Some remarks on the algorithm of Lucchesi and Osborn

HO THUAN

Let $S=\langle\Omega, F\rangle$ be a relation scheme, where $\Omega=\{A_1, A_2, \ldots, A_n\}$ is the universe of attributes and

$$F = \{L_i \to R_i | L_i, R_i \subseteqq \Omega,\ i = 1, 2, \ldots, m\}$$

is the set of functional dependencies. In [2] C. L. Lucchesi and S. L. Osborn provided a very interesting algorithm to determine the set of all keys for any relation scheme $S=\langle\Omega, F\rangle$. Following our notation, the algorithm has time complexity

$$O(|F|\,|\mathscr{K}_S|\,|\Omega|(|\mathscr{K}_S|+|\Omega|)),$$

i.e. its running time is bounded by a polynomial of $|\Omega|$, $|F|$ and $|\mathscr{K}_S|$, where

$|F|$ is the cardinality of F, and

$\mathscr{K}_S$ is the set of all keys for S.

We reproduce here this algorithm with some modifications in accordance with our notation.

Algorithm OL1. Set of all keys for $S=\langle\Omega, F\rangle$;

Comment. $\mathscr{K}_S$ is the set of keys being accumulated in a sequence which can be scanned in the order in which the keys are entered;

$$\mathscr{K}_S \leftarrow \{\text{Key}(\Omega, F, \Omega)\};^{1}$$

for each K in $\mathscr{K}_S$ **do**

  for each $FD(L_i \to R_i)$ in F **do**

    $T \leftarrow L_i \cup (K \setminus R_i)$;

    test $\leftarrow$ true;

    for each J in $\mathscr{K}_S$ **do**

[1] Let Key (Ω, F, X) be the algorithm Minimal Key in [2], which determines a key for S that is a subset of a specified superkey X.

if T includes J **then** test $\leftarrow$ false;

if test **then** $\mathcal{K}_S \leftarrow \mathcal{K}_S \cup \{\text{Key}(\Omega, F, T)\}$

end

end;

return $\mathcal{K}_S$

The following simple remarks, in some cases can be used to improve the performance of the algorithm of Lucchesi and Osborn.

Remark 1. Let L, R, H be defined as: $R=\bigcup_{i=1}^{m} R_i$, $L=\bigcup_{i=1}^{m} L_i$ $H=\bigcup_{K_j\in\mathcal{K}_S} K_j$. To find the first key for $S=\langle\Omega, F\rangle$, instead of Ω it is better to use the superkey $(\Omega\setminus R)\cup(L\cap R)$ and Algorithm 1 in [1], and instead of the algorithm Key (Ω, F, T) it is better to use Algorithm 2 in [1] for finding one key for S included in a given superkey T.

Remark 2. In [1] it is shown that

$$R\setminus L \subseteq \Omega\setminus H,$$

i.e. $R\setminus L$ consists only of non-prime attributes. Therefore, if $R_i \subseteq R\setminus L$ then $R_i\cap K=\emptyset$, $\forall K\in\mathcal{K}_S$, and $L_i\cup(K\setminus R_i)\supseteq K$. That means, when computing $T=L_i\cup \cup(K\setminus R_i)$, we can neglect all FDs $(L_i\rightarrow R_i)$ with $R_i\subseteq R\setminus L$ for every $K\in\mathcal{K}_S$.

Let us denote

$$\bar{F} = F\setminus\{L_j \rightarrow R_j | L_j \rightarrow R_j \in F \quad \text{and} \quad R_j \subseteq R\setminus L\}.$$

Remark 3. With a fixed K in $\mathcal{K}_S$, it is clear that if $K\cap R_i=\emptyset$ then $L_i\cup(K\setminus R_i)\supseteq K$. In that case, it is not necessary to continue to check whether T includes J for each J in $\mathcal{K}_S$. So, it is better to compute T by the following order

$$T = (K\setminus R_i)\cup L_i.$$

Remark 4. The algorithm of Lucchesi and Osborn is particularly effective when the number of keys for $S=\langle\Omega, F\rangle$ is small. But on what basis can we conclude that the number of keys for S is small? There is no general answer for all cases, and it is shown in [3] that the number of keys for a relation scheme $S=\langle\Omega, F\rangle$ can be factorial in $|F|$ or exponential in $|\Omega|$, and that both of these upper bounds are attainable. However, it is shown in ([1], Corollary 1) that

$$|\mathcal{K}_S| \leqq C_h^{[h/2]}$$

where h is the cardinality of $L\cap R$. Thus, if $L\cap R$ has a few elements only, then it is a good criterion for saying that S has a small number of keys. In the case $L\cap R=\emptyset$, $\Omega\setminus R$ is the unique key for $S=\langle\Omega, F\rangle$ as pointed out in ([1], Corollary 4).

Example. We take up the example in [2], Appendix 1):

$$\Omega = \{a, b, c, d, e, f, g, h\},$$

$$F = \{a \rightarrow b,\ c \rightarrow d,\ e \rightarrow f,\ g \rightarrow h\}.$$

It is clear that for this relation scheme

$$L \cap R = \emptyset,$$

and it has exactly one key, namely $\Omega \setminus R = aceg$.

Taking Remarks 1—3 into account, the algorithm of Lucchesi and Osborn now can be presented as follows:

Algorithm OL2. Set of all keys for $S = \langle \Omega, F \rangle$;

$$\mathcal{K}_S \leftarrow \{\text{Algo. } 1\,(\Omega, F, (\Omega \setminus R) \cup (L \cap R))\};$$ [2]

for each K in $\mathcal{K}_S$ **do**

for each $FD(L_i \to R_i)$ in $\bar{F}$ such that $K \setminus R_i \neq K$ **do**

$T \leftarrow (K \setminus R_i) \cup L_i$;

test $\leftarrow$ true;

for each J in $\mathcal{K}_S$ **do**

if T includes J **then** test $\leftarrow$ false;

if test **then** $\mathcal{K}_S \leftarrow \mathcal{K}_S \cup \{\text{Algo. } 2(\Omega, F, T)\}$

end

end;

return $\mathcal{K}_S$.

Remark 5. The time complexity of Algorithm OL2 is

$$0(|\mathcal{K}_S||\Omega|(|\mathcal{K}_S||\bar{F}| + |F||L \cap R|)).$$

Abstract

In [1] we have proposed two algorithms (Algorithm 1 and Algorithm 2) for finding one key of the relation scheme $S = \langle \Omega, F \rangle$ included in a given superkey. In this paper, we show that, using these algorithms and some simple remarks, the performance of the algorithm of Lucchesi and Osborn [2], in general, can be improved.

References

[1] Ho Thuan and Le van Bao, Some results about keys of relational schemes, Acta Cybernetica Tom. 7, Fasc. 1, Szeged, 1985, 99—113.

[2] Lucchesi, C. L. and Osborn, S. L., Candidate keys for relations, J. of Computer and System Sciences, 17, 1978, 270—279.

[3] Osborn, S. L., Normal forms for relational databases, Ph. D. Dissertation, University of Waterloo, 1977.

(Received Oct. 27, 1986)

[2] Algo. 1 and Algo. 2 refer to Algorithm 1 and Algorithm 2 in [1] respectively.