Tạp chí Tin học và Điều khiển học, T.21, S.4 (2005), 337–342

ANTIKEYS AND MINIMAL KEYS OF RELATION SCHEMES

VU DUC THI[1], NGUYEN HOANG SON[2]

[1]*Institute of Information Technology, VAST*

[2]*Department of Mathematics, College of Sciences, Hue University*

Abstract. Minimal keys and antikeys play a very important role in the theory of the design of relational databases. The minimal key and antikey results have been widely investigated. Hypergraphs theory [2] is an important subfield of discrete mathematics with many relevant applications in both theoretical and applied computer science. A set of minimal keys and a set of antikeys form simple hypergrahps. In this paper, we are to investigate the minimal keys of relation schemes. We characterize the set of all minimal keys of relation schemes in terms of hypergraphs. The set of antikeys is also studied in this paper.

Tóm tắt. Khóa tối tiểu và phản khóa đóng một vai trò rất quan trọng trong lý thuyết thiết kế cơ sở dữ liệu quan hệ. Các kết quả về khóa tối tiểu và phản khóa đã được nghiên cứu nhiều. Lý thuyết siêu đồ thị [2] là một trong lĩnh vực quan trọng của toán rời rạc với nhiều ứng dụng quan trọng đối với tin học. Tập các khóa tối tiểu và tập các phản khóa có dạng siêu đồ thị đơn. Trong bài báo này, chúng tôi nghiên cứu về khóa tối tiểu của sơ đồ quan hệ. Chúng tôi đặc trưng tập tất cả khóa tối tiểu của sơ đồ quan hệ theo quan điểm siêu đồ thị. Ngoài ra, tập phản khóa cũng được nghiên cứu trong bài báo này.

1. INTRODUCTION

In this section we briefly present the main concepts of the theory of relational databases which will be needed in sequel. The concepts and facts given in this section can be found in [1, 3÷5].

Let U be a nonempty finite set of *attributes* (e.g. name, age etc) and $R = \{h_1, \ldots, h_m\}$ be a relation over U. A *functional dependency* (FD for short) over U is a statement of form $X \to Y$, where $X, Y \subseteq U$. The FD $X \to Y$ *holds* in a relation R if

$$(\forall h_i, h_j \in R)((\forall a \in X)(h_i(a) = h_j(a)) \Rightarrow (\forall b \in Y)(h_i(b) = h_j(b))).$$

We also say that R *satisfies* the FD $X \to Y$.

Let F_R be a family of all FDs that holds in R. Then $F = F_R$ satisfies

(F1) $X \to X \in F$,

(F2) $(X \to Y \in F, Y \to Z \in F) \Rightarrow (X \to Z \in F)$,

(F3) $(X \to Y \in F, X \subseteq V, W \subseteq Y) \Rightarrow (V \to W \in F)$,

(F4) $(X \to Y \in F, V \to W \in F) \Rightarrow (X \cup V \to Y \cup W \in F)$.

A family of FDs satisfying (F1) - (F4) is called an $f - family$ over U.

Clearly, F_R is an f-family over U. It is known [1] that if F is an arbitrary$f$-family, then there is a relation R over U such that $F_R = F$.

Give a family F of FDs over U, 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 (F1) - (F4).

A *relation scheme* s is a pair (U, F), where U is a nonempty finite set of attributes and F is a set of FDs over U. X^+ is called the *closure* of X on s. It is obvious that $X \to Y \in F^+$ if and only if $Y \subseteq X^+$.

Let $s = (U, F)$ be a relation scheme and $K \subseteq U$. Then K is a *key* of s if $K \to U \in F^+$. K is a *minimal key* of s if K is a key of s and any proper subset of K is not a key of s.

Denote $\mathcal{K}_s$ the set of all minimal keys of s. Evidently, $\mathcal{K}_s$ is a Sperner system over U (i.e. for every $A, B \in \mathcal{K}_s$ implies $A \not\subseteq B$).

Let $\mathcal{K}$ be a Sperner system over U. We define the set of *antikeys* of $\mathcal{K}$, denoted by $\mathcal{K}^{-1}$, as follows:

$$\mathcal{K}^{-1} = \{A \in \mathcal{P}(U) \mid (B \in \mathcal{K}) \Rightarrow (B \not\subseteq A) \text{ and } (A \subset C) \Rightarrow (\exists B \in \mathcal{K})(B \subseteq C)\}.$$

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

2. HYPERGRAPHS AND TRANSVERSALS

Let U be a nonempty finite set and put $\mathcal{P}(U)$ for the family of all subsets of U. The family $\mathcal{H} = \{E_i \mid E_i \in \mathcal{P}(U), i = 1, 2, \ldots, m\}$ is called a *hypergraph* over U if $E_i \neq \emptyset$ holds for all i (in [2] it is required that the union of E_is is U, in this paper we do not require this).

The elements of U are called vertices, and the sets $E_1, \ldots, E_m$ the edges of the hypergraph $\mathcal{H}$.

A hypergraph $\mathcal{H}$ is called *simple* if it satisfies

$$\forall E_i, E_j \in \mathcal{H} : E_i \subseteq E_j \Rightarrow E_i = E_j.$$

It can be seen that simple hypergraphs are Sperner systems. Clearly, $\mathcal{K}_s$ and $\mathcal{K}_s^{-1}$ are simple hypergraphs.

Let $\mathcal{H}$ be a hypergraph over U. Then $\min(\mathcal{H})$ denotes the set of minimal edges of $\mathcal{H}$ with respect to set inclusion, i.e.,

$$\min(\mathcal{H}) = \{E_i \in \mathcal{H} \mid \nexists E_j \in \mathcal{H} : E_j \subset E_i\},$$

and $\max(\mathcal{H})$ denotes the set of maximal edges of $\mathcal{H}$ with respect to set inclusion, i.e.,

$$\max(\mathcal{H}) = \{E_i \in \mathcal{H} \mid \nexists E_j \in \mathcal{H} : E_j \supset E_i\}.$$

It is clear that, $\min(\mathcal{H})$ and $\max(\mathcal{H})$ are simple hypergraphs. Furthermore, $\min(\mathcal{H})$ and $\max(\mathcal{H})$ are uniquely determined by $\mathcal{H}$.

A set $T \subseteq U$ is called a *transversal* of $\mathcal{H}$ (sometimes it is called *hitting set*) if it meets all edges of $\mathcal{H}$, i.e.,

$$\forall E \in \mathcal{H} : T \cap E \neq \emptyset.$$

Denote by $Trs(\mathcal{H})$ the family of all transversals of $\mathcal{H}$. A transversal T of $\mathcal{H}$ is called *minimal* if no proper subset T' of T is a transversal.

The family of all minimal transversals of $\mathcal{H}$ is called the *transversal hypergraph* of $\mathcal{H}$, and denoted by $Tr(\mathcal{H})$. Clearly, $Tr(\mathcal{H})$ is a simple hypergraph.

Proposition 2.1. ([2]) *Let $\mathcal{H}$ and $\mathcal{G}$ two simple hypergraphs over U. Then $\mathcal{H} = Tr(\mathcal{G})$ if and only if $\mathcal{G} = Tr(\mathcal{H})$.*

Proposition 2.2. ([5]) *Let $\mathcal{H}$ be a hypergraph over U. Then*

$$Tr(\mathcal{H}) = Tr(\min(\mathcal{H})).$$

The following algorithm finds the family of all minimal transversals of a given hypergraph (by induction).

Algorithm 2.3. ([3])

Input: let $\mathcal{H} = \{E_1, \ldots, E_m\}$ be a hypergraph over U.

Output: $Tr(\mathcal{H})$.

Method:

Step 0. We set $\mathcal{L}_1 := \{\{a\} \mid a \in E_1\}$. It is obvious that $\mathcal{L}_1 = Tr(\{E_1\})$.

Step q+1. $(q < m)$ Assume that

$$\mathcal{L}_q = \mathcal{S}_q \cup \{B_1, \ldots, B_{t_q}\},$$

where $B_i \cap E_{q+1} = \emptyset, i = 1, \ldots, t_q$ and $\mathcal{S}_q = \{A \in \mathcal{L}_q \mid A \cap E_{q+1} \neq \emptyset\}$.

For each i $(i = 1, \ldots, t_q)$ constructs the set $\{B_i \cup \{b\} \mid b \in E_{q+1}\}$. Denote them by $A^i_1, \ldots, A^i_{r_i} (i = 1, \ldots, t_q)$. Let

$$\mathcal{L}_{q+1} = \mathcal{S}_q \cup \{A^i_p \mid A \in \mathcal{S}_q \Rightarrow A \not\subset A^i_p, 1 \leqslant i \leqslant t_q, 1 \leqslant p \leqslant r_i\}.$$

Theorem 2.4. ([3]) *For every q $(1 \leqslant q \leqslant m)$ $\mathcal{L}_q = Tr(\{E_1, \ldots, E_q\})$, i.e., $\mathcal{L}_m = Tr(\mathcal{H})$.*

It can be seen that the determination of $Tr(\mathcal{H})$ based on our algorithm does not depend on the order of $E_1, \ldots, E_m$.

Remark 2.5.([3]) Denote $\mathcal{L}_q = \mathcal{S}_q \cup \{B_1, \ldots, B_{t_q}\}$, and l_q $(1 \leqslant q \leqslant m-1)$ be the number of elements of $\mathcal{L}_q$. It can be seen that the worst-case time complexity of our algorithm is

$$\mathcal{O}(|U|^2 \cdot \sum_{q=0}^{m-1} t_q u_q),$$

where $l_0 = t_0 = 1$ and

$$u_q = \begin{cases} l_q - t_q, & \text{if } l_q > t_q; \\ 1, & \text{if } l_q = t_q. \end{cases}$$

Clearly, in each step of our algorithm $\mathcal{L}_q$ is a simple hypergraph. It is known that the size of arbitrary simple hypergraph over U cannot be greater than $C_n^{[n/2]}$, where $n = |U|$. $C_n^{[n/2]}$ is asymptotically equal to $2^{n+1/2}/(\pi.n)^{1/2}$. From this, the worst-case time complexity of our algorithm cannot be more than exponential in the number of attributes. In cases for which $l_q \leqslant l_m$ $(q = 1, \ldots, m-1)$, it is easy to see that the time complexity of our algorithm is not greater than $\mathcal{O}(|U|^2 \cdot |\mathcal{H}| \cdot |Tr(\mathcal{H})|^2)$. Thus, in these cases this algorithm finds $Tr(\mathcal{H})$ in polynomial time in $|U|, |\mathcal{H}|$ and $|Tr(\mathcal{H})|$. Obviously, if the number of elements of $\mathcal{H}$ is small, then this algorithm is very effective. It only requires polynomial time in $|U|$.

The following proposition is obvious.

Proposition 2.6.([3]) *The time complexity of finding $Tr(\mathcal{H})$ of a given hypergraph $\mathcal{H}$ is (in general) exponential in the number of elements of U.*

Proposition 2.6 is still true for a simple hypergraph.

3. MINIMAL KEYS

In this section, we investigate the minimal keys of relation schemes. We give some descriptions of the set of all minimal keys of relation schemes in terms of hypergraphs.

Let $s = (U, F)$ be a relation scheme. We set $\mathcal{L}_s = \{X^+ \mid X \subseteq U\}$, i.e., $\mathcal{L}_s$ is the set of all closures of s. We define the family $\mathcal{M}_s$ as follows

$$\mathcal{M}_s = \mathcal{L}_s - \{U\}.$$

Then $\overline{\mathcal{M}_s} = \{U - A \mid A \in \mathcal{M}_s\}$ is called the *complemented family* of $\mathcal{M}_s$.

Lemma 3.1. *Let $s = (U, F)$ be a relation scheme. Then, if $A \in \overline{\mathcal{M}_s}$ then $U - A$ is not the key of s.*

Proof. Assume that $A \in \overline{\mathcal{M}_s}$. Thus, $U - A \in \mathcal{M}_s$. By the definition of $\mathcal{M}_s$, we have

$$(U - A)^+ = U - A$$

and

$$U - A \neq U.$$

Consequently, $U - A$ is not a key of s.

The lemma is proved. ■

Lemma 3.2. *Let $s = (U, F)$ be a relation scheme. Then, $A \in Trs(\mathcal{K}_s)$ if and only if $U - A$ is not the key of s.*

Proof. Suppose that $U - A$ is a key of s. From this and the hypothesis $A \in Trs(\mathcal{K}_s)$, we have

$$A \cap (U - A) \neq \emptyset.$$

This is a conflict.

Conversely, assume that $A \notin Trs(\mathcal{K}_s)$. If there exists $K \in \mathcal{K}_s$ such that $A \cap K = \emptyset$ then $U - A$ is a key of s, which contradicts the hypothesis $U - A$ is not the key of s.

The lemma is proved. ■

Theorem 3.3. *Let $s = (U, F)$ be a relation scheme. Then*

$$Tr(\mathcal{K}_s) = \min(\overline{\mathcal{M}_s}).$$

Proof. Suppose that $A \in Tr(\mathcal{K}_s)$. By Lemma 3.2 we obtain which $U - A$ is not a key of s. Clearly, $A \neq \emptyset$ and $(U - A)^+ \neq A$. On the other hand, we have also

$$U - (U - A)^+ \cap K \neq \emptyset \quad \forall K \in \mathcal{K}_s.$$

Hence, if

$$U - A \subset (U - A)^+$$

then

$$A \supset U - (U - A)^+.$$

This contradicts with the hypothesis $A \in Tr(\mathcal{K}_s)$. Consequently, $(U - A)^+ = U - A$, i.e., $U - A \in \mathcal{M}_s$. Thus, $A \in \overline{\mathcal{M}_s}$.

Now we assume that there exists a $B \subset A$ and $B \neq \emptyset$ such that $B \in \overline{\mathcal{M}_s}$. Then, according to Lemma 3.1 we have $U - B$ is not a key of s. By Lemma 3.2 we obtain $B \in Trs(\mathcal{K}_s)$, which contradicts the fact that $A \in Tr(\mathcal{K}_s)$. Therefore, $A \in \min(\overline{\mathcal{M}_s})$ holds.

Conversely, assume that $A \in \min(\overline{\mathcal{M}_s})$. Hence, $A \in \overline{\mathcal{M}_s}$. By Lemma 3.1 we have $U - A$ is not a key of s. Thus, according to Lemma 3.2 we obtain $A \in Trs(\mathcal{K}_s)$. Suppose that there is a $B \subset A$ such that $B \in Tr(\mathcal{K}_s)$. By the above proof we obtain $B \in \overline{\mathcal{M}_s}$. This contradicts with the fact that $A \in \min(\overline{\mathcal{M}_s})$. Hence, $A \in Tr(\mathcal{K}_s)$ holds.

The theorem is proved. ■

By Proposition 2.1 and Theorem 3.3, the following corollary is immediate.

Corollary 3.4. *Let $s = (U, F)$ be a relation scheme. Then*

$$\mathcal{K}_s = Tr(\min(\overline{\mathcal{M}_s})).$$

Theorem 3.5. *Let $s = (U, F)$ be a relation scheme. Then*

$$\mathcal{K}_s = Tr(\min(\overline{\mathcal{L}_s} - \{\emptyset\})).$$

Proof. It is clear that from the definiton of $\mathcal{M}_s$ and Corollary 3.4.

The theorem is proved. ■

4. ANTIKEYS

In this section, we study the set of antikeys by hypergraphs. We present connections between the set of antikeys and the set of closures of relation schemes.

Let $\mathcal{A}$ be a family of subsets of U. We define

$$\min(\mathcal{A}) = \{A_i \in \mathcal{A} \mid \nexists A_j : A_j \subset A_i\}$$

and

$$\max(\mathcal{A}) = \{A_i \in \mathcal{A} \mid \nexists A_j : A_j \supset A_i\}.$$

Lemma 4.1. *Let $\mathcal{A}$ be a family of subsets of U. Then*

$$\min(\overline{\mathcal{A}}) = \overline{\max(\mathcal{A})}.$$

Proof. We shall prove that $\overline{\min(\overline{\mathcal{A}})} = \max(\mathcal{A})$. Suppose $A \in \overline{\min(\overline{\mathcal{A}})}$. Hence, $\overline{A} \in \min(\overline{\mathcal{A}})$. This means that

$$\forall B \in \overline{\mathcal{A}} : B \not\subset \overline{A}$$

or

$$\forall \overline{B} \in \mathcal{A} : \overline{B} \not\supset A.$$

Thus, we obtain $A \in \max(\mathcal{A})$.

On the other hand, let $A \in \max(\mathcal{A})$. By an argument analogous to the previous one, we get $A \in \overline{\min(\overline{\mathcal{A}})}$.

The lemma is proved. ■

Theorem 4.2. *Let $s = (U, F)$ be a relation scheme. Then*

$$\overline{Tr(\mathcal{K}_s)} = \max(\mathcal{M}_s).$$

Proof According to Theorem 3.3 we have

$$Tr(\mathcal{K}_s) = \min(\overline{\mathcal{M}_s}).$$

From this and Lemma 4.1, we obtain

$$\overline{Tr(\mathcal{K}_s)} = \max(\mathcal{M}_s).$$

The theorem is proved. ■

The Theorem 4.2 means that

$$\forall X^+ \subset U, \exists A \in \overline{Tr(\mathcal{K}_s)} : X^+ \subseteq A.$$

Note that the following result is known [4].

Proposition 4.3. *Let $s = (U, F)$ be a relation scheme. Then*

$$\mathcal{K}_s^{-1} = \overline{Tr(\mathcal{K}_s)}.$$

Therefore, by Theorem 4.2 and Proposition 4.3, the following corollary is evident.

Corollary 4.4. *Let $s = (U, F)$ be a relation scheme. Then*

$$\mathcal{K}_s^{-1} = \max(\mathcal{M}_s).$$

5. CONCLUSION

We have characterized the set of all minimal keys of relation schemes in terms of hypergraphs. Futhermore, the set of antikeys is also studied in this paper. We present connections between the set of antikeys and the set of closures of relation schemes.

TÀI LIỆU THAM KHẢO

[1] Armstrong W. W., Dependency structure of database relationship, *Information Processing 74*, North-Holland Pub. Co., (1974) 580–583.

[2] Berge C., *Hypergraphs: combinatorics of finite sets*, North - Holland, Amsterdam (1989).

[3] Demetrovics J., Thi V. D., Describing candidate keys by hypergraphs, *Computers and Artificial Intelligence* **18** (2) (1999) 191–207.

[4] Thi V. D., Son N. H., Some problems related to keys and the Boyce-Codd normal form, *Acta Cybernetica* **16** (3) (2004) 473–483.

[5] Thi V. D., Son N. H., Some results related to dense families of database relations, *Acta Cybernetica* **17** (1) (2005) 173–182.

Received on June 06, 2005