Tạp Chí Tin Học và Điều Khiển Học
Tập X(1994), Số 2, 36-42

Một số kết quả trên phụ thuộc mạnh

Vũ Đức Thi
Viện Công nghệ thông tin

I. Mở đầu

Phụ thuộc mạnh được đề xuất và tiên đề hóa trong [2,3,4] và đã được cài đặt trong hệ thiết kế các hệ quản trị cơ sở dữ liệu có đặc tính nổi bật thuận lợi là phân tách các mảng dữ liệu lớn thành các mảng dữ liệu nhỏ hơn. Bài báo này trình bày các kết quả mới thu được về phụ thuộc mạnh. Trước tiên, chúng tôi đưa ra thuật toán tìm bao đóng của một thuộc tính với độ phức tạp là tuyến tính $O(p)$, ở đây $p = l(S)$ - chiều dài của tập phụ thuộc mạnh S. Dựa vào thuật toán này, chúng tôi đưa ra thuật toán giải quyết bài toán thành viên. Khái niệm khóa được đưa ra với ý nghĩa xác định duy nhất một đối tượng. Chúng tôi đưa ra thuật toán tìm một khóa với độ phức tạp là đa thức, chúng tôi cũng chỉ ra được rằng đối với phụ thuộc mạnh mọi khóa đều có lực lượng như nhau và bài toán tìm tập tất cả các khóa tối thiểu của S (Ks) được giải quyết bằng một thuật toán có độ phức tạp là đa thức nếu lực lượng của Ks là đa thức.

2. Các định nghĩa

Định nghĩa 1.1: Cho $R = \{h_1, ..., h_m\}$ là quan hệ trên Ω và $A, B \subseteq \Omega$. Chúng ta nói rằng B phụ thuộc mạnh vào A trong R (kí hiệu $A \xrightarrow[R]{S} B$) nếu

$$(\forall h_i,\ h_j \in R)((\exists a \in A)\ (h_i(a) = h_j(a)) \to (\forall b \in B)\ (h_i(b) = h_j(b))).$$

Đặt $S_R = \{A \xrightarrow[R]{S} B\}$. S_R được gọi là họ đủ các phụ thuộc mạnh của R.

Định nghĩa 1.2: Cho R là quan hệ trên $\Omega. A, B \subseteq \Omega$. Chúng ta nói $A \rightarrow B$ là một phụ thuộc mạnh của R nếu B phụ thuộc mạnh vào A trong R, hay $A \underset{R}{\xrightarrow{S}} B$. Chúng ta cũng nói rằng R thỏa mãn $A \rightarrow B$.

Định nghĩa 1.3: Cho S là một tập phụ thuộc mạnh trên Ω và $X \rightarrow Y$ là một phụ thuộc mạnh, với $X, Y \subseteq \Omega$. Chúng ta nói S kéo theo logic ra $X \rightarrow Y$, và viết $S \models X \rightarrow Y$ nếu với mọi quan hệ R trên Ω thỏa mãn các phụ thuộc mạnh trong S thì cũng thỏa mãn phụ thuộc mạnh $X \rightarrow Y$.

Định nghĩa 1.4: Cho Ω là một tập hữu hạn các thuộc tính, với $\forall A, B, C, D \subseteq \Omega$, $a \in \Omega$. Các S - luật suy dẫn là:

$S1$: $\{a\} \rightarrow \{a\}$;
$S2$: Nếu $A \rightarrow B$ và $B \rightarrow C$ với $B \neq \emptyset$ thì $A \rightarrow C$;
$S3$: Nếu $A \rightarrow B$ và $C \subset A$, $D \subseteq B$ thì $C \rightarrow D$;
$S4$: Nếu $A \rightarrow B$ và $C \rightarrow D$ thì $A \cup C \rightarrow B \cap D$;
$S5$: Nếu $A \rightarrow B$ và $C \rightarrow D$ với thì $A \cap C \rightarrow B \cup D$;

Dễ dàng nhận thấy rằng S_R thỏa mã các S-luật suy diễn, do vậy các S-luật suy diễn này cho phép ta suy diễn ra tất cảcác phụ thuộc mạnh khác từ S.

3. Bao đóng của tập phụ thuộc mạnh, bao đóng của các thuộc tính

Định nghĩa 2.1: Cho S là tập các phụ thuộc mạnh trên Ω. Đặt S^+ là bao đóng của S, là tập tất cả các phụ thuộc mạnh được kéo theo logic từ S, có nghĩa là

$$S^+ = \{X \rightarrow Y : S \models X \rightarrow Y\}.$$

Chúng ta không hy vọng liệt kê toàn bộ S^+, nhưng ít ra. chúng ta có thể biết được khi nào một phụ thuộc mạnh $X \rightarrow Y$ cho trước thuộc vào S^+, hay được suy diễn ra từ S.

Trước hết, chúng ta đưa ra khái niệm bao đóng của tập các thuộc tính ứng với tập các phụ thuộc mạnh đã cho.

Định nghĩa 2.2: Cho S là tập các phụ thuộc mạnh trên Ω, X là tập con của Ω, $A \subseteq \Omega$. Chúng ta gọi X^+, bao đóng của X ứng với S, là tập các thuộc tính a sao cho $X \rightarrow \{a\}$ có thể suy diễn từ S bởi các S-luật suy diễn, có nghĩa là: $X^+ = \{a \in \Omega : X \rightarrow \{a\} \in S^+\}$.

Ta có bổ đề sau:

Bổ đề 2.3: *Phụ thuộc mạnh $X \rightarrow Y$ được suy diễn từ S bởi các S-luật suy diễn nếu và chỉ nếu $Y \subseteq X^+$.*

Chứng minh. Giả sử $Y = a_1 a_2 \ldots a_k$ với $a_i \in \Omega$, $i = \overline{1,k}$. Nếu $Y \subseteq \Omega$, do định nghĩa của X^+, ta có $X \rightarrow \{a_i\}$ với $i = \overline{1,k}$. Vì vậy theo $S5$ ta có $X \rightarrow Y$. Ngược lại, giả sử $X \rightarrow Y$ được suy diễn từ S bởi các S-luật suy diễn. Với mỗi $i = \overline{1,k}$, $X \rightarrow \{a_i\}$ là đúng bởi ($S3$). Vậy $Y \subseteq X^+$. Bổ đề được chứng minh.

Liên quan đến khái niệm bao đóng và các tính chất của bao đóng, chúng ta đưa ra các hàm mạnh.

Định nghĩa 2.4: Cho Ω là một tập hữu hạn, ánh xạ $F: P(\Omega) \rightarrow P(\Omega)$ được gọi là hàm mạnh trên Ω nếu với mọi $a, b \in \Omega$ và $A \subseteq \Omega$, các tính chất sau là đúng:

(1) $a \in F(\{a\})$.
(2) $b \in F(\{a\}) \rightarrow F(\{b\}) \subseteq F(\{a\})$.
(3) $F(A) = \bigcap_{a \in A} F(\{a\})$.

Nhận xét 2.5: Dễ dàng nhận thấy rằng các tính chất sơ đẳng sau của hàm mạnh là đúng

(1) Với $A, B \in P(\Omega)$, ta có $F(A \cup B) = F(A) \cap F(B)$;
(2) Bởi quy ước $\cap\emptyset = \Omega$, ta có $F(\emptyset) = \cap\emptyset = \Omega$;
(3) Với $A \subseteq B$, thì $F(B) \subseteq F(A)$.

Bổ đề 2.6: *Cho S là tập các phụ thuộc mạnh trên Ω và $A \subseteq \Omega$. Chúng ta xây dựng ứng ánh xạ $F_s: P(\Omega) \rightarrow P(\Omega)$ như sau:*

$$F_s(A) = \{a \in \Omega : A \rightarrow \{a\} \in S^+\}$$

thì F_s là một hm mạnh.

Chứng minh. Gỉa sử S là tập các phụ thuộc mạnh, ta có $F_s(\{a\}) = \{b \in \Omega : \{a\} \rightarrow \{b\} \in S^+\}$ do luật $(S1)$, ta có $a \in F_s(\{a\})$. Do định nghĩa của F_s và luật $(S5)$, ta có $\{b\} \rightarrow F_s(\{b\})$ và nếu $b \in F(\{a\})$ thì $\{a\} \rightarrow \{b\}$. Do vậy, ta có $\{a\} \rightarrow \{b\}$ và $\{b\} \rightarrow F_s(\{b\})$, sử dụng luật $(S2)$, ta được $\{a\} \rightarrow F_s(\{b\})$ và $F_s(\{b\}) \subseteq F_s(\{a\})$ do định nghĩa của F_s.

Cho A là tập các thuộc tính, ta có $A \rightarrow F_s(A) \in S^+$, do vậy với mọi $a \in A$, ta có $\{a\} \rightarrow F_s(a)$ bởi luật $(S3)$ do đó $F_s(A) \subseteq F_s(\{a\})$ do định nghĩa của $F_s(\{a\})$. Vì vậy,

$$F_s(A) = \bigcap_{a \in A} F_s(\{a\}).$$

Vậy F_s là một hàm mạnh.

Nhận xét 2.7. Dễ dàng nhận thấy rằng $F_s(A)$ chính là bao đóng của tập A các thuộc tính ứng với tập các phụ thuộc mạnh S trên Ω do định nghĩa của F_s và định nghĩa bao đóng của một tập thuộc tính, có nghĩa là $A^+ = F_s(A)$ hay $A^+ = \bigcap_{a \in A} \{a\}^+$. Do vậy để tính bao đóng của một tập thuộc tính, trước hết ta cần phải tính bao đóng của mọi thuộc tính nằm trong thuộc tính đó.

Thuật toán 2.8: Tìm bao đóng của một thuộc tính cho trước ứng với tập các phụ thuộc mạnh đã cho.

Vào: Tập hữu hạn các thuộc tính Ω, tập hữu hạn các phụ thuộc mạnh S và $a \in \Omega$.
Ra: $\{a\}^+$, bao đóng của $\{a\}$ ứng với S.

Cách tính: Chúng ta tính dãy các tập thuộc tính $X^{(0)}, X^{(1)}, \dots$ theo quy tắc sau:

1. $X^{(0)} = \{a\}$,

2. $X^{(i+1)} = X^{(i)}$, nếu có một ph thuộc mạnh $Y \twoheadrightarrow Z \in S$ nào đó sao cho $Y \cap X^{(i)} \neq \emptyset$ và $A = Z \setminus X^{(i)} \neq \emptyset$, và $X^{(i+1)} = X^{(i)}$, trong trường hợp ngược lại.

Do vậy, chúng ta sẽ có một dãy các tập bao nhau $X^{(0)} \subseteq X^{(1)} \subseteq \dots \subseteq X^{(i)} \subseteq X^{(i+1)} \dots \subseteq \Omega$ và vì Ω là hữu hạn, các tập phụ thuộc mạnh là hữu hạn, nên chúng ta sẽ đi đến i mà $X^{(i)} = X^{(i+1)}$. Điều đó kéo theo rằng $X^{(i)} = X^{(i+1)} = X^{(i+2)} \dots$

Khi đó chúng ta lấy $\{a\}^+ = X^{(i)}$.

Nhận xét 2.9: Thuật toán có thể cài đặt để chạy trong thời gian tỷ lệ số các phụ thuộc mạnh trong S và chiều dài các phụ thuộc mạnh đó (số các thuộc tính trong các phụ thuộc mạnh của S). Do vậy, thuật toán tìm bao đóng của một thuộc tính cho trước có độ phức tạp là tuyến tính $O(p)$, ở đây $p = l(S)$ (p là chiều dài của S, tức là số các thuộc tính xuất hiện trong S kể cả số lần lặp các thuộc tính giống nhau).

Chúng ta phải chỉ ra rằng thuật toán trên là đúng. Dễ dàng nhận thấy rằng với mọi $i : X^{(i)} \subseteq \{a\}^+$ nhưng ngược lại thì khó hơn.

Định lý 2.10: *Thuật toán 2.8 tính đúng* $\{a\}^+$.

Chứng minh. Trước tiên chúng ta chỉ ra bằng quy nạp theo j rằng $X^{(j)} \subseteq \{a\}^+$. Ta phải chứng minh mệnh đề trên thỏa mãn 2 điều kiện:

1. Với $j = 0$ thì $X^{(0)} \subseteq \{a\}^+$.

Thực vậy, $X^{(0)} = \{a\}$, do đó hiển nhiên là $X^{(0)} \subseteq \{a\}^+$.

2. Cho $j \geq 1$, giả sử rằng $X^{(j-1)} \subseteq \{a\}^+$, ta phải chứng minh rằng $X^{(j)} \subseteq \{a\}^+$.

Ta chỉ ra cho trường hợp ngược lại, nếu $b \in \{a\}$ thì $b \in X^{(j)}$ nào đó. Thuật toán (2.8) kết thúc trước khi tính $X^{(j)}$ hay không là không quan trọng bởi vì nếu thuật toán này dừng khi $X^{(i)} = X^{(i+1)}$ với $i < j$ thì chúng ta có $X^{(i)} = X^{(j)}$. Trong thực tế ta có $i \leq m$, với m là phụ thuộc mạnh của S. $m = |S|$, vì chỉ có m phụ thuộc mạnh trong S thôi.

Chúng ta sẽ chỉ ra rằng nếu $\{a\} - Y \in S$, tức là $\{a\} - Y$ được suy diễn từ S bởi các S-luật suy diễn, thì nếu $b \in Y$, ta phải có $b \in X^{(j)}$ nào đó. Chứng minh được tiến hành bằng phương pháp quy nạp theo số dòng suy diễn ra Y từ S. Ở đây mỗi dòng là một phụ thuộc mạnh của S, hoặc là một phụ thuộc mạnh được suy diễn từ các dòng trước và dòng cuối cùng là $\{a\} \rightarrow Y$.

Chúng ta chứng minh hai điều kiện sau:

1. Với 1 dòng:

- $\{a\} \rightarrow Y \in S$, vì $\{a\} \cap X^{(j-1)} = \{a\} \neq \emptyset$ nên theo thuật toán (2.8) sẽ phải tồn tại một $j > 0$ nào đó sao cho hoặc là $A = Y \setminus X^{(j-1)} \neq \emptyset$ thì $X^{(j)} = X^{(j-1)} \cup A$ hoặc là $Y \subseteq X^{(j)}$ do đó $Y \subseteq X^{(j)}$.

- $\{a\} \rightarrow Y$ được suy diễn bởi $S1$, tức là $Y = \{a\}$ thì hiển nhiên là $Y \subseteq X^{(0)}$.

2. Giả sử các chứng minh với số dòng ít hơn p là đúng và $\{a\} - Y$ được suy diễn từ S bởi p dòng.

- $\{a\} - Y$ được suy diễn bởi $(S1)$, thì chúng ta chứng minh tương tự như điều kiện 1.

- $\{a\} - Y$ được suy diễn bởi $(S2)$, có nghĩa là có $\{a\} - B$ và $B \neq \emptyset$. Ta có $\{a\} - B$ được suy diễn từ S bởi số dòng ít hơn p. Theo giả thiết quy nạp có một $X^{(j)}$ nào đó để $B \subseteq X^{(j)}$.

Vì $B \cap X^{(j)} \neq \emptyset$ nên theo thuật toán 2.8 sẽ phải tồn tại một $i > j$ nào đó sao cho hoặc là $A = Y \setminus X^{(i-1)} = \emptyset$ thì $X^{(i)} = X^{(i-1)} \cup A$ hoặc là $Y \subseteq X^{(i)}$, do đó $Y \subseteq X^{(i)}$.

- $\{a\} - Y$ được suy diễn bởi $(S3)$, có nghĩa là có $B - C$ và $a \in B$, $Y \subseteq C$. Vì $B \cap X^{(0)} = \{a\} \neq \emptyset$ nên theo thuật toán 2.8 sẽ phải tồn tại một $j > 0$ nào đó sao cho hoặc là $A = C \setminus X^{(j-1)} = \emptyset$ thì $X^{(j)} = X^{(j-1)} \cup A$ hoặc là $C \subseteq X^{(j)}$, do đó $Y \subseteq X^{(j)}$.

- $\{a\} - Y$ được suy diễn bởi $(S4)$, có nghĩa là có $\{a\} - B$ và $\{a\} - D$ và $Y = B \cap D$, vì $\{a\} - B$ và $\{a\} - D$ $\{a\} - Y$ được suy diễn từ S bởi số dòng ít hơn p. Do vậy, ta có $Y \subseteq X^{(j)}$.

- $\{a\} - Y$ được suy diễn bởi $(S5)$, có nghĩa là có $A - B$, $C - D$ và $A \cap C = \{a\}$, $Y = B \cup D$. Như trong trường hợp với $(S3)$ với $A - Y$, $a \in C$ sẽ có $X^{(j)}$ sao cho $B \subseteq X^{(j)}$ và với $C - D$, $a \in C$, sẽ có $X^{(i)}$ sao cho $D \subseteq X^{(i)}$, do vậy $Y \subseteq X^{\max(i,j)}$. Định lý được chứng minh.

Thuật toán 2.11: Bài toán thành viên

Vào: Tập hữu hạn các thuộc tính Ω, tập hữu hạn các phụ thuộc mạnh S và phụ thuộc mạnh $X - Y$.

Ra: $X - Y \in S^+$?

Bước 1: Mỗi thuộc tính $a \in X$, tính $\{a\}^+$ theo thuật toán 2.8 $(O(np))$

Bước 2: Tính $X^+ = \cap\{a\}^+$ $(O(n^2))$

Bước 3 Kiểm tra $Y \subseteq X^+$? $(O(n))$

Chứng minh. Theo bổ đề 2.3 ta có $X - Y \in S^+ \Leftrightarrow Y \subseteq X^+$.

Nhận xét 2.12: Bài toán thành viên với tập S các phụ thuộc mạnh trên Ω và phụ thuộc mạnh $X - Y$ cho trước, quyết định $X - Y \in S^+$ hay không được giải quyết trong thời gian đa thức $O(n^2) + O(np)$, ở đây $n = |\Omega|$, $p = l(S)$ - chiều dài của tập phụ thuộc mạnh S.

Khóa

Khi nói về tập đối tượng, chúng ta giả sử rằng có một khóa tức là tập các thuộc tính mà nó xác định duy nhất một đối tượng.

Định nghĩa 3.1: Cho R là quan hệ trên Ω, và $K \subseteq \Omega$. Chúng ta nói K là khoá tối tiểu của R (Khóa tối tiểu của S) nếu:

Đặt $K = \{a_1, a_2, ..., a_k\}$. $F_R(\{a_j\}) = \{a \in \Omega : \{a_j\} \xrightarrow[S]{R} \{a\}\}$. $F_s(\{a_j\}) = \{a \in \Omega : \{a_j\} \to \{a\} \in S^+\}$.

$$1. \qquad \bigcup_{j=1}^{k} F_R(\{a_j\}) = \Omega (\bigcup_{j=1}^{k} F_s(\{a_j\}) = \Omega)$$

2. Với mọi tập con thực sự K' của K thì $\bigcup_{a_j \in K'} F_R(\{a_j\}) \subset \Omega$ ($\bigcup_{a_j \in K'} F_s(\{a_j\}) \subset \Omega$). Ta nhận thấy với mỗi quan hệ hay tập các phụ thuộc mạnh, có thể là một hay nhiều khóa tối tiểu. Kí hiệu K_R (K_s) là tập tất cả các khóa tối tiểu của $R(S)$. Rõ ràng rằng K_R, K_s là các hệ Sperner trên Ω.

Thuật toán 3.2: Tìm một khóa tối tiểu của tập các phụ thuộc mạnh S trên Ω.
Vào: Tập hữu hạn các thuộc tính Ω, tập hữu hạn các phụ thuộc mạnh S trên Ω.
Ra: $K \in K_s$.

Cách tìm: Giả sử $\Omega = \{a_1, a_2, ...a_n\}$.
Với mọi a_i, $i = \overline{1,n}$, sử dụng thuật toán 2.8 tính $F_s(\{a_i\}) = \{a_i\}^+$ $(O(np))$
Đặt $K_0 = \Omega$

$$K_i = \begin{cases} K_{i-1} \setminus \{a_i\} & \text{nếu } \exists a_j \in K_{i-1} \setminus \{a_i\} | a_i \in F_s(\{a_j\}) \\ K_{i-1} & \text{nếu ngược lại} \end{cases} \qquad (O(n^3))$$

thì K_n chính là khóa tối tiểu của S tức là $K_n \in K_s$.

Chứng minh. Do cách xây dựng K_j, ta có

$$\bigcup_{a_i \in K_j} F_s(\{a_i\}) = \Omega, \forall j = \overline{1,n},$$

hơn thế nữa không thể loại một phần tử nào của K_n được nữa, do vậy K_n là khóa tối tiểu. Dễ dàng nhận thấy thuật toán tìm một khóa tối tiểu được giải quyết trong thời gian đa thức $(O(np) + O(n^3))$, ở đây $n = |\Omega|$, $p = l(S)$ - chiều dài của tập phụ thuộc mạnh S.

Định nghĩa 3.3: Cho S là tập các phụ thuộc mạnh trên Ω và a, b là các thuộc tính của Ω. Chúng ta nói rằng a và b là tương đương (kí hiệu $\equiv$) nếu $F_s(\{a\}) = F_s(\{b\})$.
Kí hiệu $L(a) = \{b \in \Omega : b \equiv a\}$.

Định lý 3.4: *Cho S là tập các phụ thuộc mạnh trên Ω. $K = \{a_1, a_2, ..., a_k\}$ là khóa tối tiểu của S. Thì K' là khoá tối tiểu của S khi và chỉ khi $K' = \{b_1, b_2, ..., b_k\}$ với $b_i \in L(a_i)$, $i = \overline{1,k}$.*

Chứng minh. Do định nghĩa của khóa tối tiểu và định nghĩa phần tử tương đương, dễ dàng nhận thấy K' là khóa tối tiểu của S. Ngược lại giả sử $K'' = \{c_1, ..., c_l\}$ là khóa tối tiểu của S Đặt $L = \Omega \setminus \bigcup_{i=1}^{n} L(a_i)$. Xét $a_j \in L$, vì $a_j \notin K$ nên phải tồn tại $a_i \neq a_j$ sao cho $a_j \in F_s(\{a_i\})$ và $F_s(\{a_j\}) \neq F_s(\{a_i\})$. Vì K'' là khóa tối tiểu của S, có nghĩa là $\bigcup_{i=1} F_s(\{a_i\}) = \Omega$. do vậy tồn tại c_p để $a_i \in F_s(\{c_p\})$. Vì vậy. ta có $F_s(\{a_i\}) \subseteq F_s(\{c_p\})$ và $a_j \in F_s(\{c_p\})$, $a_j \neq c_p$ và $F_s(\{a_i\}) \neq F_s(\{c_p\})$. Do vậy $a_j \notin K''$, và K'' không chứa L. Do đó sẽ tồn tại $b_1, ... b_k$ với $b_i \in L(a_i)$, $i = \overline{1,k}$ để $K'' = \{b_1, ..., b_k\}$ với $b_i \in L(a_i)$, $i = \overline{1,k}$. Định lý được chứng minh.

Nhận xét 3.7: Các bước1,2,3 của thuật toán trên chỉ đòi hỏi thời gian là đa thức đối với Ω và chjiều dài của các tập phụ thuộc mạnh S. ở đây $n = |\Omega|$ và $p = l(S)$. Do vậy. bài toán tìm tập tất cả các khóa tối tiểu của S được giải quyết bằng một thuật toán có độ phức tạp là đa thức nếu lược lượng củaK_s là đa thức.

Tài liệu tham khảo

1. Armstrong W.W., *Dependency structures of data base relationships*, Information Processing 74, Holland Publ. Co. 1974, p. 580-583.
2. Czédli G., *On dependencies in the relational model of data*, EIK v. 17, 1981, p. 103-112.
3. Demetrovics J., *Relácios adatmodell logikai es structurális vizsgálata* MTA-SZTAKI Tanulmányok, Budapest, v. 114, 1980, p. 1-97.
4. Demetrovics J. & Gyepesi G., *On the functional dependency and some generalizations of it*, Acta Cybernetica Hungary V. 3, 1983, p. 295-305.
5. Demetrovics J & Vu Duc Thi, *Relations and minimal keys*, Acta Cybernetica v. 8, 1988, p. 279-285.
6. Vu Duc Thi, *Strong dependencies and s-semilattices*, Acta Cybernetica v. 8, 1987, p. 175-202.

Abstract

Some results on strong dependencies

Strong dependencies have been introduced and axiomatized in [2,3,4]. In this paper, we give some results, that are related to strong dependencies. We give an algorithm for computing closures of one attribute in lineast arme O(p) where p is the length of S, and then we give a membership algorithm. We also give an algorithm for finding a minimal key and finding a set of all minimal keys of S.