

INVESTIGATION ON ANTIKEYS AND MINIMAL KEYS OF RELATION SCHEMES BY HYPERGRAPHS

Nguyen Hoang Son (Hue, Vietnam)

Abstract. Minimal keys and antikeys play a very important role in the theory of the design of relational databases (see, e.g. [3, 6, 7]). The minimal key and antikey results have been widely investigated. Hypergraphs theory (see, e.g. [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 hypergraphs. The aim of this paper is to investigate the minimal keys of relation schemes in term of hypergraphs. 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.

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, 7].

Let U be a finite set of *attributes* (e.g. name, age, etc). The elements of U will be denoted by a, b, c, \dots, x, y, z , if an ordering on U is needed, by a_1, \dots, a_n . A map dom associates with each $a \in U$ its *domain* $dom(a)$. A *relation* R on U is a subset of Cartesian product $\prod_{a \in U} dom(a)$.

We can think of a relation R on U as a set of tuples: $R = \{h_1, \dots, h_m\}$,

$$h_i : U \rightarrow \bigcup_{a \in U} dom(a), \quad h_i(a) \in dom(a), \quad i = 1, 2, \dots, m.$$

A *functional dependency* (FD for short) on U is a statement of form $X \rightarrow Y$, where $X, Y \subseteq U$. The FD $X \rightarrow Y$ *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 $X \rightarrow Y$. Let F_R be a family of all FDs that holds in R . Then $F = F_R$ satisfies

- (F1) $X \rightarrow X \in F$,
- (F2) $(X \rightarrow Y \in F, Y \rightarrow Z \in F) \Rightarrow (X \rightarrow Z \in F)$,
- (F3) $(X \rightarrow Y \in F, X \subseteq V, W \subseteq Y) \Rightarrow (V \rightarrow W \in F)$,
- (F4) $(X \rightarrow Y \in F, V \rightarrow W \in F) \Rightarrow (X \cup V \rightarrow Y \cup W \in F)$.

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

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

Given a family F of FDs on 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 set of attributes and F is a set of FDs on U . Denote $X^+ = \{a \in U \mid A \rightarrow \{a\} \in F^+\}$. X^+ is called the *closure* of X on s . It is obvious that $X \rightarrow Y \in F^+$ iff $Y \subseteq X^+$.

Let $s = (U, F)$ be a relation scheme and $K \subseteq U$. Then K is a *key* of s if $K \rightarrow U \in F^+$. 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 . Now we define the set of *antikeys* of \mathcal{K}_s , denoted by \mathcal{K}_s^{-1} , as follows:

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

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 $\mathcal{H} = \{E_i \mid E_i \in \mathcal{P}(U), i = 1, 2, \dots, m\}$ is called a *hypergraph*

on U if $E_i \neq \emptyset$ holds for all i (in [2] it is required that the union of E_i 's is U , in this paper we do not require this).

The elements of U are called vertices, and the sets E_1, \dots, 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 system [4]. Clearly, \mathcal{K}_s and \mathcal{K}_s^{-1} are simple hypergraphs.

In this paper we always assume that if a simple hypergraph plays the role of set of minimal keys (antikeys), then this simple hypergraph is not empty (does not contain U).

Let \mathcal{H} be a hypergraph on 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 transversal 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} be two simple hypergraphs on U . Then*

- (1) $\mathcal{H} = Tr(\mathcal{G})$ if and only if $\mathcal{G} = Tr(\mathcal{H})$,
- (2) $Tr(\mathcal{H}) = Tr(\mathcal{G})$ if and only if $\mathcal{H} = \mathcal{G}$,
- (3) $Tr(Tr(\mathcal{H})) = \mathcal{H}$.

By the definition of minimal transversal the following proposition is obvious.

Proposition 2.2. [8] *Let \mathcal{H} be a hypergraph on 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. [4]

Input: let $\mathcal{H} = \{E_1, \dots, E_m\}$ be a hypergraph on 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

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

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

For each i ($i = 1, \dots, t_q$) constructs the set $\{B_i \cup \{b\} \mid b \in E_{q+1}\}$. Denote them by $A_1^i, \dots, A_{r_i}^i$ ($i = 1, \dots, t_q$). Let

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

Theorem 2.4. ([4]) *For every q ($1 \leq q \leq m$) $\mathcal{L}_q = Tr(\{E_1, \dots, E_q\})$, 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, \dots, E_m .

Remark 2.5. ([4]) Denote $\mathcal{L}_q = \mathcal{S} \cup \{B_1, \dots, B_{t_q}\}$, and l_q ($1 \leq q \leq m-1$) be the number of elements of \mathcal{L}_q . It can be seen that the worst-case time complexity of our algorithm is

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

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 on U cannot be greater than $C_n^{\lfloor n/2 \rfloor}$, where $n = |U|$. $C_n^{\lfloor n/2 \rfloor}$ is asymptotically equal to $2^{n+1/2}/(\pi \cdot n)^{1/2}$. From this, the worst-case time complexity of our algorithm cannot be more than exponential in the number of attributes. In cases for which $l_q \leq l_m$ ($q = 1, \dots, m-1$), 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. ([4]) *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. Let \mathcal{H} be a simple hypergraph on U . Now we define a set \mathcal{H}^{-1} as follows:

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

It is easy to see that if \mathcal{H}^{-1} is a hypergraph on U , then \mathcal{H}^{-1} is a simple hypergraph.

Proposition 2.7. *Let \mathcal{H} be a simple hypergraph on U . Then*

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

Proof. Suppose that $A \in \mathcal{H}^{-1}$. By the definition of \mathcal{H}^{-1} we have $\bar{A} \cap E \neq \emptyset$ for every $E \in \mathcal{H}$, which mean that $\bar{A} \in Trs(\mathcal{H})$. On the other hand, according to the definition of \mathcal{H}^{-1} , there exists an $E \in \mathcal{H}$ such that

$$A \cup \{a\} \supseteq E \quad \forall a \in \bar{A},$$

i.e. $(\bar{A} \setminus \{a\}) \cap E = \emptyset$. Therefore we obtain $\bar{A} \in Tr(\mathcal{H})$ or $A \in \overline{Tr(\mathcal{H})}$.

Conversely, assume that $T \in Tr(\mathcal{H})$. Thus

$$\bar{T} \not\supseteq E \quad \forall E \in \mathcal{H}.$$

By the definition of the transversal hypergraph, it is obvious that $T \setminus \{a\} \notin \text{Trs}(\mathcal{H})$ for every $a \in T$, or $(T \setminus \{a\}) \cap E = \emptyset$ for some $E \in \mathcal{H}$. This means that

$$\overline{T} \cup \{a\} \supseteq E \quad \forall a \in T.$$

Consequently, according to the definition of \mathcal{H}^{-1} , we have $\overline{T} \in \mathcal{H}^{-1}$. The proposition is proved.

Note that, if $\mathcal{H} = \emptyset$ then $\overline{\text{Tr}(\mathcal{H})} = \{\emptyset\} = \{U\}$. On the other hand, according to the definition of \mathcal{H}^{-1} , we have $\mathcal{H}^{-1} = \{U\}$. Consequently, if $\mathcal{H} = \emptyset$ then we have also $\mathcal{H}^{-1} = \overline{\text{Tr}(\mathcal{H})}$.

3. Minimal keys

In this section we investigate the minimal keys of relation schemes. We give two descriptions of the set of all minimal keys of relation schemes in term 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* 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, $A \in \text{Trs}(\mathcal{K}_s)$ 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 $A \in \text{Trs}(\mathcal{K}_s)$ we have

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

This is a contradiction.

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 also have

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

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 $B \subset A$ and $B \neq \emptyset$ such that $B \in \overline{\mathcal{M}_s}$. Then, according to Lemma 3.1, $U - B$ 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, $A \in Trs(\mathcal{K}_s)$. Suppose that there is a $B \subset A$ such that $B \in Trs(\mathcal{K}_s)$.

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.

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

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

Proof. (1) It is obvious from Proposition 2.2 and Theorem 3.3.

(2) It is clear that from the definition of \mathcal{M}_s and (1).

The theorem is proved.

4. Antikeys

In this section, firstly, 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

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

and

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

where $\overline{\mathcal{A}}$ is the complemented of \mathcal{A} .

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 $A \in \overline{\max(\mathcal{A})}$. By an argument analogous to the previous one, we get $A \in \min(\overline{\mathcal{A}})$. The lemma is proved.

Lemma 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

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

The lemma is proved.

The Lemma 4.2 means that

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

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

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

Proof. Because \mathcal{K}_s is a simple hypergraph, and by Proposition 2.7, we have

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

From this and Lemma 4.2, we immediately get

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

The theorem is proved.

Note that a set of minimal keys and set of antikeys form simple hypergraphs. Let \mathcal{H} be a simple hypergraph on U . The time complexity of finding \mathcal{H}^{-1} is (in general) exponential in the number of elements of U [4]. However, if we restrict the number of elements of \mathcal{H} , then the time complexity of finding \mathcal{H}^{-1} is polynomial time.

Lemma 4.4. *Let \mathcal{H} be a simple hypergraph on U . If $|\mathcal{H}| \leq c$ (c is a constant) then \mathcal{H}^{-1} is computable in polynomial time.*

Proof. Assume that $\mathcal{H} = \{E_1, \dots, E_c\}$ where $c \geq 1$. Certainly, $E_i \neq \emptyset$ for all $i = 1, 2, \dots, c$. We construct the set

$$\mathcal{G} = \{\{a_1\} \cup \dots \cup \{a_k\} \mid a_i \in E_i, 1 \leq i \leq c\}.$$

Denote elements of \mathcal{G} by G_1, \dots, G_t . Clearly, $G_i \in Trs(\mathcal{H})$ for all $i = 1, 2, \dots, t$. Then, we compute

$$\min(\mathcal{G}) = \{G_i \in \mathcal{G} \mid \nexists G_j \in \mathcal{G} : G_j \subset G_i\}.$$

According to the definition of transversal hypergraphs, we have

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

By Proposition 2.7, we obtain

$$\mathcal{H}^{-1} = \overline{\min(\mathcal{G})}.$$

Obviously, $|\mathcal{G}| \leq |U|^c$. Consequently, $\overline{\min(\mathcal{G})}$ is computable in polynomial time. The lemma is proved.

Algorithm 4.5. (Finding \mathcal{H}^{-1})

Input: let $\mathcal{H} = \{E_1, \dots, E_c\}$ be a simple hypergraph on U , where c is a constant.

Output: \mathcal{H}^{-1} .

Method:

Step 1. We construct the set

$$\mathcal{G} = \{\{a_1\} \cup \dots \cup \{a_c\} \mid a_i \in E_i, 1 \leq i \leq c\}.$$

Step 2. Compute

$$\min(\mathcal{G}) = \{G_i \in \mathcal{G} \mid \nexists G_j \in \mathcal{G} : G_j \subset G_i\}.$$

Step 3. Let $\mathcal{H}^{-1} = \overline{\min(\mathcal{G})}$.

By Lemma 4.4, it is clear that Algorithm 4.5 computes \mathcal{H}^{-1} . Furthermore, the time complexity of Algorithm 4.5 is polynomial time in the size of U . Obviously, if c is small then our algorithm is very effective.

5. Conclusion

We have characterized the set of all minimal keys of relation schemes in term of hypergraphs. Furthermore, 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.

It can be seen that, if the number of elements of \mathcal{H} is constant, i.e. $|\mathcal{H}| \leq c$ for some constant c , then the time complexity of finding \mathcal{H}^{-1} of a given hypergraph \mathcal{H} is polynomial time.

References

- [1] **Armstrong W.W.**, Dependency structures of database relationships, *Information Processing*, **74** (1974), 580-583.

- [2] **Berge C.**, *Hypergraphs: combinatorics of finite sets*, North-Holland, Amsterdam, 1989.
- [3] **Demetrovics J. and Thi V.D.**, Keys, antikeys and prime attributes, *Annales Univ. Sci. Budapest. Sect. Comp.*, **8** (1987), 35-52.
- [4] **Demetrovics J. and Thi V.D.**, Describing candidate keys by hypergraphs, *Computers and Artificial Intelligence*, **18** (2) (1999), 191-207.
- [5] **Demetrovics J. and Thi V.D.**, Some computational problems related to Boyce-Codd normal form, *Annales Univ. Sci. Budapest. Sect. Comp.*, **19** (2000), 119-132.
- [6] **Lucchesi C.L. and Osborn S.L.**, Candidate keys for relations, *J. Comput. Syst. Scien.*, **17** (2) (1978), 270-279.
- [7] **Thi V.D.**, Minimal keys and antikeys, *Acta Cybernetica*, **7** (4) (1986), 361-371.
- [8] **Thi V.D. and Son N.H.**, Some results related to dense families of database relations, *Acta Cybernetica*, **17** (1) (2005), 173-182.

(Received December 1, 2005)

Nguyen Hoang Son
Department of Mathematics
College of Sciences
Hue University
77 Nguyen Hue, Hue, Vietnam
nhson@hueuni.edu.vn