij} : 1 \leq i \leq j \leq |r|\}$, where $E_{ij} = \{a \in R : h_i(a) = h_j(a)\}$.
Let $T_r = \{A \in P(R) : \exists E_{ij} = A, \exists E_{pq} : A \subset E_{pq}\}$. Then T_r is called the maximal equality system of r.

Definition 9. Let r be a relation, and K a Sperner system over R. We say that r represents K iff $K_r = K$.

The following theorem is known ([6]).

Theorem 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 K_s^{-1} the set of all an antikey antikey of s. From Theorem 1 we obtain the following proposition.

Proposition 1. *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.*

2. RESULTS

Let $s = \langle R, F \rangle$ be a relation scheme over R. Denote $V_s = \{A : A$ is not a key of $s\}$. V_s is called the set of all nonkeys of s. From Proposition 1 we can see that K_s^{-1} is the set of all maximal nonkeys of s. It is known [3, 5, 6, 8, 13, 15, 16] that maximal nonkeys play an important role for extremal problems in the relational datamodel.

Now we give the NP-complete problem concerning nonkeys.

Theorem 2. *The following problem is NP-complete:*

Let $s = \langle R, F \rangle$ be a relation scheme over R and an integer m $(m \leq |R|)$, decide whether there is a nonkey A such that $m \leq |A|$.

Proof: We nondeterministically choose a set A so that $m \leq |A|$ and decide whether A is not a key of s. Clearly, by the polynomial time algorithm finding the closure (see [2]) our algorithm is nondeterministic polynomial. Thus, the problem lies in NP. Thus, our problem lies in NP.

It is known [12] that the independent set problem is NP-complete:

Given integer m and a non-directed graph $G = \langle V, E \rangle$, where V is subset $A \subseteq V$ such that for all a, $b \in A$ the edge (a, b) is not in E. The independent set problem is deciding whether G contains an independent set A having cardinality greater than or equal to m.

We shall prove that the independent set problem is polynomially reducible to our problem.

Let $G = \langle V, E \rangle$ be a non-directed graph, $m \leq |V|$. Set $s = \langle R, F \rangle$, where $R = V$ and $F = \{\{a_i, a_j\} \rightarrow \{a\} : (a_i, a_j) \in E, \text{ for all } a \in V - \{a_i, a_j\}\}$. Clearly, s is constructed in polynomial time in the size of G.

According to the definition of the set of edges. E is a Sperner system over R. From this, we can see that s is in BCNF. Because E is the set of edges, and by the definition of the minimal key, we can see that if $(a_i, a_j) \in E$, then $\{a_i, a_j\}$ is a minimal key of s. Conversely. If $B \in K_s$. Then have $\{a_i, a_j\} = B$. Hence, $K_s = E$ holds.

Consequently, A is not a key of s if and only if $\{a_i, a_j\} \notin A$ for all $(a_i, a_j) \in E$. Thus, A is a nonkey of s if and only if A is an independent set of G. The theorem is proved.

It is known that K_s^{-1} is the set of all maximal nonkeys of the relation scheme s. By Theorem 2 we can obtain

Proposition 2. *If $NP \neq P$, then there is no a polynomial time algorithm finding K from a given relation scheme s.*

However, from a given relation r we can compute the maximal equality system T_r. On the other hand, by Theorem 1 we have $K_r^{-1} = T_r$ where K_r^{-1} is the set of all antikeys of r. Consequently, T_r is the set of all maximal nonkeys of r. We obtain the following proposition

Proposition 3. *For relation, the problem deciding whether there is a nonkey having cardinality greater than or equal to a given integer m can be solved a polynomial time algorithm.*

REFERENCES

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

2. Beeri C., Bernstein P. A., *Computational problem 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. Demetrovics J., *Logical and structural Investigation of Relational Datamodel*, MTA - SZTAKI Tanulmanyok, Budapest, **114** (1980), 1 - 97.

5. Demetrovics J., Thi V. D., *Some results about functional dependencies*, Acta Cybernetica, **8** (3) (1988), 273 - 278.

6. Demetrovics J., Thi V. D., *Relations and minimal keys*, Acta Cybernetica, **8** (3) (1988), 279 - 285.

7. Demetriovics J., Thi V. D., *On keys in the Relational Datamodel*, Inform. Process Cybern. EIK, **24** (10) (1988), 515 - 519.

8   Demetrovics J., Thi V. D., *Algorithm for generating Armstrong relations and inferring functional dependencies in the relational datamodel*, Computers and Mathematics with Applications, Great Britain, **26** (4) (1993), 43 - 45.

9. Demetrovics J., Thi V. D., *Some problem concerning Keys for relation Schemes and Relationals in the Relational Datamodel*, Information Processing Letters, North Holland, **46** (4) (1993), 179 - 183.

10. Demetrovics J., Thi V. D., *Some Computational Problems Related to the functional Dependency in the Relational Datamodel*, Acta Scientiarum Mathematicarum, **57** (1-4) (1993), 627 - 628.

11. Demetrovics J., Thi V. D., *Armstrong Relation, Functional Dependencies and strong Dependencies*, Comput. and AI. (submitted for publication).

12. Garey M. R., Johnson D. S., *Computers and Intractability: A Guide to theory of NP - Completeness*, Bell Laboratories, W. H Freeman and Company, San Francisco, 1979.

13. Gottlob G., Libkin L., *Investigation on Armstrong relations dependency inference, and excluded functional dependencies*, Acta Cybernetica Hungary, **IX**/4 (1990), 385 - 402.

14. Lucchesi C. L., Osbonrn S. L., *Candidate keys for relations*, J. Comput. Syst. Sci., **17** (2) (1978), 270 - 279.

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

16. Thi V. D., *Minimal keys and Antikeys*, Acta Cybernetica, 7(4) (1986), 361 - 371.

17. Demetrovics J., Thi V. D., *Some results about normal forms for functional dependency in the relational datamodel*, Discrete Applied Mathematics, **69** (1996), 61 - 74.

Institute of Information Technology

National Centre for Science and Technology of Vietnam.