TẠP CHÍ KHOA HỌC TÍNH TOÁN VÀ ĐIỀU KHIỂN Tập II (1986), Số 2 (23–29)

MỘT SỐ KẾT QUẢ VỀ CẤU TRÚC CỦA PHỦ TỐI THIỂU TRONG MÔ HÌNH CƠ SỞ DỮ LIỆU DẠNG QUAN HỆ

TRẦN THÁI SƠN – ĐINH THỊ NGỌC THANH

Tóm tắt. Bài này đưa ra một số kết quả nghiên cứu về tính bất biến trong các phủ không dư thừa. Tiếp tục nghiên cứu của D. Maier [1], chứng minh một số kết quả liên quan đến cấu trúc vế phải của các phụ thuộc hàm trong phủ tối thiểu. Trong bài này cũng đưa ra thuật toán tìm phủ «gần tối ưu» theo nghĩa tiết kiệm bộ nhớ.

1. Mở đầu.

Khái niệm phụ thuộc hàm có vai trò quan trọng trong vấn đề thiết kế và quản lý một cơ sở dữ liệu (CSDL) dạng quan hệ. Nhiều tác giả như W.W. Armstrong, D. Maier đã tiến hành nghiên cứu các phủ tối thiểu - tập tối thiểu có chọn lọc các phụ thuộc hàm mà từ đó sinh ra mọi phụ thuộc hàm xác định trong CSDL đã cho nhờ các quy tắc suy dẫn. D.Maier trong [1] đã nghiên cứu cấu trúc vế trái của các phụ thuộc hàm trong phủ tối thiểu, đưa ra thuật toán tìm phủ tối thiểu của một tập phụ thuộc hàm trong thời gian đa thức. Tuy nhiên, quy luật biến đổi của các vế phải chưa được tìm hiểu. W.W. Armstrong trong [2,3], khi nghiên cứu các phủ có tập thuộc tính bên vế phải tối đại, đã chỉ ra các vế phải tối đại này xác định duy nhất bởi vế trái và có thể tìm tất cả các phủ tối thiểu của các thành phần tối đại đó. Rõ ràng các phủ đó mang nhiều thông tin dư thừa bên vế phải. Trong bài này chúng tôi trình bày nghiên cứu quy luật biến đổi các vế phải của các phủ tối thiểu đã rút gọn trái, phải.

Để tiện trình bày, xin nhắc lại một số khái niệm cơ bản đưa ra trong [1]. Cho F, G là các tập phụ thuộc hàm, nói F phủ G nếu $F^+ = G^+$, nghĩa là F và G suy ra cùng một họ các phụ thuộc hàm theo các quy tắc suy diễn. Khi đó ta nói F và G là tương đương và ký hiệu $F \equiv G$.

Tập phụ thuộc hàm F được gọi là không dư thừa (hàm) nếu không tồn tại tập G thực sự chứa trong F sao cho $G^+ = F^+$. Tập phụ thuộc hàm F gọi là tối thiểu nếu không tồn tại tập phụ thuộc hàm G có ít số phụ thuộc hàm hơn tương đương với nó.

Thuộc tính A trong phụ thuộc hàm f của tập F gọi là rút gọn được nếu khi loại bỏ nó khỏi f vẫn thu được phủ tương đương với phủ ban đầu. F gọi là L–rút gọn (R–rút gọn) nếu không có thuộc tính nào ở vế trái (phải) của F là rút gọn được. F là LR–rút gọn nếu là L–rút gọn và R–rút gọn. Ta ký hiệu LF và RF là các tập thuộc tính bên vế trái và vế phải của các phụ thuộc hàm của F.

Các tập thuộc tính X và Y gọi là tương đương theo tập phụ thuộc hàm F, ký hiệu $X \leftrightarrow Y$, nếu $X \to Y$ và $Y \to X$ thuộc F^+. Ký hiệu $E_F(X)$ là tập các phụ thuộc hàm trong F có vế trái tương đương với X theo F. X gọi là suy ra trực tiếp Y, ký hiệu $X \twoheadrightarrow Y$, nếu $X \to Y$ thuộc $(F \setminus E_F(X))^+$. X và Y tương đương trực tiếp nếu $X \twoheadrightarrow Y$ và $Y \twoheadrightarrow X$, ký hiệu $X \twoheadleftrightarrow Y$.

Ngoài ra bài này còn sử dụng hệ tiên đề về họ đầy đủ và định lý sau do Armstrong đưa ra trong [2].

Định lý 1. Cho $G = \{X_i \to Y_i \; i = 1 \ldots k\}$ là tập các phụ thuộc hàm. Gọi J là tập mọi phụ thuộc hàm $X \to Y$ sao cho tồn tại dãy các phụ thuộc hàm trong G,

$$X_{i_j} \to Y_{i_j}, \; j = 1 \ldots m, \; m \geqslant 0 \text{ để:}$$

$$X \supseteq X_{i_1}, \; XY_{i_1} \supseteq X_{i_2}, \ldots, XY_{i_1} \ldots Y_{i_k} \supseteq Y \; (*)$$

Khi đó J là họ đầy đủ nhỏ nhất các phụ thuộc hàm chứa G, $J = G^+$. Mỗi phụ thuộc hàm $X_{i_j} \to Y_{i_j}$ được gọi là tham gia vào dãy suy dẫn trong G ra $X \to Y$.

2. Các bất biến trong phủ không dư thừa.

Để đi sâu nghiên cứu cấu trúc của các phủ tối thiểu, trước hết ta xét một số bất biến trong các phủ không dư thừa.

Định lý 2. Nếu F_1 và F_2 là hai phủ không dư thừa tương đương bất kỳ thì $LF_1 \cup RF_1 = LF_2 \cup RF_2$.

Chứng minh: Ta chứng minh $RF_1 \subseteq LF_2 \cup RF_2$. Với bất kỳ thuộc tính $A \in RF_1$, tồn tại X, Y sao cho $X \to AY$ thuộc $F_1 \subseteq F_2^+$. Theo định lý 1, có dãy dẫn xuất trong F_2 ra $X \to AY$, $X_{i_j} \to Y_{i_j}$, $j = 1 \ldots m$ thỏa (*). Dễ dàng thấy $A \in XY_{i_1} Y_{i_2} \ldots Y_{i_m}$. Nếu $A \notin X$ thì $A \in Y_{i_1} \ldots Y_{i_m} \subseteq RF_2 \subseteq LF_2 \cup RF_2$. Nếu $A \in X$, bỏ qua trường hợp đơn giản ta giả thiết $A \notin X_{i_1} X_{i_2} \ldots X_{i_m}$ thì khi thay X bởi $X \setminus A$ trong (*) ta thu được cùng một dãy suy dẫn ra $X \setminus A \to Y$ trong F_2. Lại áp dụng định lý 1 cho phụ thuộc hàm $X \setminus A \to Y \in F_2^+ = F_1^+$ ta có trong F_1 dãy suy dẫn $Z_{i_j} \to W_{i_j}$, $j = 1, \ldots, k$ trong F_1 cho $X \setminus A \to Y$. Trong dãy đó phải xuất hiện $X \to AY \in F_1$ vì ngược lại có thể nhận được $X \to AY$ từ $X \setminus A \to Y$ bởi từ tập con không chứa $X \to AY$ của F_1, vi phạm tính không dư thừa của F_1. Gọi j là chỉ số nhỏ nhất trong dãy mà $X \to AY = Z_{i_j} \to W_{i_j}$. Rõ ràng $j > 1$, $A \notin X \setminus A \supseteq Z_{i_j}$ và $A \in W_{i_1} \ldots W_{i_{j-1}}$. Vậy tồn tại chỉ số $e \leqslant j$ để $Z_{i_e} \to W_{i_e} \in F_1$ mà $A \notin Z_{i_e}$ và $A \in W_{i_e}$. Lặp lại chứng minh trường hợp $A \notin X$ cho $Z_{i_e} \to W_{i_e}$ ta thu được kết quả $A \in LF_2 \cup RF_2$. Kết hợp các trường hợp ta đều có $RF_1 \subseteq LF_2 \cup RF_2$.

Bây giờ ta chứng minh $LF_1 \subseteq LF_2 \cup RF_2$. Lấy bất kỳ thuộc tính $B \in LF_1$, tồn tại Z, W sao cho $ZB \to W$ thuộc F_1. Vì F_1 không dư thừa và $F_2 \equiv F_1$ nên $ZB \to W$ phải tham gia vào dãy suy dẫn $U_{i_j} \to W_{i_j}$, $j = 1, \ldots, k$ trong F_1 của một phụ thuộc hàm $U \to V \in F_2$, nghĩa là tồn tại n để $ZB \to W = U_{i_n} \to V_{i_n}$. Ta có $ZB \subseteq UV_{i_1} \ldots V_{i_{n-1}}$. Vậy hoặc $B \in U \subseteq LF_2$ hoặc $B \in V_{i_1} \ldots V_{i_{n-1}} \subseteq RF_1 \subseteq LF_2 \cup RF_2$. Bởi thế $LF_1 \subseteq LF_2 \cup RF_2$. Kết hợp các trường hợp lại, với vai trò đối xứng của F_1 và F_2, ta chứng minh được định lý.

Định lý 3. Nếu F_1 và F_2 là hai phủ tương đương, F_1 không dư thừa thì $LF_1 \setminus LF_2 \subseteq RF_1$.

Chứng minh: Giả sử $A \in LF_1 \setminus LF_2$, khi đó tồn tại X, Y sao cho $AX \to Y \in F_1$. Vì F_1 không dư thừa nên $AX \to Y$ phải tham gia vào dãy suy dẫn $X_{i_j} \to Y_{i_j}$, $j = 1, \ldots, k$ trong F_1 của một phụ thuộc hàm $Z \to W \in F_2 \subseteq F_1^+$. Vậy theo định lý 1, $AX \subseteq ZY_{i_1} \ldots Y_{i_{j-1}}$ mà j là chỉ số nhỏ nhất trong dãy sao cho $AX \to Y = X_{i_j} \to Y_{i_j}$. Vì $A \in LF_2 \supseteq Z$ nên $A \in Y_{i_1} \ldots Y_{i_{j-1}} \subseteq RF_1$.

Định lý 4. Nếu F_1 và F_2 là hai phủ tương đương R-rút gọn thì $RF_1 = RF_2$.

Chứng minh: Lấy bất kỳ $A \in RF_1$, khi đó tồn tại X, Y để $X \to AY \in F_1$ và $A \notin X$. Từ dãy suy dẫn $X_{i_j} \to Y_{i_j}$, $j = 1, \ldots, m$, trong F_2 của $X \to AY$ theo định lý 1, ta có $A \in Y_{i_1} \ldots Y_{i_m} \subseteq RF_2$. Tương tự có $RF_2 \subseteq RF_1$. Bởi vậy $RF_1 = RF_2$.

Hệ quả 1. Nếu F_1 và F_2 là hai phủ tương đương không dư thừa và R-rút gọn thì $LF_1 \setminus RF_1 = LF_2 \setminus RF_2$.

Định lý 5. Nếu F_1 và F_2 là hai phủ tương đương không dư thừa và L – rút gọn thì $LF_1 = LF_2$.

Chứng minh: Giả sử $LF_1 \neq LF_2$. Không mất tổng quát ta giả thiết $A \in LF_1 \setminus LF_2 \neq \phi$ Khi đó tồn tại X, Y sao cho $AX \to Y \in F_1$, $Y \setminus A \neq \phi$ và $AX \to (Y \setminus A)$ thuộc $F_1^+ = F_2^+$. Theo định lý 1, tồn tại dãy suy dẫn trong F_2 ra $AX \to Y \setminus A$ thỏa (*). Vì $A \overline{\in} LF_2$ nên khi bỏ A khỏi các điều kiện trong (*) ta vẫn thu được cùng dãy suy dẫn trong F_2 của $X \to Y \setminus A$. Vậy $X \to Y \setminus A \in F_1^+$ và A là thuộc tính vế trái rút gọn được của F_1, trái với giả thiết Bởi thế định lý được chứng minh.

Từ các định lý 2–5 cho phép kết luận rằng sau khi loại các thuộc tính rút gọn được, các tập LF và RF là như nhau đối với các phủ tương đương không dư thừa. Việc rút gọn các thuộc tính có ý nghĩa tự nhiên vì các thuộc tính rút gọn được không có vai trò quan trọng trong việc suy diễn các phụ thuộc hàm. Bởi thế ta có thể chỉ cần đi sâu nghiên cứu cấu trúc của các vế phải và trái với các thuộc tính nằm trong các tập bất biến và ký hiệu L, R thay cho LF và RF đối với các phủ tương đương đã rút gọn được xét đến sau này.

3. Các lớp tương đương theo vế trái.

Trong [1] D.Maier đã đưa ra khái niệm các lớp tương đương theo vế trái $E_F(X)$ và chứng minh với hai phủ tương đương tối thiểu bất kỳ F_1 và F_2 số lượng các phụ thuộc hàm trong $E_{F_1}(X)$ và $E_{F_2}(X)$ là bằng nhau. Hơn thế, mỗi phụ thuộc hàm $X_i \to \overline{X}_i \in E_{F_1}(X)$ có duy nhất một phụ thuộc hàm $Y_i \to \overline{Y}_i \in E_{F_2}(X)$ sao cho $X_i \leftrightarrow Y_i$. Bởi thế các vế trái có thể thay thế cho nhau mà không ảnh hưởng tới phủ, nghĩa là $(F_1 \setminus (X_i \to \overline{X}_i)) \cup Y_i \to \overline{X}_i \equiv F_1$. Để nghiên cứu cấu trúc của vế phải, tìm quy luật biến đổi của chúng, nhằm thuận tiện và tiết kiệm trong việc quản lý, từ kết quả nhắc đến trên của D.Maier, ta có thể giả thiết các vế trái của các phủ tối thiểu tương đương là như nhau. Ta chứng minh bổ đề sẽ dùng nhiều ở sau.

Bổ đề 1. Cho $Z \to W$ là phụ thuộc hàm trong F^+, giả sử $X \to Y \in F$ tham gia vào dãy suy dẫn $X_{ij} \to Y_{ij}$, $j = 1, \ldots, m$, của $Z \to W$ theo định lý 1. Khi đó có

$$Z \to X,\ ZY \to W \in (F \setminus (X \to Y))^+.$$

Chứng minh: Gọi k là chỉ số nhỏ nhất trong dãy suy dẫn ra $Z \to W$ mà $X \to Y = X_{ik} \to Y_{ik}$ theo giả thiết. Khi đó $ZY_{i_1} \ldots Y_{i_{k-1}} \supseteq X$ và cũng theo định lý 1 ta thu được dãy suy dẫn của $Z \to X$ trong $F \setminus (X \to Y)$, nghĩa là $Z \to X \in (F \setminus (X \to Y))^+$. Xét tiếp dãy bằng cách loại bỏ các phụ thuộc dạng $X_{i_j} \to Y_{i_j} = X \to Y$ nhờ gia tăng Z bởi ZY, ta thu được dãy suy dẫn trong F cho $ZY \to W$ mà không chứa $X \to Y$. Bởi vậy theo định lý 1 cũng có $ZY \to W \in (F \setminus X \to Y)^+$, kết thúc chứng minh bổ đề.

Từ bổ đề ta thu được hệ quả hiển nhiên:

Hệ quả 2. Với mọi $Z \to W \in F^+$ và $X \to Y \in F$, ta luôn có $ZY \to W \in (F \setminus X \to Y)^+$.

Định lý 6. Cho F_1 và F_2 là hai phủ tối thiểu tương đương. Khi đó $(F_1 \setminus E_{F_1}(X)) \cup E_{F_2}(X) \equiv$ $\equiv F_1 \equiv (F_2 \setminus E_{F_2}(X)) \cup E_{F_1}(X)$.

Chứng minh: Ta chứng minh $F_1 \setminus E_{F_1}(X)) \cup E_{F_2}(X) \equiv F_1$. Trước hết, nếu $X' \to Y' \in$ $\in (F_1 \setminus E_{F_1}(X)) \cup E_{F_2}(X)$ thì $X' \to Y' \in (F_1 \cup F_2)^+ = F_1^+$. Ta chỉ cần chứng minh chiều ngược lại. Lấy bất kỳ $Z \to Y \in F_1$, bỏ qua trường hợp hiển nhiên ta giả thiết $Z \to Y \in E_{F_1}(X)$ và $Z \to Y$ không suy dẫn được chỉ từ các phụ thuộc hàm trong $E_{F_2}(X)$. Khi đó $Z \to Y \in F_2^+ =$ $= ((F_2 \setminus E_{F_2}(X)) \cup E_{F_2}(X))^+$ và có $\phi \neq \{ W_i \to V_i \mid i = 1, \ldots, k \} \subseteq F_2 \setminus E_{F_2}(X)$ là tập tất cả các phụ thuộc hàm trong $F_2 \setminus E_{F_2}(X)$ tham gia vào dãy suy dẫn ra $Z \to Y$. Theo bổ đề 1, có $Z \to W_i \in F_2^+ = F_1^+$ với mọi $i = 1, \ldots, k$. Giả sử tồn tại $i_0 \leqslant k$ mà $W_{i_0} \to V_{i_0}$ có $T \to U \in E_{F_1}(X)$ tham gia vào dãy suy dẫn trong F_1 ra $W_{i_0} \to V_{i_0}$. Theo bổ đề 1, có $W_{i_0} \to T \in F_1^+$. Vì $T \to U \in E_{F_1}(X)$ nên $T \leftrightarrow X \in F_1^+$ và theo tính chất bắc cầu có $V_{i_0} \to X \in F_1^+$. Do $Z \to W_{i_0} \in F_1^+$, vậy $X \leftrightarrow W_{i_0} \in F_1$, nhưng điều này mâu thuẫn với giả thiết $W_{i_0} \to W_{i_0} \in F_2 \setminus E_{F_2}(X)$. Bởi thế có $Z \to Y \in ((F_1 \setminus E_{F_1}(X)) \cup E_{F_2}(X))$. Định lý được chứng minh

Định lý trên cho phép giới hạn việc nghiên cứu cấu trúc vế phải của phủ nói chung về nghiên cứu cấu trúc vế phải của các phụ thuộc hàm trong một lớp tương đương. Hơn thế, có thể nhận được phủ tương đương khi thay cả lớp tương đương bằng tập phụ thuộc hàm thích hợp trong khi không thể chỉ thay đổi từng phụ thuộc hàm một. Thí dụ phủ $\{A \to B, B \to C, C \to A\}$ không thể thay riêng lẻ bất kỳ phụ thuộc hàm nào tuy nhiên nó có phủ tương đương $\{A \to C, C \to B, B \to A\}$.

4. Một số quy luật biến đổi các tập thuộc tính vế phải của các phụ thuộc hàm trong phủ tối thiểu.

Ở đây ta chứng minh định lý có phát biểu hình thức tương tự định lý trong [1] về các tập thuộc tính bên vế phải. Ký hiệu $LE_F(X)$ và $RE_F(X)$ là các tập thuộc tính vế trái và vế phải của các phụ thuộc hàm trong $E_F(X)$. Ta viết $F \setminus X \to Z_0$ thay cho việc viết $F \setminus X \to \overline{X}) \cup (X \to \overline{X} \setminus Z_0)$ nếu $Z_0 \subseteq \overline{X}$ và $X \to \overline{X} \in F$.

Định lý 7. Cho F_1 và F_2 là hai phủ tối thiểu tương đương, $X_1 \to \overline{X_1} \in E_{F_1}(X)$. Giả sử $Z_0 \subseteq \overline{X_1}$, $Z_0 \cap E_F(X) = \phi$. Khi đó tồn tại Z sao cho $X_1 Z \leftrightarrow X_1 Z_0 \in (F_1 \setminus X_1 \to Z_0)^+$. Ta nói khi đó Z_0 và Z tương đương nhờ X_1.

Chứng minh: Ta tách $X_1 \to \overline{X_1} \in E_{F_1}(X)$ thành hai phụ thuộc hàm $X_1 \to Z_0$ và $X_1 \to \overline{X_1} \setminus Z_0$ và cho

$$E_{F_2}(X) = \{X_1 \to \overline{Y_1},\ X_2 \to \overline{Y_2}, \ldots, X_k \to \overline{Y_k}\}$$

Xét phụ thuộc hàm $X_1 \overline{Y_1} \to \overline{X_1}$, thuộc $(F_2 \setminus \overline{X_1} \to \overline{Y_1})^+$ theo hệ quả 1. Nếu $X_1 \to Z_0$ không tham gia suy dẫn $X_1 \overline{Y_1} \to \overline{X_1}$, trong F_1, thì $X_1 \overline{Y_1} \to \overline{X_1} \in (F_1 \setminus X_1 \to Z_0)^+$ và chọn được $\overline{Y_1}$ là tập Z phải tìm. Thật vậy, ta cũng có $X_1 \overline{X_1} \to \overline{Y_1} \in (F_1 \setminus X_1 \to \overline{X_1})^+ \subseteq (F_1 \setminus X_1 \to Z_0)^+$ theo hệ quả 1, mà $X_1 \to \overline{X_1} \setminus Z_0 \in (F_1 \setminus X_1 \to Z_0)^+$ nên theo quy tắc bắc cầu thu được $X_1 Z_0 \to \overline{Y_1} \in (F_1 \setminus X \to Z_0)^+$.

Xét trường hợp $X_1 \to Z_0$ tham gia suy dẫn $X_1 \overline{Y_1} \to \overline{X_1}$, khi đó lấy phụ thuộc hàm $X_2 \to \overline{Y_2} \in E_{F_2}(X)$ và gia tăng để xét phụ thuộc hàm $X_1 \overline{Y_1}\,\overline{Y_2} \to \overline{X_1} \in (F_2 \setminus \{X_1 \to \overline{Y_1}, X_2 \to \overline{Y_2}\})^+$ thay cho $X_1 \overline{Y_1} \to \overline{X_1}$. Tương tự dẫn dắt trên ta có thể chọn được $\overline{Y_1}\,\overline{Y_2}$ làm Z phải tìm khi $X_1 \to Z_0$ không tham gia suy dẫn ra $X_1 \overline{Y_1}\,\overline{Y_2} \to \overline{X_1}$ trong F_1, hoặc lại gia tăng tiếp bởi phụ thuộc hàm $X_3 \to \overline{Y_3} \in E_{F_2}(X)$, v.v... Vì $E_{F_2}(X)$ hữu hạn nên trường hợp xấu nhất ta đi đến xét phụ thuộc hàm $X_1 \overline{Y_1} \ldots \overline{Y_k} \to \overline{X_1} \in (F_2 \setminus E_{F_2}(X))^+$. Đến đây $X_1 \to Z_0$ không thể tham gia suy dẫn trong F_1 ra $X_1 \overline{Y_1} \ldots \overline{Y_k} \to \overline{X_1}$. Thật vậy, giả sử $X_1 \to Z_0$ tham gia suy dẫn ra $X_1 \overline{Y_1} \ldots \overline{Y_k} \to \overline{X_1}$ trong F_1, nghĩa là tồn tại $W_j \to \overline{V_j} \in F_2 \setminus E_{F_2}(X)$ tham gia suy dẫn $X_1 \overline{Y_1} \cdots \overline{Y_k} \to \overline{X_1}$ mà $X_1 \to Z_0$ lại tham gia suy dẫn ra $W_j \to \overline{V_j}$ trong F_1. Sử dụng bổ đề 1 ta có $W_j \to X_1$ và $X_1 \overline{Y_1} \ldots \overline{Y_k} \to W_j$, trong khi có $X_1 \to \overline{Y_1} \ldots \overline{Y_k}$

theo tính chất của lớp tương đương $E_{F_2}(X)$ và $E_{F_1}(X)$ nên $W_1 \leftrightarrow X_1 \in F_1^+ = F_2^+$, trái với giả thiết $W_1 \to \overline{V_1}$ thuộc $F_2 \setminus E_{F_2}(X)$. Ta thu được chứng minh của định lý. Lưu ý rằng trong quá trình chọn Z ở trên, vì $Z_o \cap R E_{F_2}(X) = \phi$ theo giả thiết, nên $Z_o \neq Z \subseteq \overline{Y_1}\ \overline{Y_2} \ldots \overline{Y_k} = RE_{F_2}(X)$.

Hệ quả 3. Với Z_o và Z có tính chất như trong định lý 7, ta có thể thay Z_o trong $\overline{X_1}$ của $X_1 \to \overline{X_1} \in F$ bởi Z mà vẫn nhận được một phủ tối thiểu mới tương đương.

Qua định lý và hệ quả, tập thuộc tính vế phải của một phụ thuộc hàm trong $E_F(X)$ có thể chia thành hai lớp theo quy luật biến đổi như sau. Một là lớp thuộc tính vế phải có thể thay bởi tập tương đương theo vế trái của nó. Hai là lớp thuộc tính bất biến theo vế phải mà giữ nguyên vị trí hoặc chỉ chuyển đổi vị trí trong các vế phải của các phụ thuộc hàm trong cùng lớp tương đương mà ta sẽ xét tiếp sau này. Bây giờ ta nêu một trường hợp riêng đáng quan tâm.

Hệ quả 4. Nếu $Z_o \subseteq \overline{X_1}$ ở đó $X_1 \to \overline{X_1} \in E_{F_1}(X)$ và $Z_o \subseteq R \setminus L$ thì trong mọi phủ tối thiểu không dư thừa LR - rút gọn $F_1 \equiv F_2$, ta có $Z_o \subseteq RE_{F_2}(X)$.

Chứng minh: Ta thấy nếu tồn tại Z để $X_1Z_o \leftrightarrow X_1Z \in F_1^+$ như trong định lý thì phải có phụ thuộc hàm $Y \to W \in F_1$ tham gia suy dẫn $X_1Z_o \to Z$ mà $Y \cap Z_o \neq \phi$. Khi đó $Z_o \notin R \setminus L$.

Bây giờ, theo cách chia lớp ở trên, xét $E_F(X)$ là tập phụ thuộc hàm dạng:

$$\{X_1 \to \overline{X_1}\ \overline{X'_1},\ X_2 \to \overline{X_2}\ \overline{X'_2}, \ldots, X_k \to \overline{X_k}\ \overline{X'_k}\}$$

với $X_i \subseteq LE_F(X)$ còn $\overline{X'_i} \cap LE_F(X) = \phi$ với $i = 1, \ldots, k$. Bằng cách thay các vế trái vào các tập vế phải $\overline{X_i}$ tương ứng như sau đối với $E_F(X)$:

$$\{X_1 \to X_{i_1}\ \overline{X'_1},\ X_2 \to X_{i_2}\ \overline{X'_2}, \ldots, X_k \to X_{i_k}\ \overline{X'_k}\} \quad (**)$$

sao cho $X_{i_1}, X_{i_2}, \ldots, X_{i_k}$ là hoán vị của $X_1, X_2, \ldots X_k$ mà $i_j \neq j$ với mọi $j = 1, \ldots, k$, ta thu được phủ mới tương đương với F.

Lấy bất kỳ $A \in RE_F(X)$, phải có $A \in \overline{X_i}\ \overline{X'_i}$ mà $X_i \to \overline{X_i}\ \overline{X'_i} \in E_F(X)$. Nếu $A \in \overline{X_i} \subseteq LE_F(X)$ thì tồn tại $j \leq k$ để $A \in X_j$, $X_j \to \overline{X_j}\ \overline{X'_j} \in E_F(X)$. Nếu tồn tại $m \neq j$ để $A \in X_m$, $X_m \to \overline{X_m}\ \overline{X'_m} \in E_F(X)$ thì ta có thể tìm được hoán vị tương đương (**) trong đó xuất hiện $X_i \to X_m\ \overline{X'_i}$ và $X_m \to X_j\ \overline{X'_m}$ mà sau khi rút gọn phủ tương đương mới thu được, thuộc tính A chuyển đổi vị trí từ phụ thuộc hàm thứ i sang vế phải của phụ thuộc hàm m khác trong cùng lớp tương đương nếu A thuộc lớp thuộc tính bất biến của vế phải.

Nếu $A \in \overline{X'_i} \notin LE_F(X)$ thì có thể chọn hoán vị tương đương trong đó lớp $E_F(X)$ được thay chẳng hạn bởi:

$$X_1 \to X_2,\ X_2 \to X_3, \ldots, X_{k-1} \to X_k,\ X_k \to X_1\ \overline{X'_1} \ldots \overline{X'_k}\} \quad (***)$$

Sau khi rút gọn phủ mới tương đương nhận được, thuộc tính A chuyển vị trí từ phụ thuộc hàm thứ i sang k nếu như A là thuộc tính bất biến của các vế phải trong $E_F(X)$.

Tuy các thuộc tính vế phải có thể chuyển đổi vị trí trong các phụ thuộc hàm của lớp tương đương $E_F(X)$ như đã xét ở trên, chúng vẫn phải thỏa mãn tính chất sau để đảm bảo thông tin suy dẫn cho tính tương đương của các vế trái trong $E_F(X)$.

Tính chất 1: Giả sử $E_F(X) = \{X_i \to Y_i \ i = 1,\dots, k\}$. Khi đó với mọi $i \leq k$, tồn tại $j \leq k$ sao cho $X_i \to X_j$ thuộc $((F \setminus E_F(X)) \cup X_i \to Y_i)^+$.

Chứng minh: Lấy i, m bất kỳ, ta có $X_i \to X_m \in F^+$.

Theo định lý 1, ta có dãy suy dẫn trong F, $X_{i_1} \to Y_{i_1}$, $1 = 1,\dots, p$, của $X_i \to X_m$ mà $X_i \supseteq X_{i_1}$, $X_i X_{i_1} \dots X_{i_p} \supseteq X_m$. Chọn chỉ số n nhỏ nhất có thể sao cho trong dãy đó $X_{i_n} \to Y_{i_n} \in E_F(X)$, với X_{i_n} đó ta có kết quả phải chứng minh. Trường hợp ngược lại X_m thỏa mãn điều kiện đòi hỏi.

Trong [1], D. Maier đã đưa ra khái niệm phủ tối ưu trong đó số lần xuất hiện của các thuộc tính trong phủ là ít nhất và chứng minh bài toán tìm phủ tối ưu của một tập phụ thuộc hàm đã cho là NP–đầy đủ. Tuy nhiên về mặt quản lý trong máy thì vẫn chưa phải là tiết kiệm nhất, vì đối với $E_F(X)$ phải quản tất cả các thông tin về các vế trái lẫn phải, trong khi đó có phủ tương đương với nó, chẳng hạn dạng (***), rút gọn được việc quản lý các $\overline{X_i}$ trong vế phải của $E_F(X)$ nhờ gia tăng chúng bằng các vế trái. Ta gọi phủ thu được bằng cách đó là phủ « gần tối ưu ».

Để kết thúc, có thể đưa ra thuật toán tìm phủ gần tối ưu nêu trên với lưu ý có rút gọn thuộc tính trong tập $\overline{X}'_1, \overline{X}'_2, \dots, \overline{X}'_k = RE_F(X) \setminus LE_F(X)$.

Thuật toán

VÀO: Họ các phụ thuộc hàm G.

RA : Phủ gần tối ưu F, $F^+ = G^+$.

Bước 1. Từ G tìm phủ tối thiểu F_1 theo [1] và rút gọn trái, phải được phủ với các lớp tương đương:

$$E_F(X^i) = \{X^i_j \to \overline{Y}^i_j \mid j = 1,\dots, k_i\},\ i = 1;\dots, m,$$

Bước 2. Với mỗi lớp tương đương.

Đặt $M_i = RE_F(X^i) \setminus L$.

Cho $j = 1,\dots, k_i$, xét $X^i_j \to \overline{Y}^i_j$:

Với mọi $A \in \overline{Y}^i_j$, kiểm tra

nếu $A \in LE_F(X^i)$ thì bỏ qua,

nếu $A \notin LE_F(X^i)$ mà $\exists Z \supseteq LE_F(X^i) \setminus X^i_j$

thỏa $X^i_j Z \to A \in (F \setminus X^i_j \to A)^+$

thì bỏ qua, ngược lại đặt $M_i = M_i \cup A$.

Bước 3. Kết thúc chương trình khi đã duyệt hết A của mọi lớp tương đương và thu được phủ F gần tối ưu có dạng:

$$\bigcup_{i=1}^{m} \{X^i_1 \to X^i_2;\ X^i_2 \to X^i_3, \dots,\ X^i_k \to X^i_1 M_i\}$$

Kết luận: Trong bài này chúng tôi đã chứng minh một số kết quả liên quan đến tính bất biến của các tập thuộc tính bên vế trái và vế phải trong các phủ tương đương không dư thừa, đồng thời chỉ ra một số quy luật biến đổi và khả năng phân bố của các thuộc tính bên vế phải của lớp phụ thuộc hàm tương đương theo vế trái. Chúng tôi hy vọng các kết quả này giúp thêm cho việc hiểu cấu trúc của các phủ tối thiểu để có thể thiết kế và quản lý CSCL được thuận tiện và tiết kiệm.

Chúng tôi xin bày tỏ sự biết ơn chân thành Phó tiến sĩ Nguyễn Cát Hồ đã đặt vấn đề và hướng dẫn viết bài báo này.

Nhận ngày 12-1-1985.

TÀI LIỆU THAM KHẢO

1. D. Maier,. Minimum Cover in the Relational Database Model, JACM 1980 Vol 27, N°4.

2. W.W. Armstrong,. On the Generation of Dependency Structures of Relational Data Bases. pub. 272 Université de Montréal 1977.

3. W.W. Armstrong,. Dependency structures of Database relation ships. Proc. IFIP 74 North Holland (1974).

ABSTRACT

Some results on structures of a minimum cover in the relational database model

In this paper several properties of a minimum cover in the relational database are provided. It is shown that set of attributes on the left side and the right side of functional dependencies (FD) are the same for all equivalent minimum covera.

D. Maier has studied equivalence and transformations of left side of FDs [1]. In this paper we have studied equivalence and transformations of right side of one. We also obtained an algorithm for finding pseudooptimal cover from a given set of FDs.