

Một số kết quả về thuật toán tính bao đóng và rút gọn bài toán tìm khóa của lược đồ quan hệ

Vũ Quốc Tuấn, Hồ Thuần

Học viện Khoa học và Công nghệ, Viện Hàn lâm Khoa học và Công nghệ Việt Nam

E-mail: vqtuanhd@gmail.com, hothuan1812@yahoo.com

Tác giả liên hệ: Vũ Quốc Tuấn

Ngày nhận: 27/03/2017, ngày sửa chữa: 04/08/2017, ngày duyệt đăng: 13/11/2017

Tóm tắt: Trong lĩnh vực trí tuệ nhân tạo và cơ sở dữ liệu quan hệ, khóa của lược đồ quan hệ và bao đóng của một tập thuộc tính đối với một tập phụ thuộc hàm có vai trò quan trọng và được sử dụng trong nhiều bài toán như tối ưu hóa truy vấn, chuẩn hóa lược đồ quan hệ, loại bỏ ràng buộc dư thừa, v.v. Do đó, độ phức tạp của thuật toán tìm bao đóng và việc rút gọn bài toán tìm khóa là các vấn đề luôn được quan tâm. Trong vài năm gần đây, các vấn đề này được xới lại với hàng loạt các công trình mới nhằm giải quyết bài toán tính bao đóng và tìm tập các khóa của lược đồ quan hệ một cách hiệu quả hơn theo nhiều tiếp cận khác nhau. Trong bài báo này, chúng tôi đề xuất một thuật toán cải tiến tính bao đóng và đưa ra một số kết quả về rút gọn bài toán tìm khóa nhằm nâng cao hiệu năng tính toán khi giải quyết các vấn đề có liên quan.

Từ khóa: Cơ sở dữ liệu quan hệ, lược đồ quan hệ, phụ thuộc hàm, bao đóng của một tập thuộc tính, khóa của lược đồ quan hệ.

Title: Some Results for the Closure Computing Algorithm and Reducing the Key Finding Problem of a Relation Schema

Abstract: In artificial intelligence and relational databases, the key for a relation schema and the closure of a set of attributes under a set of functional dependencies are important and used in several problems such as query optimization, relational schema normalization, removing redundant constraints, etc. Therefore, the complexity of closure computing algorithms and reducing the key finding problem are always interesting problems. In recent years, these problems have been revisited with a series of new studies, to be solved more efficiently by several different approaches. In this paper, we propose an improved closure computing algorithm and provide some results for reducing the key finding problem to enhance the computing performance for solving related problems.

Keywords: Relational database, relation schema, functional dependency, closure of a set of attributes, key for a relation schema.

I. MỞ ĐẦU

Phần này nhắc lại một số khái niệm quan trọng trong lý thuyết cơ sở dữ liệu quan hệ nhằm mục đích sử dụng cho các phần tiếp theo.

Lược đồ quan hệ. Một lược đồ quan hệ S là một cặp có thứ tự $S = \langle \Omega, F \rangle$, trong đó Ω là tập hữu hạn các thuộc tính của quan hệ, F là tập các ràng buộc giữa các thuộc tính.

Cho lược đồ quan hệ $S = \langle \Omega, F \rangle$ với $\Omega = \{A_1, A_2, \dots, A_n\}$. Nếu không quan tâm đến tập các ràng buộc F thì ta sẽ dùng ký hiệu $S(\Omega)$ thay cho $S = \langle \Omega, F \rangle$.

Ta dùng ký hiệu $r(S)$ để chỉ một quan hệ r (hay một thể hiện r) của lược đồ quan hệ S . Với một bộ t của $r(S)$ và $X \subseteq \Omega$, ta ký hiệu $t[X]$ là bộ chỉ chứa các giá trị của bộ t tại các thuộc tính trong X .

Phụ thuộc hàm. Cho Ω là tập thuộc tính và $S(\Omega)$ là một lược đồ quan hệ trên Ω . Giả sử $X, Y \subseteq \Omega$, khi đó Y được gọi là *phụ thuộc hàm* vào X trên lược đồ $S(\Omega)$, ký hiệu là $X \rightarrow Y$, nếu với mọi quan hệ r trên lược đồ $S(\Omega)$, với hai bộ bất kỳ $t_1, t_2 \in r$ mà $t_1[X] = t_2[X]$ thì $t_1[Y] = t_2[Y]$.

Nếu Y phụ thuộc hàm vào X thì ta cũng nói “ X xác định hàm Y ”. Với mỗi quan hệ r trên lược đồ $S(\Omega)$, ta nói r *thỏa mãn* (hay *thỏa*) phụ thuộc hàm $X \rightarrow Y$ (hay phụ thuộc hàm $X \rightarrow Y$ đúng trên r) nếu và chỉ nếu với mọi bộ $t_1, t_2 \in r$, $t_1[X] = t_2[X]$ kéo theo $t_1[Y] = t_2[Y]$. Trong bài báo này, ta hạn chế F (của lược đồ $S = \langle \Omega, F \rangle$) chỉ gồm các phụ thuộc hàm.

Hệ quy tắc suy diễn Armstrong. Với lược đồ quan hệ $S = \langle \Omega, F \rangle$ và $X, Y \subseteq \Omega$, ta ký hiệu XY thay cho $X \cup Y$. Với mọi $X, Y, Z \subseteq \Omega$, hệ quy tắc suy diễn Armstrong đối

Thuật toán 1: Tính bao đóng X^+

Input:
 $\Omega, F, X \subseteq \Omega$
Output: X^+
begin
 $X^+ = X$
repeat
 for each $(Y \rightarrow Z) \in F$ **do**
 if $Y \subseteq X^+$ **then**
 $X^+ = X^+ \cup Z$;
 end if
 end for each
until không còn thuộc tính nào được thêm vào X^+ ;
return X^+ ;
end

với các phụ thuộc hàm gồm ba quy tắc sau đây:

- A1. (Phản xạ):** Nếu $Y \subseteq X$ thì $X \rightarrow Y$;
A2. (Gia tăng): Nếu $X \rightarrow Y$ thì $XZ \rightarrow YZ$;
A3. (Bắc cầu): Nếu $X \rightarrow Y$ và $Y \rightarrow Z$ thì $X \rightarrow Z$.

Ký hiệu F^+ là tập tất cả các phụ thuộc hàm được suy diễn từ F bằng cách áp dụng một số hữu hạn lần các quy tắc của hệ quy tắc suy diễn Armstrong.

Bao đóng của một tập thuộc tính. Cho tập phụ thuộc hàm F xác định trên tập thuộc tính Ω (phụ thuộc hàm (FD: functional dependency) $Y \rightarrow Z$ xác định trên tập thuộc tính Ω nếu $Y, Z \subseteq \Omega$) và $X \subseteq \Omega$. Ta gọi *bao đóng* của tập thuộc tính X đối với tập phụ thuộc hàm F , ký hiệu là X_F^+ , là tập tất cả các thuộc tính A của Ω sao cho $X \rightarrow A$ được suy diễn từ F nhờ hệ quy tắc suy diễn Armstrong,

$$X_F^+ = \{A \in \Omega \mid (X \rightarrow A) \in F^+\}.$$

Khóa của lược đồ quan hệ. Cho lược đồ quan hệ $S = \langle \Omega, F \rangle$ và $K \subseteq \Omega$. Ta nói K là một *khóa* của S nếu hai điều kiện sau đây đồng thời được thỏa mãn:

- i) $(K \rightarrow \Omega) \in F^+$;
ii) Nếu $K' \subset K$ thì $(K' \rightarrow \Omega) \notin F^+$.

II. THUẬT TOÁN CẢI TIẾN TÍNH BAO ĐÓNG CỦA MỘT TẬP THUỘC TÍNH

1. Thuật toán chuẩn tính bao đóng X^+

Thuật toán 1 là thuật toán tính bao đóng X^+ . Để chứng minh được rằng độ phức tạp thời gian của Thuật toán 1 là $O(np \min\{n, p\})$, trong đó n là số thuộc tính trong Ω và p là số phụ thuộc hàm trong F . Như vậy, thuật toán trên không là tuyến tính theo tích np . Để có các thuật toán tính bao đóng với độ phức tạp tuyến tính, cần sử dụng một số cấu trúc dữ liệu thích hợp nhằm giảm chi phí của việc duyệt các tập F và Ω . Các chiến lược rút gọn bao gồm:

Thuật toán 2: Tính bao đóng X^+ [4]

Input:
 $\Omega, F, X \subseteq \Omega$
Output: X^+
begin
 $X_{\text{new}} = X$;
repeat
 $X_{\text{old}} = X_{\text{new}}$;
 for each $(Y \rightarrow Z) \in F$ **do**
 if $Y \subseteq X_{\text{new}}$ **then**
 $X_{\text{new}} = X_{\text{new}} \cup Z$; ▷ (I)
 $F = F - \{Y \rightarrow Z\}$;
 else if $Z \subseteq X_{\text{new}}$ **then**
 $F = F - \{Y \rightarrow Z\}$; ▷ (II)
 else
 $F = F - \{Y \rightarrow Z\}$; ▷ (III)
 $F = F \cup \{Y - X_{\text{new}} \rightarrow Z - X_{\text{new}}\}$; ▷ (III)
 end if
 end for each
until $(X_{\text{new}} = X_{\text{old}})$ hoặc $(|F| = 0)$;
return $X^+ = X_{\text{new}}$;
end

- i) Dùng một tập để lưu giữ các thuộc tính còn phải thêm vào bao đóng;
- ii) Dùng một mảng được đánh chỉ số bởi các thuộc tính A_i để lưu giữ các phụ thuộc hàm có vế trái chứa A_i ;
- iii) Lưu giữ số thuộc tính thuộc vế trái của mỗi phụ thuộc hàm còn chưa có mặt trong bao đóng.

Các chiến lược này đã giúp một số tác giả xây dựng được các thuật toán tuyến tính tính bao đóng, tức có độ phức tạp thời gian là $O(np)$. Đó là các thuật toán của Beerli trong [1], của Diederich trong [2] và của Paredaens trong [3].

2. Thuật toán của Mora và cộng sự

Thuật toán 2 và các thuật toán tính bao đóng của Diederich, Beerli và Paredaens đã được chạy thử nghiệm và cho kết quả được trình bày trong [4], cụ thể như sau: sinh ngẫu nhiên tập thuộc tính Ω , tập phụ thuộc hàm F và tập con $X \subseteq \Omega$; các tập phụ thuộc hàm được sinh ngẫu nhiên lần lượt có 25, 50, 75, 100, 125, 150, 175 và 200 phụ thuộc hàm, số lượng các thuộc tính ở vế trái và vế phải của các phụ thuộc hàm biến đổi từ 1 đến 300, kích thước của các tập phụ thuộc hàm từ 50 đến 61770 thuộc tính (kích thước của một tập phụ thuộc hàm F được định nghĩa là tổng các kích thước của các phụ thuộc hàm trong F ; kích thước của $X \rightarrow Y$ là $|X| + |Y|$; $|X|$ là số thuộc tính của tập X); thực hiện thử nghiệm các thuật toán 1817 lần. Kết quả thử nghiệm cho thời gian tính toán của 1817 lần thực hiện được cho trong Bảng I, trong đó thời gian trung bình tính

Bảng I
KẾT QUẢ THỬ NGHIỆM

Thuật toán	Trung bình
Thuật toán tính bao đóng X^+ của Diederich	4593,48
Thuật toán tính bao đóng X^+ của Beeri	7013,56
Thuật toán tính bao đóng X^+ của Paredaens	5863,35
Thuật toán tính bao đóng của nhóm A. Mora và cộng sự (Thuật toán 2)	1262,41

theo đơn vị giây. Từ bảng kết quả trên, ta thấy Thuật toán 2 tiêu tốn ít thời gian hơn ba thuật toán còn lại.

3. Thuật toán cải tiến tính bao đóng

Ta nhận thấy rằng, so với các thuật toán tính bao đóng đã nêu trong mục II-1, Thuật toán 2 có ưu điểm là làm đơn giản hóa tập F bằng cách loại bỏ FD $Y \rightarrow Z$ trong F sau khi đã dùng nó để tính giá trị mới của X_{new} hoặc thay thế FD $Y \rightarrow Z$ trong F bằng một FD đơn giản hơn trong trường hợp cả hai vế của FD này đều không là những tập con của X_{new} . Tuy nhiên, tính đúng đắn của Thuật toán 2 không được chứng minh một cách tường minh. Hơn nữa, nhược điểm của nó là mỗi lần duyệt tập F , tất cả các FD có vế trái và vế phải cùng chứa trong X_{new} vẫn được kiểm tra vế trái để từ đó tính giá trị mới của X_{new} (điều này làm mất thời gian không cần thiết vì giá trị X_{new} thực chất không thay đổi). Thuật toán 3 tránh được những phép kiểm tra và tính toán không cần thiết này vì thực hiện loại bỏ ngay từ đầu các FD có vế phải chứa trong X_{new} nên thời gian thực hiện nhanh hơn so với Thuật toán 2.

Với Thuật toán 3, ta có bố đề sau đây.

Bổ đề 1. *Thuật toán 3 là đúng đắn, có nghĩa nó tính đúng bao đóng X^+ của X đối với F .*

Chứng minh: Vì Thuật toán 3 là một cải tiến của Thuật toán 2 và do đó cũng là cải tiến của Thuật toán 1 nên để chứng minh tính đúng đắn của Thuật toán 3, ta chỉ cần chỉ ra rằng việc thay thế

$$Y \rightarrow Z \text{ bởi } Y - X_{\text{new}} \rightarrow Z - X_{\text{new}} \quad (1)$$

không có ảnh hưởng gì đến kết quả của việc tính bao đóng.

Thật vậy, Thuật toán 3 sẽ thay thế $Y \rightarrow Z$ bởi $Y - X_{\text{new}} \rightarrow Z - X_{\text{new}}$ trong trường hợp cả Y và Z đều không phải là tập con của X_{new} . Do đó, từ (1) ta suy ra $Y - X_{\text{new}} \neq \emptyset$ vì nếu $Y - X_{\text{new}} = \emptyset$ thì Y là tập con của X_{new} (mâu thuẫn).

Mặt khác, ta luôn có

$$(Y - X_{\text{new}}) \cup (Y \cap X_{\text{new}}) = Y. \quad (2)$$

Giả sử sau phép thay thế (1), X_{new} thay đổi và nhận giá trị mới là $X_{\text{new}1}$ với $X_{\text{new}} \subseteq X_{\text{new}1}$.

Thuật toán 3: Tính bao đóng X^+

Input:

$\Omega, F, X \subseteq \Omega$

Output: X^+

begin

$X_{\text{new}} = X;$

repeat

$X_{\text{old}} = X_{\text{new}};$

for each $(Y \rightarrow Z) \in F$ **do**

if $Z \subseteq X_{\text{new}}$ **then**

$F = F - \{Y \rightarrow Z\};$ ▷ (I)

else if $Y \subseteq X_{\text{new}}$ **then**

begin ▷ (II)

$X_{\text{new}} = X_{\text{new}} \cup Z;$

$F = F - \{Y \rightarrow Z\};$

end

else

begin ▷ (III)

$F = F - \{Y \rightarrow Z\};$

$F = F \cup \{Y - X_{\text{new}} \rightarrow Z - X_{\text{new}}\};$

end

end if

end for each

until $(X_{\text{new}} = X_{\text{old}})$ or $(|F| = 0);$

return $X_{\text{new}};$

end

Bây giờ ta phải chứng minh

$$(Y - X_{\text{new}}) \subseteq X_{\text{new}1} \iff Y \subseteq X_{\text{new}1}. \quad (3)$$

Thật vậy, từ $Y \subseteq X_{\text{new}1}$ ta có

$$(Y - X_{\text{new}}) \subseteq X_{\text{new}1}.$$

Từ $(Y - X_{\text{new}}) \subseteq X_{\text{new}1}$, $(Y \cap X_{\text{new}}) \subseteq X_{\text{new}} \subseteq X_{\text{new}1}$ và (2), ta có

$$Y = (Y - X_{\text{new}}) \cup (Y \cap X_{\text{new}}) \subseteq X_{\text{new}1}.$$

Từ hai kết quả trên, ta suy ra (3) được chứng minh. □

Ví dụ 1. Cho $F = \{d \rightarrow a, ad \rightarrow c, e \rightarrow bi, ke \rightarrow m, ce \rightarrow ik, d \rightarrow bei, h \rightarrow cde\}$. Chúng ta tính bao đóng của tập thuộc tính $X = acd$ theo Thuật toán 3. Trong Bảng II, ký hiệu (I), (II), (III) tương ứng với các đoạn mã (I), (II), (III) trong Thuật toán 3 được áp dụng; ký hiệu $\times$ là phụ thuộc hàm trong cột tương ứng bị loại bỏ khỏi F . Kết quả ta được $X^+ = acdbeikm$. So với Thuật toán 2, ta thấy có 4 vị trí (các ô có màu xám) chứng tỏ Thuật toán 3 thực hiện hiệu quả hơn. Chẳng hạn, với $(d \rightarrow a)$ hoặc $(ad \rightarrow c)$ hoặc $(e \rightarrow bi)$ thì Thuật toán 3 chỉ cần kiểm tra vế phải và loại bỏ ngay trong khi Thuật toán 2 phải kiểm tra vế trái rồi hợp vế phải vào X_{new} (nhưng thực sự X_{new} không thay đổi) sau đó mới loại bỏ. Như vậy, chỉ với 7 phụ thuộc hàm trong

Bảng II
MINH HỌA THUẬT TOÁN 3

F		$d \rightarrow a$	$ad \rightarrow c$	$e \rightarrow bi$	$ke \rightarrow m$	$ce \rightarrow ik$	$d \rightarrow bei$	$h \rightarrow cde$
X_{new}	acd	(I)	(I)			(III)	$acdbei$ (II)	(I)
F		$\times$	$\times$	$e \rightarrow bi$	$ke \rightarrow m$	$e \rightarrow ik$	$\times$	$\times$
X_{new}	$acdbei$			(I)	(III)	$acdbeik$ (II)		
F				$\times$	$k \rightarrow m$	$\times$		
X_{new}	$acdbeik$				$acdbeikm$ (II)			
F					$\times$			

ví dụ trên, Thuật toán 3 đã tiết kiệm được 4 thao tác tính toán vô ích so với Thuật toán 2.

III. MỘT SỐ KẾT QUẢ VỀ RÚT GỌN BÀI TOÁN TÌM KHÓA

Trong [5], dựa trên ngữ nghĩa quen thuộc của các phụ thuộc hàm trong mô hình cơ sở dữ liệu quan hệ và thuật toán tính bao đóng của một tập thuộc tính, các tác giả đã xây dựng được một điều kiện cần để một tập thuộc tính thuộc Ω là khóa của S . Tiếp đó, một số hướng cải tiến cho điều kiện cần thu được cũng đã được xem xét. Trong [6], dựa trên việc nghiên cứu các toán tử lý tưởng không tắt định (ideal non-deterministic operators) trong khuôn khổ của lý thuyết dàn, các tác giả của [6] cũng đưa ra một điều kiện cần để một tập thuộc tính là khóa. Như vậy, chúng ta có hai kết quả cho cùng một bài toán được công bố cách nhau 26 năm mà thoạt nhìn dường như khác nhau.

Trong phần này, chúng tôi sẽ chứng minh rằng điều kiện cần trong [6] chính là một dạng cải tiến của điều kiện cần trong [5]. Mỗi quan hệ giữa các dạng của điều kiện cần để một tập thuộc tính là khóa của một lược đồ quan hệ với việc rút gọn bài toán tìm khóa cũng được chỉ ra.

1. Nhắc lại một số kết quả đã biết

Trong mục này, một số kết quả trong [5] và [6] được nhắc lại để tiện so sánh. Lưu ý rằng thuật ngữ khóa dùng ở đây được hiểu theo nghĩa khóa tối thiểu.

Cho $S = \langle \Omega, F \rangle$ là một lược đồ quan hệ, trong đó $\Omega = \{A_1, A_2, \dots, A_n\}$ là tập hữu hạn các thuộc tính và $F = \{L_1 \rightarrow R_1, \dots, L_m \rightarrow R_m \mid L_i, R_i \subseteq \Omega, \forall i = 1, \dots, m\}$ là tập hữu hạn các phụ thuộc hàm đúng trên S .

Ký hiệu $L = \cup_{i=1}^m L_i$, $R = \cup_{i=1}^m R_i$, K_S là tập tất cả các khóa của S , $K_S = \{K_i \mid K_i \text{ là khóa của } S\}$, $G = \cap_{K_j \in K_S} K_j$ là giao của tất cả các khóa của S , $H = \cup_{K_j \in K_S} K_j$ là tập tất cả các thuộc tính khóa của S , $\overline{H} = \Omega \setminus H$ là tập tất cả các thuộc tính không khóa của S .

Trong [5] đã chứng minh các kết quả sau:

Định lý 1 ([5, Định lý 1]). Cho $S = \langle \Omega, F \rangle$ là một lược đồ quan hệ và X là một khóa của S , khi đó

$$\Omega \setminus R \subseteq X \subseteq (\Omega \setminus R) \cup (L \cap R). \quad (4)$$

Định lý 2 ([5, Định lý 4]). Cho $S = \langle \Omega, F \rangle$ là một lược đồ quan hệ, khi đó

$$G = \Omega \setminus R. \quad (5)$$

Mệnh đề 1 ([5, Trong chứng minh Định lý 1]). Ta có $R \setminus L \subseteq \overline{H}$, có nghĩa các thuộc tính trong $R \setminus L$ đều là các thuộc tính không khóa.

Từ (4), dễ dàng suy ra các nhận xét sau:

Nhận xét 1. $(\Omega \setminus R) \cup (L \cap R)$ là siêu khóa chứa tất cả các khóa của S . Lưu ý là trong phân tích $\Omega = (\Omega \setminus R) \cup (R \cap L) \cup (R \setminus L)$, chỉ tập $(L \cap R)$ có khả năng chứa cả hai loại thuộc tính là thuộc tính khóa và thuộc tính không khóa. Thêm vào đó, nếu có $(R \setminus L) \neq \emptyset$ thì siêu khóa $(\Omega \setminus R) \cup (L \cap R) \subset \Omega$ và việc tìm tập tất cả các khóa chứa trong một siêu khóa nhỏ hơn thực sự Ω sẽ ít tốn kém hơn. Điều này rõ ràng liên quan đến việc rút gọn bài toán tìm khóa của một lược đồ quan hệ. Thật vậy, giả sử đã xác định được $Z \subset \Omega$ là tập chứa tất cả các khóa của lược đồ quan hệ $S = \langle \Omega, F \rangle$. Khi đó việc rút gọn bài toán cho việc tìm khóa của S được tiến hành qua các bước sau:

- 1) Xác định lược đồ quan hệ $S' = \langle \Omega', F' \rangle$ trong đó $\Omega' = Z \setminus (\Omega \setminus R)$ và $F' = \{L_i \cap \Omega' \rightarrow R_i \cap \Omega' \mid (L_i \rightarrow R_i) \in F, i = 1, 2, \dots, m\}$;
- 2) Tìm $K_{S'}$ theo một thuật toán nào đó;
- 3) Dễ thấy rằng $K_S = \{(\Omega \setminus R) \cup K \mid K \in K_{S'}\}$.

Nhận xét 2. Các khóa $K_j \in K_S$ không chứa nhau và có cấu trúc chung là $K_j = (\Omega \setminus R) \cup Z_j$ với $Z_j \subseteq L \cap R$. Điều này tạo thuận lợi cho việc xác định các khóa của S .

Nhận xét 3. Trường hợp tồn tại tập $Z \subseteq \overline{H}$ sao cho $(L \cap R) \cap Z \neq \emptyset$ thì $(\Omega \setminus R) \cup [(L \cap R) \setminus Z]$ sẽ là một siêu khóa chứa tất cả các khóa của S và siêu khóa này rõ ràng chứa thực sự trong siêu khóa $(\Omega \setminus R) \cup (L \cap R)$.

Khi đó

$$(\Omega \setminus R) \subseteq K_j \subseteq (\Omega \setminus R) \cup [(L \cap R) \setminus Z], \quad \forall K_j \in K_S,$$

sẽ là một dạng cải tiến của điều kiện cần (4).

Trong [6, 7], có đưa ra định nghĩa và định lý sau (các ký hiệu được sửa lại cho phù hợp với hệ thống ký hiệu đã dùng ở trên).

Định nghĩa 1 ([7, Định nghĩa 3.3]). Cho $S = \langle \Omega, F \rangle$ là một lược đồ quan hệ, khi đó lõi (core) và thân (body) của S được định nghĩa như sau:

$$\begin{aligned} \text{core}(\Omega, F) &= \Omega \setminus \left(\bigcup_{(L_i \rightarrow R_i) \in F} R_i \right), \\ \text{body}(\Omega, F) &= \left(\bigcup_{(L_i \rightarrow R_i) \in F} L_i \right) \cap [\Omega \setminus \text{core}(\Omega, F)^+]. \end{aligned}$$

Bằng những tính toán đơn giản, ta nhận được

$$\text{core}(\Omega, F) = \Omega \setminus R, \quad (6)$$

$$\text{body}(\Omega, F) = L \cap [\Omega \setminus (\Omega \setminus R)^+]. \quad (7)$$

Ví dụ 1 ([7, Ví dụ 3.1]) Cho $S = \langle \Omega, F \rangle$ là một lược đồ quan hệ, trong đó tập thuộc tính $\Omega = \{a, b, c, d, e, f, g, h\}$ và tập phụ thuộc hàm $F = \{ab \rightarrow c, a \rightarrow g, g \rightarrow c, b \rightarrow h, bh \rightarrow d, c \rightarrow d, e \rightarrow f, f \rightarrow e\}$.

Ta có $L = abce fgh$, $R = cde fgh$, $\Omega \setminus R = ab$, $(\Omega \setminus R)^+ = abcdgh$, $L \cap [\Omega \setminus (\Omega \setminus R)^+] = ef$. Từ đó, $\text{core}(\Omega, F) = ab$ và $\text{body}(\Omega, F) = ef$.

Định lý 3 ([6, 7, Định lý 3.4]). Cho $S = \langle \Omega, F \rangle$ là một lược đồ quan hệ và K là một khóa (tối thiểu) của S , khi đó, ta có $\text{core} \subseteq K \subseteq (\text{core} \cup \text{body})$, có nghĩa là

$$\Omega \setminus R \subseteq K \subseteq (\Omega \setminus R) \cup [L \cap [\Omega \setminus (\Omega \setminus R)^+]]. \quad (8)$$

Rõ ràng (8) là phát biểu của một điều kiện cần để K là khóa của S . Chứng minh của (8) được cho trong [6] cùng với một số ví dụ minh họa.

2. Một dạng cải tiến cho điều kiện cần (4)

Trong [5] có chứng minh bổ đề sau:

Bổ đề 2 ([5, Bổ đề 3]). Cho $S = \langle \Omega, F \rangle$ là một lược đồ quan hệ và X là một khóa của S , khi đó

$$X \cap R \cap (L \setminus R)^+ = \emptyset.$$

Bổ đề 2 dễ dàng được mở rộng thành Bổ đề 3 sau đây.

Bổ đề 3. Cho $S = \langle \Omega, F \rangle$ là một lược đồ quan hệ, khi đó

$$K \cap R \cap (\Omega \setminus R)^+ = \emptyset \quad \forall K \in K_S \Rightarrow R \cap (\Omega \setminus R)^+ \subseteq \overline{H}.$$

Chứng minh: Giả sử điều ngược lại, tức là tồn tại $K \in K_S$ sao cho $K \cap R \cap (\Omega \setminus R)^+ \neq \emptyset$, có nghĩa là tồn tại $A \in \Omega$ sao cho $A \in K$, $A \in R$ và theo định nghĩa bao đóng, $\Omega \setminus R \rightarrow A$. Vì $A \in R$ nên $A \notin \Omega \setminus R$. Từ điều kiện (4), có $(\Omega \setminus R) \subseteq K$. Kết hợp với $A \notin \Omega \setminus R$, suy ra $\Omega \setminus R \subseteq K \setminus A$. Từ đó, $K \setminus A \rightarrow \Omega \setminus R$. Mặt khác, $\Omega \setminus R \rightarrow A$. Kết quả là, $K \setminus A \rightarrow A$ với $A \in K$, chứng tỏ K không là khóa của S . Vậy, $K \cap R \cap (\Omega \setminus R)^+ = \emptyset$, có nghĩa là $R \cap (\Omega \setminus R)^+ \subseteq \overline{H}$. $\square$

Từ Nhận xét 3, định lý sau đây là hiển nhiên.

Định lý 4. Cho $S = \langle \Omega, F \rangle$ là một lược đồ quan hệ, khi đó

$$\Omega \setminus R \subseteq K \subseteq (\Omega \setminus R) \cup [(L \cap R) \setminus (R \cap (\Omega \setminus R)^+)], \quad (9)$$

với mọi $K \in K_S$.

Ta xem (9) như một dạng cải tiến của (4). Sau đây là một ví dụ trong đó $(L \cap R) \cap (R \cap (\Omega \setminus R)^+) \neq \emptyset$ có nghĩa là

$$(\Omega \setminus R) \cup [(L \cap R) \setminus (R \cap (\Omega \setminus R)^+)] \subset (\Omega \setminus R) \cup (L \cap R).$$

Ví dụ 2. Xét lược đồ quan hệ $S = \langle \Omega, F \rangle$, trong đó $\Omega = \{a, b, c, d, e, f, g, h, i\}$ và $F = \{a \rightarrow b, b \rightarrow c, d \rightarrow e, h \rightarrow i, i \rightarrow h\}$.

Với lược đồ quan hệ này, ta có: $L = abdhi$, $R = bcehi$, $L \cap R = bhi$, $\Omega \setminus R = adfg$; $(\Omega \setminus R)^+ = abcdefg$, $R \cap (\Omega \setminus R)^+ = bce$. Để thấy rằng $K_S = \{ad fgh, ad fgi\}$. Từ đó $H = \{a, d, f, g, h, i\}$ và $\overline{H} = \{b, c, e\}$. Bổ đề 3 được nghiệm đúng vì $R \cap (\Omega \setminus R)^+ = bce \subseteq \overline{H}$.

Hơn nữa, ta còn có $(L \cap R) \cap (R \cap (\Omega \setminus R)^+) = b \neq \emptyset$. Và như vậy, với lược đồ quan hệ S được cho trong Ví dụ 2, ta có

$$\Omega \setminus R \subseteq K \subseteq (\Omega \setminus R) \cup [(L \cap R) \setminus (R \cap (\Omega \setminus R)^+)],$$

với mọi $K \in K_S$. Cụ thể là, $ad f g \subseteq K \subseteq ad f ghi$, với $K \in \{ad f gh, ad f gi\}$.

3. So sánh hai điều kiện cần (8) và (9)

Nhận xét 4. Trong (8) dễ thấy rằng

$$L \cap [\Omega \setminus (\Omega \setminus R)^+] = L \setminus (\Omega \setminus R)^+.$$

Thật vậy, giả sử $x \in L \cap [\Omega \setminus (\Omega \setminus R)^+]$. Lúc đó, $x \in L$, $x \in \Omega \setminus (\Omega \setminus R)^+$, suy ra $x \in L$, $x \notin (\Omega \setminus R)^+$, vì thế $x \in L \setminus (\Omega \setminus R)^+$. Ngược lại, giả sử $x \in L \setminus (\Omega \setminus R)^+$. Khi đó, $x \in L$, $x \notin (\Omega \setminus R)^+$, suy ra $x \in L$, $x \in \Omega \setminus (\Omega \setminus R)^+$, vì vậy $x \in L \cap [\Omega \setminus (\Omega \setminus R)^+]$.

Do Nhận xét 4, để đơn giản, ta vẫn đánh số bao hàm thức kép

$$\Omega \setminus R \subseteq K \subseteq (\Omega \setminus R) \cup [L \setminus (\Omega \setminus R)^+], \quad \forall K \in K_S, \quad (8)$$

là (8).

Định nghĩa 2. Ta nói rằng điều kiện (8) tốt hơn điều kiện (9) nếu

$$L \setminus (\Omega \setminus R)^+ \subseteq (L \cap R) \setminus (R \cap (\Omega \setminus R)^+)$$

và tồn tại một lược đồ quan hệ sao cho

$$L \setminus (\Omega \setminus R)^+ \subset (L \cap R) \setminus (R \cap (\Omega \setminus R)^+).$$

Hiểu theo nghĩa đó, ta thấy điều kiện (9) là một dạng cải tiến của (4). Tương tự, ta có định nghĩa khi nào thì (9) tốt hơn (8).

Để so sánh (8) với (9) ta có định lý sau.

Định lý 5. Hai điều kiện (8) và (9) chỉ là một và được diễn đạt bằng những biểu thức khác nhau.

Chứng minh: Để chứng minh Định lý 5, rõ ràng chỉ cần chứng minh

$$L \setminus (\Omega \setminus R)^+ = (L \cap R) \setminus (R \cap (\Omega \setminus R)^+). \quad (10)$$

Giả sử x là một thuộc tính bất kỳ thuộc $L \setminus (\Omega \setminus R)^+$. Khi đó, với mọi $x \in L \setminus (\Omega \setminus R)^+$, ta có $x \in L$ và $x \notin (\Omega \setminus R)^+$. Từ đó, $x \in L$, $x \notin (\Omega \setminus R)$ và $x \notin (\Omega \setminus R)^+$, suy ra $x \in L$, $x \in R$ và $x \notin (\Omega \setminus R)^+$, suy ra $x \in (L \cap R)$ và $x \notin [R \cap (\Omega \setminus R)^+]$, suy ra $x \in (L \cap R) \setminus (R \cap (\Omega \setminus R)^+)$, có nghĩa là

$$L \setminus (\Omega \setminus R)^+ \subseteq (L \cap R) \setminus (R \cap (\Omega \setminus R)^+). \quad (11)$$

Bây giờ ta chứng minh điều ngược lại. Với mọi $x \in (L \cap R) \setminus (R \cap (\Omega \setminus R)^+)$, ta có $x \in L$, $x \in R$ và $x \notin [R \cap (\Omega \setminus R)^+]$, suy ra $x \in L$, $x \in R$ và $x \notin (\Omega \setminus R)^+$, suy ra $x \in L \setminus (\Omega \setminus R)^+$, có nghĩa là

$$(L \cap R) \setminus (R \cap (\Omega \setminus R)^+) \subseteq L \setminus (\Omega \setminus R)^+. \quad (12)$$

Kết hợp (11) và (12), ta có

$$L \setminus (\Omega \setminus R)^+ = (L \cap R) \setminus (R \cap (\Omega \setminus R)^+).$$

Định lý được chứng minh. $\square$

Để minh họa cho Định lý 5, ta trở lại với Ví dụ 1 và Ví dụ 2. Với Ví dụ 1, $\Omega = \{a, b, c, d, e, f, g, h\}$, $F = \{ab \rightarrow c, a \rightarrow g, g \rightarrow c, b \rightarrow h, bh \rightarrow d, c \rightarrow d, e \rightarrow f, f \rightarrow e\}$. Ta có, $L = abcefg, R = cdefgh, L \cap R = cefgh, \Omega \setminus R = ab, (\Omega \setminus R)^+ = abcdgh, R \cap (\Omega \setminus R)^+ = cdgh$. Từ đó $L \setminus (\Omega \setminus R)^+ = ef$ và $(L \cap R) \setminus (R \cap (\Omega \setminus R)^+) = ef$.

Với Ví dụ 2, $\Omega = \{a, b, c, d, e, f, g, h, i\}$, $F = \{a \rightarrow b, b \rightarrow c, d \rightarrow e, h \rightarrow i, i \rightarrow h\}$. Ta có, $L = abdhi, R = bcehi, L \cap R = bhi, \Omega \setminus R = adfg, (\Omega \setminus R)^+ = abcdefg, R \cap (\Omega \setminus R)^+ = bce$. Từ đó $L \setminus (\Omega \setminus R)^+ = hi$ và $(L \cap R) \setminus (R \cap (\Omega \setminus R)^+) = hi$.

Liên quan tới các điều kiện cần để một tập thuộc tính $K \subseteq \Omega$ là khóa của lược đồ quan hệ $S = \langle \Omega, F \rangle$, ta có thể xem xét và giải quyết bài toán sau.

4. Một bài toán quyết định

Cho $S = \langle \Omega, F \rangle$ là một lược đồ quan hệ và cho $Z \subset \Omega$. Bài toán đặt ra là quyết định xem Z có phải là tập chứa tất cả các khóa của S không.

Giả sử Z chứa tất cả các khóa của S . Điều đó có nghĩa là

$$Z \supseteq H = \bigcup_{K_j \in K_S} K_j.$$

Từ đó $\Omega \setminus Z \subseteq \Omega \setminus H = \overline{H}$.

Bổ đề 4. Cho $S = \langle \Omega, F \rangle$ là một lược đồ quan hệ và cho $Z \subset \Omega$. Khi đó Z chứa tất cả các khóa của S khi và chỉ khi $\Omega \setminus Z$ chỉ gồm các thuộc tính không khóa, có nghĩa là

$$\Omega \setminus Z \subseteq \overline{H}.$$

Để thấy được ý nghĩa của Bổ đề 4, ta trở lại với điều kiện (8), là Định lý 3.4 trong [7]. Rõ ràng, điều kiện này khẳng định rằng

$$Z = (\Omega \setminus R) \cup [L \setminus (\Omega \setminus R)^+] \quad (13)$$

là tập (siêu khóa) chứa tất cả các khóa của S .

Để kiểm tra tính chất trên, ta có thể dùng Bổ đề 4. Ta có¹

$$\begin{aligned} \Omega \setminus Z &= \Omega \setminus [(\Omega \setminus R) \cup (L \setminus (\Omega \setminus R)^+)] \\ &= R \setminus (L \setminus (\Omega \setminus R)^+) \\ &= (R \setminus L) \cup (R \cap (\Omega \setminus R)^+). \end{aligned}$$

Như vậy

$$\begin{aligned} \Omega \setminus Z &= R \setminus (L \setminus (\Omega \setminus R)^+) \\ &= (R \setminus L) \cup (R \cap (\Omega \setminus R)^+) \subseteq \overline{H}, \end{aligned}$$

do đã có $(R \setminus L) \subseteq \overline{H}$ (theo [5]) và $R \cap (\Omega \setminus R)^+ \subseteq \overline{H}$ (theo Bổ đề 3). Điều này chứng tỏ rằng $Z = (\Omega \setminus R) \cup [L \setminus (\Omega \setminus R)^+]$ là siêu khóa chứa tất cả các khóa của S .

IV. KẾT LUẬN

Trong bài báo này, chúng tôi đã đề xuất một thuật toán cải tiến (Thuật toán 3) tính bao đóng của một tập thuộc tính đối với một tập phụ thuộc hàm. Vì tất cả các FD có vẻ phải chứa trong X_{new} đều bị loại bỏ trước khi thực sự tiến hành tính bao đóng nên Thuật toán 3 rõ ràng là hiệu quả hơn Thuật toán 2. Đặc biệt là trong trường hợp tập F ban đầu gồm một số phụ thuộc hàm có vẻ phải chứa trong X_{new} hoặc trong quá trình tính giá trị mới của X_{new} , việc thay thế một phụ thuộc hàm bằng một phụ thuộc hàm đơn giản hơn, làm xuất hiện một phụ thuộc hàm mới có vẻ phải chứa trong X_{new} . Hơn nữa, tính đúng đắn của Thuật toán 2 không được chứng minh tường minh khi thực hiện phép thay thế các phụ thuộc hàm bằng các phụ thuộc hàm đơn

¹ Ở đây, với ba tập bất kỳ $A, B, C \subseteq \Omega$, ta đã áp dụng biến đổi quen thuộc $A \setminus (B \setminus C) = (A \setminus B) \cup (A \cap C)$.

giản hơn. Với bổ đề 0, chúng tôi đã chứng minh tính đúng đắn này trong Thuật toán 3. Với việc rút gọn bài toán tìm khóa, chúng tôi cũng đã chứng minh được điều kiện cần (8) trùng với điều kiện cần (9) là một dạng cải tiến của điều kiện cần (4). Đây là những điều kiện cần để một tập con của Ω là khóa tối thiểu của lược đồ quan hệ $S = \langle \Omega, F \rangle$. Việc tìm một điều kiện cần tốt hơn (8) hoặc (9) nhằm rút gọn hơn nữa bài toán tìm khóa là một vấn đề đáng quan tâm.

TÀI LIỆU THAM KHẢO

- [1] C. Beeri and P. A. Bernstein, “Computational problems related to the design of normal form relational schemas,” *ACM Transactions on Database Systems (TODS)*, vol. 4, no. 1, pp. 30–59, 1979.
- [2] J. Diederich and J. Milton, “New methods and fast algorithms for database normalization,” *ACM Transactions on Database Systems (TODS)*, vol. 13, no. 3, pp. 339–365, 1988.
- [3] J. Paredaens, P. De Bra, M. Gyssens, and D. Van Gucht, “The Structure of the Relational Database Model,” *EATCS Monographs on Theoretical Computer Science*, vol. 17, 1989.
- [4] A. Mora, G. Aguilera, M. Enciso, P. Cordero, and I. P. de Guzmán, “A new closure algorithm based in logic: SLFD-Closure versus classical closures,” *Inteligencia Artificial. Revista Iberoamericana de Inteligencia Artificial*, vol. 10, no. 31, pp. 31–40, 2006.
- [5] H. Thuan and L. V. Bao, “Some results about key of relational schemas,” *Acta Cybernetica*, vol. 7, no. 1, pp. 99–113, 1985.
- [6] A. Mora, I. P. de Guzmán, M. Enciso, and P. Cordero, “Ideal non-deterministic operators as a formal framework to reduce the key finding problem,” *International Journal of Computer Mathematics*, vol. 88, no. 9, pp. 1860–1868, 2011.
- [7] P. Cordero, M. Enciso, and A. Mora, “Automated reasoning to infer all minimal keys,” in *Proceedings of the Twenty-Third International Joint Conference on Artificial Intelligence (IJCAI)*, 2013, pp. 817–823.



Vũ Quốc Tuấn sinh năm 1982 tại Hải Dương. Ông tốt nghiệp Trường Đại học Khoa học Tự nhiên, Đại học Quốc gia Hà Nội, ngành Toán-Tin ứng dụng, năm 2005. Năm 2010, ông nhận bằng Thạc sĩ Công nghệ Thông tin tại Trường Đại học Sư phạm Hà Nội. Hiện nay, ông đang công tác tại Trường Cao đẳng Hải Dương và là nghiên cứu sinh tại Học viện Khoa học và Công nghệ, Viện Hàn lâm Khoa học và Công nghệ Việt Nam. Lĩnh vực nghiên cứu của ông bao gồm khai phá dữ liệu, các hệ cơ sở dữ liệu và cơ sở tri thức.



Hồ Thuần sinh năm 1933 tại Bắc Ninh, nguyên là Trưởng phòng Lập trình, Viện Khoa học tính toán và Điều khiển, Viện Hàn lâm Khoa học và Công nghệ Việt Nam. Ông tốt nghiệp Trường Đại học Tổng hợp năm 1960, bảo vệ Tiến sĩ ngành Toán học tính toán năm 1979 và được phong Phó Giáo sư năm 1984. Hiện nay, ông là cán bộ nghỉ hưu. Lĩnh vực nghiên cứu của ông gồm các hệ cơ sở dữ liệu và cơ sở tri thức, phân tích và thiết kế thuật toán, lý thuyết mật mã và an toàn dữ liệu.