Tạp chí Tin học và Điều khiển học, T. 15, S. 3 (1999), 46-49

MỘT SỐ CẢI TIẾN TRONG QUÁ TRÌNH XÂY DỰNG PHỦ TỐI TIỂU

NGUYỄN XUÂN HÀO

Abstract. The problem of finding the minimal cover plays an important role in the design of a relational database. In this paper, some ameliorations are proposed for reducing the complexity of the algorithm for finding a minimal cover.

1. ĐẶT VẤN ĐỀ

Giả sử $S = \langle R, S\rangle$, R là một lược đồ quan hệ có tập thuộc tính là $\Omega = \{A_1, A_2, \ldots, A_n\}$, $F = \{f_i : i = 1, \ldots, m\}$ là tập các phụ thuộc hàm. Ta giả sử $f_i = X_i \rightarrow Y_i$ với $X_i, Y_i \subseteq \Omega$. Khi cần chỉ rõ f, ta kí hiệu $f = X \rightarrow Y$ là $f.X \rightarrow f.Y$.

Kí hiệu F^+ là tập tất cả các phụ thuộc hàm được suy diễn từ F theo các tiên đề của Armstrong.

Để đơn giản, ta kí hiệu $X \cup Y$ là XY, với X, Y là các tập hợp, kể cả khi chúng chỉ gồm 1 phần tử.

Liên quan đến các phụ thuộc hàm là các bài toán tìm bao đóng, bài toán thành viên, bài toán tìm phủ tối thiểu, bài toán tìm khóa và tập các khóa. Bài toán tìm bao đóng có tải trọng lớn nhất, vì các bài toán sau đó đều ứng dụng kết quả của nó.

Sau đây là những kết quả đã biết sẽ được dùng trong bài báo này.

a. Bao đóng

Định nghĩa. Giả sử $X \subseteq \Omega$. Bao đóng của X đối với tập các phụ thuộc hàm F, kí hiệu là X_F^+ : $X_F^+ = \{A \in \Omega \mid (X \rightarrow A) \in F^+\}$. Khi không cần chỉ rõ F, có thể viết là X^+.

Beeri và Bernstein đã nêu ra thuật toán với thời gian tuyến tính với việc sử dụng các cấu trúc dữ liệu thích hợp để tìm X^+. Thuật toán dưới đây tuy có độ phức tạp cao hơn nhưng khá trực quan khi tính bao đóng.

1) Xây dựng dãy $X^{(0)}, X^{(1)}, \ldots$ như sau:

$$X^{(0)} = X,$$
$$X^{(i+1)} = X^{(i)} \bigcup \Big(\bigcup_{\substack{f \in F \\ f.X \subseteq X^{(i)}}} f.Y \Big),$$

2) Từ kết quả trên ta được dãy:

$$X^{(0)} \subseteq X^{(1)} \subseteq \cdots \subseteq X^{(i)} \subseteq \cdots \subseteq \Omega.$$

Do Ω là tập hữu hạn nên tồn tại một giá trị t nguyên không âm nhỏ nhất sao cho: $X^{(t)} = X^{(t+1)}$.

3) Kết luận: $X^+ = X^{(t)}$.

Mệnh đề 1. *Nếu $X \subseteq Y$ thì $X^+ \subseteq Y^+$.*

b. Bài toán thành viên

Việc xác định phụ thuộc hàm g có thuộc F^+ hay không được gọi là bài toán thành viên.

Định lý sau đã giải quyết trọn vẹn bài toán thành viên [4].

Định lý. *Giả sử $\langle R, F\rangle$ là một lược đồ quan hệ, $g = X \rightarrow Y$ là một phụ thuộc hàm. $g \in F^+$ khi và chỉ khi $Y \subseteq X^+$.*

2. PHỦ TỐI TIỂU

Định nghĩa. Giả sử G là tập các phụ thuộc hàm trên lược đồ $\langle R, F\rangle$. G được gọi là phủ của F nếu $G^+ = F^+$.

Định nghĩa. G được gọi là phủ tối tiểu của F nếu G là phủ của F và:

i. $\forall g \in G$, $g = X \to Y$ thì Y chỉ có 1 phần tử.

ii. $\forall g \in G$ $\quad \{G \setminus \{g\}\}^+ \neq F^+$.

iii. $\forall g \in G$, $\forall g'$ thỏa mãn $g'.X \subseteq g.X$, $g'.Y = g.Y$, ta luôn có:

$$\{(G \setminus \{g\}) \cup \{g'\}\}^+ \neq F^+.$$

Các tính chất đó được gọi là: (i) - vế phải tối giản; (ii) - tính không thừa; và (iii) - tính phụ thuộc đầy đủ.

Ý nghĩa của phủ tối tiểu có thể thấy qua định nghĩa. Đó là nếu thay thế tập phụ thuộc F ban đầu bằng phủ tối tiểu thì sẽ tăng hiệu quả tìm bao đóng:

- Giảm số lần duyệt các phụ thuộc hàm của tập F do F không dư thừa.
- Giảm khối lượng tính toán khi xây dựng dãy các tập $X^{(i)}$ vì phép kiểm tra vế trái của mỗi phụ thuộc hàm (có ít thuộc tính nhất) đó là tập con của $X^{(i)}$ hay không sẽ là hiệu quả nhất.

2.1. Tính chất của phủ tối tiểu

Mệnh đề 2. *Cho hai tập các phụ thuộc hàm F và $G, F \supseteq G$. Khi đó với mọi X, biểu thức sau luôn được thỏa mãn: $X_F^+ \supseteq X_G^+$.*

Chứng minh. Giả sử $x \in X_G^+$. Ta phải chứng minh $x \in X_F^+$.

Thật vậy, nếu $x \in X_G^+$ và $x \in X$ thì ta có ngay $x \in X_F^+$. Còn nếu $x \in X_G^+$ và $x \notin X$ thì theo thuật toán tìm bao đóng X_G^+, có một dãy $X^{(0)}, X^{(1)}, ..., X^{(t)}$ sao cho $X^{(t)} = X^{(t+1)}$. Giả sử rằng có được dãy đó là do áp dụng các phụ thuộc hàm $g_1, g_2, \dots, g_t$.

Do giả thiết $F \supseteq G$, nên $g_i \in F$, suy ra quá trình tìm bao đóng X_G^+ là một "quá trình con" khi tìm bao đóng X_F^+. Vậy $x \in X_F^+$. □

Mệnh đề 3. *Cho 2 tập các phụ thuộc hàm F và $G, F \supseteq G$. Khi đó ta luôn có $F^+ \supseteq G^+$.*

Chứng minh. Thật vậy, giả sử $g = X \to Y \in G^+$, ta chứng minh $g \in F^+$. Điều đó tương đương với: nếu $Y \subseteq X_G^+$ thì $Y \subseteq X_F^+$. Theo Mệnh đề 1, vì $G \subseteq F$ nên $X_G^+ \subseteq X_F^+$, suy ra điều cần chứng minh. □

Mệnh đề 4. *Với mọi tập các phụ thuộc hàm F, ta luôn có $(F^+)^+ = F^+$.*

Chứng minh. Dễ thấy $F^+ \supseteq F$. Áp dụng Mệnh đề 1 ta có $(F^+)^+ \supseteq F^+$.

Ngược lại, ta chứng minh $(F^+)^+ \subseteq F^+$. Thật vậy nếu $f \in (F^+)^+$, giả sử f được suy dẫn từ các $f_1^+, f_2^+, ..., f_k^+$, mà $f_i^+ \in F^+$, nên f_i^+ lại được suy dẫn từ $f_{i_1}, f_{i_2}, \dots, f_{i_k}$, với $f_{i_j} \in F$, nên ghép các suy dẫn đó, ta thấy f được suy dẫn từ F, tức là $f \in F^+$. □

2.2. Xây dựng phủ tối tiểu

Theo định nghĩa, từ một tập phụ thuộc hàm bất kỳ, ta luôn luôn biến đổi để nó thỏa mãn điều kiện (i): Nếu $f = X \to Y$ có $Y = \{A_{i1}, A_{i2}, ..., A_{ik}\}$ $(k > 1)$ thì ta thay f bằng các phụ thuộc hàm $f_i = X \to A_{ij}$ $(j = 1, ..., k)$.

Sau đây là những kết quả để phủ thỏa mãn các điều kiện (ii) và (iii).

Mệnh đề 5. *Giả sử F là tập các phụ thuộc hàm, $f \in F$. Đặt $G = F \setminus \{f\}$. Nếu $f \in G^+$ thì $F^+ = G^+$.*

Chứng minh. Dễ thấy $G \subseteq F$. Theo Mệnh đề 3 thì $G^+ \subseteq F^+$.

Ngược lại, do $G \subseteq G^+$ và theo giả thiết $G = F \setminus \{f\}$ nên $F \setminus \{f\} \subseteq G^+$. Thêm phụ thuộc hàm f vào cả 2 vế, chú ý có giả thiết $f \in G^+$ nên ta có $F \subseteq G^+$. Lấy bao đóng cả 2 vế, ta có $F^+ \subseteq (G^+)^+ = G^+$. Và như vậy, nếu $f \in G^+$ thì $F^+ = G^+$. □

Mệnh đề 5 cho ta thuật toán loại các phụ thuộc dư thừa có trong F: $\forall f \in F$, nếu $f \in (F \setminus \{f\})^+$ thì tập $(F \setminus \{f\})$ là phủ của F.

Quá trình duyệt mọi f và loại bỏ nó khỏi F dẫn đến tập F là tập các phụ thuộc hàm không dư thừa. Vậy ta có thuật toán sau:

Thuật toán 1:

For each f in F
    $F := F \setminus \{f\}$
    if not $(f \in F^+)$ then $F := F \cup \{f\}$.

Mệnh đề 6. *Giả sử F là tập các phụ thuộc hàm, $f \in F$, $f = XA \to Y$ ($X \subseteq \Omega$, $A \in \Omega$). Đặt $g = X \to Y$ (như vậy, phụ thuộc hàm g có được từ phụ thuộc hàm f bằng cách bớt đi một thuộc tính trong vế trái của f). Đặt $G = (F \setminus \{f\}) \cup \{g\}$. Ta có các kết luận sau:*

a. $f \in G^+$.

b. $F^+ \subseteq G^+$.

c. Nếu $g \in F^+$ *thì* $F^+ = G^+$.

Chứng minh.

a. Do $g = (X \to Y) \in G$ nên $f = (XA \to Y) \in G^+$.

b. Từ giả thiết, ta có $(F \setminus \{f\}) \subseteq G$. Và từ kết luận a, ta có $f \in G^+$. Suy ra mọi phụ thuộc hàm thuộc F đều thuộc G^+. Vậy $F^+ \subseteq G^+$.

c. Từ giả thiết $g \in F^+$, ta thấy mọi phụ thuộc hàm thuộc G đều thuộc F^+ nên ta có $G^+ \subseteq F^+$. Kết hợp với kết luận b, ta có $G^+ = F^+$. □

Mệnh đề 6 cho ta thuật toán xây dựng tập các phụ thuộc hàm đầy đủ:

$$\forall \in F,\ f = X \to Y,\ \forall A \in X,$$
$$g := (X \setminus \{A\}) \to Y.$$

Nếu $g \in F^+$ thì $(F \setminus \{f\} \cup \{g\}$ là phủ của F.

Thuật toán 2:

For each f in F
    For each A in $f.X$
        $g.X := f.X \setminus \{A\}$
        $g.Y := f.Y$
        If $g \in F^+$ then $F := (F \setminus \{f\}) \cup \{g\}$.

Tuy nhiên, độ phức tạp của thuật toán trên còn lớn do việc duyệt mọi thuộc tính có trong vế trái của từng phụ thuộc hàm. Nội dung sau đây cho điều kiện cần để $g \in F^+$, từ đó giảm được số phép duyệt các thuộc tính có trong vế trái của f.

3. GIẢM QUÁ TRÌNH DUYỆT TRONG THUẬT TOÁN 2

Mệnh đề 7. *Giả sử F là tập các phụ thuộc hàm, $f \in F$, $f = XA \to Y$ ($X \subseteq \Omega$, $A \in \Omega$, $A \notin X$). Đặt $g = X \to Y$, $G = (F \setminus \{f\}) \cup \{g\}$. Nếu F không dư thừa và $g \in F^+$ thì $A \in \bigcup_{f \in F} (f.Y)$.*

Chứng minh. Kí hiệu $R = \bigcup_{f \in F} (f.Y)$. Từ giả thiết $g \in F^+$, ta có $Y \subseteq X_F^+$. Xét quá trình tìm bao đóng X_F^+, ta có dãy $X = X^{(0)}, X^{(1)}, \ldots, X^{(i)}, \ldots$ Vì $Y \subseteq X_F^+$ nên $\exists j \geq 0$, nhỏ nhất sao cho $Y \subseteq X^{(j)}$.

+ Ta chứng minh $j \neq 0$. Thật vậy, nếu $j = 0$ thì chứng tỏ $Y \subseteq X$, hay $Y \subseteq XA$, tức F là dư thừa, trái với giả thiết.

+ Với $j > 0$, do j nhỏ nhất nên ta có $Y \subseteq X^{(j)}$ và $Y \not\subseteq X^{(j-1)}$. Ta chứng minh $A \in X^{(j-1)}$. Giả sử trái lại, $A \notin X^{(j-1)}$. Như vậy, quá trình tính đến $X^{(j)}$ không có sự tham gia của f vì vế trái của f là XA, không phải là tập con của $X^{(k)}$ với $k = 1, ..., j$. Do đó nếu loại f ra khỏi F, tính $X^{(+)}_{F\setminus\{f\}}$, ta vẫn có $Y \subseteq X^{(+)}_{F\setminus\{f\}}$. Theo Mệnh đề 1, ta có $Y \subseteq (XA)^{(+)}_{F\setminus\{f\}}$, hay $f = (XA \to Y) \in (F \setminus \{f\})^+$, trái với giả thiết F không dư thừa.

+ Do $A \notin X$, $A \in X^{(j-1)}$ nên A đã được bổ sung trong quá trình tìm bao đóng. Vậy A là một thuộc tính trong vế phải của một phụ thuộc hàm nào đó, hay $A \in R$. □

Mệnh đề 7 cho ta điều kiện cần để $g \in F^+$ là $A \in R$. Kết hợp với Mệnh đề 6, ta có thuật toán cải tiến sau:

Thuật toán 3:

For each f in F
  For each A in $(f.X \cap R)$
    $g.X := f.X \setminus \{A\}$
    $g.Y := f.Y$
    If $g \in F^+$ then $F := (F \setminus \{f\}) \cup \{g\}$,

trong đó $R = \bigcup_{f \in F} (f.Y)$.

4. KẾT LUẬN

Để xây dựng phủ tối tiểu của tập các phụ thuộc hàm cho trước, chúng tôi đã đưa ra hai thuật toán 1 và 2. Tuy rằng việc chỉ ra một cách tường minh độ phức tạp của Thuật toán 3 thấp hơn Thuật toán 2 là khó khăn, song về trực quan ta thấy Thuật toán 3 đã giảm quá trình duyệt các phần tử có trong vế trái của f. Vậy nên dùng Thuật toán 1 và Thuật toán 3 để tìm phủ tối tiểu.

Cũng cần lưu ý rằng phải thực hiện Thuật toán 1 trước để đảm bảo tập các phụ thuộc hàm F thỏa mãn điều kiện của Mệnh đề 7 là không dư thừa.

Cuối cùng, tác giả xin cảm ơn PGS. Hồ Thuần đã đọc và giúp hoàn thiện bài báo này.

TÀI LIỆU THAM KHẢO

[1] C. Beeri and P. A. Bernstein, *Computational Problem Related to the Design of Normal Form Relation Schemes*, Trans. on DBSys. 4:1, March 1979.

[2] E. F. Codd, A relational model of data for large shared datas banks, *Comm. ACM* **13** (5) (1970).

[3] Ho Thuan, Some remarks on the algorithm of Lucchesi and Osborn, *Acta Cybernetica, Hungary* **8** (2) (1987).

[4] Ho Thuan and Le Van Bao, Some results about keys of relation schemes, *Acta Cybernetica, Hungary* **7** (1) (1985) 99-113.

[5] Lucchesi C. L. and Osborn S. L., Candidate keys for relations, *J. of Computer and System Sciences* **17** (1978) 270-279.

Nhận bài ngày 9 - 9 - 1998
Nhận sau khi sửa ngày 3 - 6 - 1999

Đại học Nông Lâm, Đại học Huế.