Tạp chí Tin học và Điều khiển học, T.27, S.3(2011), 199–205

THUẬT TOÁN TÌM TẤT CẢ CÁC RÚT GỌN TRONG BẢNG QUYẾT ĐỊNH*

NGUYỄN LONG GIANG, VŨ ĐỨC THI

Viện Công nghệ thông tin, Viện Khoa học và Công nghệ Việt Nam

Tóm tắt. Rút gọn thuộc tính là bài toán quan trọng trong lý thuyết tập thô. Cho đến nay, nhiều bài báo khoa học về các thuật toán rút gọn thuộc tính đã được đề xuất. Tuy nhiên, các thuật toán này đều tìm một tập rút gọn tốt nhất theo một tiêu chí đánh giá nào đó. Bài báo đề xuất một thuật toán mới tìm tất cả các tập rút gọn trong bảng quyết định và đánh giá thuật toán này có độ phức tạp thời gian là hàm mũ. Tuy vậy, trong nhiều trường hợp cụ thể, thuật toán này có độ phức tạp là đa thức.

Abstract. Attribute reduction is one of the most important issues in rough set theory. There have been many scientific papers that suppose algorithms on attribute reduction. However, these algorithms are all heuristic which find the best attribute reduction based on a kind of heuristic information. In this paper, we present a new algorithm for finding all attribute reductions of a decision and we show that the time complexity of the algorithm is exponential in the number of attributes. We also show that this complexity is polynomial in many special cases.

1. MỞ ĐẦU

Rút gọn thuộc tính trong bảng quyết định là quá trình loại bỏ các thuộc tính dư thừa trong tập thuộc tính điều kiện mà không ảnh hưởng đến việc phân lớp các đối tượng. Dựa vào tập rút gọn thu được, việc sinh luật và phân lớp đạt hiệu quả cao nhất. Cho đến nay, có rất nhiều công trình nghiên cứu về các thuật toán rút gọn thuộc tính trong lý thuyết tập thô. Tuy nhiên, các thuật toán này đều tìm được một tập rút gọn tốt nhất theo một tiêu chí đánh giá nào đó với độ phức tạp đa thức (các thuật toán theo hướng tiếp cận heuristic) mà chưa giải quyết bài toán tìm tất cả các tập rút gọn.

Với bảng quyết định nhất quán $DT = (U, C \cup d, V, f)$ trong lý thuyết tập thô, theo định nghĩa của Pawlak nếu $B \subseteq C$ là một rút gọn của C nếu B là tập tối thiểu thỏa mãn phụ thuộc hàm $B \to d$. Với quan hệ r trên tập thuộc tính R trong lý thuyết cơ sở dữ liệu, B là một tập tối thiểu của thuộc tính $a \in R$ trên r nếu B là tập thuộc tính nhỏ nhất thỏa mãn phụ thuộc hàm $B \to a$. Do đó, nếu xem bảng quyết định $DT = (U, C \cup d, V, f)$ là quan hệ r trên tập thuộc tính $R = C \cup d$ thì khái niệm tập rút gọn tương đương với khái niệm tập tối thiểu của thuộc tính $\{d\}$. Vì vậy, bài toán tìm tất cả các rút gọn trong lý thuyết tập thô trở thành bài toán tìm họ tất cả các tập tối thiểu của một thuộc tính trên quan hệ và bài toán này được giải quyết dựa trên các kết quả đã chứng minh trong lý thuyết cơ sở dữ liệu quan hệ.

*Nghiên cứu được hoàn thành dưới sự hỗ trợ từ Quỹ phát triển KHCNQG NAFOSTED, dự án số 102.01-2010.09

Bài báo này xây dựng một thuật toán tìm tất cả các tập rút gọn trong bảng quyết định dựa trên các kết quả đã công bố của giáo sư J. Demetrovics và Vũ Đức Thi trong cơ sở dữ liệu quan hệ và chứng minh độ phức tạp của thuật toán trong trường hợp xấu nhất là hàm mũ theo số thuộc tính điều kiện, tuy nhiên bài báo cũng chỉ ra trong một số trường hợp đặc biệt, độ phức tạp của thuật toán là đa thức theo kích thước của bảng quyết định. Phần còn lại của bài báo gồm: Mục 2 trình bày một số khái niệm cơ bản trong cơ sở dữ liệu quan hệ và trong lý thuyết tập thô; Mục 3 trình bày một số thuật toán cơ bản trong cơ sở dữ liệu quan hệ trong [5, 10]; Mục 4 đề xuất thuật toán tìm tất cả các rút gọn trong bảng quyết định và ví dụ minh họa thuật toán; Cuối cùng là kết luận.

2. CÁC KHÁI NIỆM CƠ BẢN

2.1. Các khái niệm cơ bản trong cơ sở dữ liệu quan hệ

Phần này sẽ trình bày một số khái niệm cơ bản trong lý thuyết cơ sở dữ liệu quan hệ. Các khái niệm này có thể xem trong [1, 3, 4, 6, 7, 10].

Cho $R = \{a_1, ..., a_n\}$ là tập hữu hạn khác rỗng các thuộc tính, mỗi thuộc tính a_i có miền giá trị là $D(a_i)$. Quan hệ r trên R là tập các bộ $\{h_1, ..., h_m\}$ với $h_j : R \to \bigcup_{a_i \in R} D(a_i), 1 \le j \le m$ là một hàm sao cho $h_j(a_i) \in D(a_i)$.

Cho $r = \{h_1, ..., h_m\}$ là một quan hệ trên $R = \{a_1, ..., a_n\}$. Phụ thuộc hàm (PTH) trên R là một dãy ký tự có dạng $A \to B$ với $A, B \subseteq R$. PTH $A \to B$ thỏa mãn quan hệ r trên R nếu $(\forall h_i, h_j \in r)\left((\forall a \in A)(h_i(a) = h_j(a)) \Rightarrow (\forall b \in B)(h_i(b) = h_j(b))\right)$. Đặt $F_r = \{(A, B) : A, B \subseteq R, A \to B\}$ là họ đầy đủ các PTH thỏa mãn quan hệ r. Gọi $P(R)$ là tập các tập con của R, nếu $F = P(R) \times P(R)$ thỏa mãn:

(1) $(A, A) \in F$

(2) $(A, B) \in F, (B, C) \in F \Rightarrow (A, C) \in F$

(3) $(A, B) \in F, A \subseteq C, D \subseteq B \Rightarrow (C, D) \in F$

(4) $(A, B) \in F, (C, D) \in F \Rightarrow (A \cup C, B \cup D) \in F$

thì F được gọi là một họ f trên R. Rõ ràng F_r là một họ f trên R. Theo [1] nếu F là một họ f trên R thì có một quan hệ r trên R sao cho $F_r = F$. Ký hiệu F^+ là tập tất cả các PTH được dẫn xuất từ F bằng việc áp dụng các quy tắc (1)-(4).

Một sơ đồ quan hệ (SĐQH) s là một cặp $< R, F >$ với R là tập thuộc tính và F là tập các phụ thuộc hàm trên R. Ký hiệu $A^+ = \{a : A \to \{a\} \in F^+\}$, A^+ được gọi là bao đóng của A trên s. Dễ thấy $A \to B \in F^+$ khi và chỉ khi $B \subseteq A^+$. Theo [1], nếu $s =< R, F >$ là sơ đồ quan hệ thì có quan hệ r trên R sao cho $F_r = F^+$, quan hệ r như vậy gọi là quan hệ Armstrong của s.

Cho r là một quan hệ, $s =< R, F >$ là một SĐQH và $A \subseteq R$. Khi đó A là một khóa của r (một khóa của s) nếu $A \to R\,(A \to R \in F^+)$. A là một khóa tối thiểu của $r(s)$ nếu A là một khóa của $r(s)$ và bất kỳ một tập con thực sự nào của A không phải là khóa của $r(s)$. Ký hiệu $K_r\,(K_s)$ là tập tất cả các khóa tối thiểu của $r(s)$. Gọi $K \subseteq P(R)$ là một hệ Sperner trên

R nếu với mọi $A, B \in K$ kéo theo $A \not\subset B$, dễ thấy $K_r\,(K_s)$ là các hệ Sperner trên R. Cho K là một hệ Sperner trên R đóng vai trò là tập tất cả các khóa tối thiểu. Ta định nghĩa tập các phản khóa của K, ký hiệu là K^{-1}, như sau:

$$K^{-1} = \{A \subset R : (B \in K) \Rightarrow (B \not\subset A)\} \text{ và nếu } (A \subset C) \Rightarrow (\exists B \in K)\,(B \subseteq C).$$

Dễ thấy K^{-1} cũng là một hệ Sperner trên R. Theo định nghĩa, nếu K là tập các khóa tối thiểu của một SĐQH nào đó thì K^{-1} là tập tất cả các tập không phải khóa lớn nhất.

Cho r là một quan hệ trên R. Đặt $E_r = \{E_{\text{ij}} : 1 \le i < j \le |r|\}$ với $E_{\text{ij}} = \{a \in R : h_i\,(a) = h_j\,(a)\}$, khi đó E_r được gọi là hệ bằng nhau của r. Theo [4], nếu $A_r \subseteq R$ thì $A_r^+ = \cap E_{\text{ij}}$ nếu tồn tại $E_{\text{ij}} \in E_r : A \subseteq E_{\text{ij}}$ và $A_r^+ = R$ trong trường hợp ngược lại. Tiếp theo, ta đưa ra định nghĩa họ các tập tối thiểu của một thuộc tính trên quan hệ và SĐQH.

Định nghĩa 2.1. [6]. Cho $s = (R, F)$ là SĐQH trên R và $a \in R$.

Đặt $K_a^s = \{A \subseteq R : A \to \{a\}, \exists B \subseteq R : (B \to \{a\})\,(B \not\subset A)\}$.

K_a^s được gọi là họ các tập tối thiểu của thuộc tính a trên SĐQH.

Tương tự, ta định nghĩa họ các tập tối thiểu của một thuộc tính trên quan hệ.

Định nghĩa 2.2. Cho r là một quan hệ trên R và $a \in R$.

Đặt $K_a^r = \{A \subseteq R : A \to \{a\}, \exists B \subseteq R : (B \to \{a\})\,(B \not\subset A)\}$.

K_a^r được gọi là họ các tập tối thiểu của thuộc tính a trên quan hệ r.

Rõ ràng $R \notin K_a^s, R \notin K_a^r, \{a\} \in K_a^s, \{a\} \in K_a^r$ và K_a^s, K_a^r là các hệ Sperner trên R.

2.2. Các khái niệm cơ bản trong lý thuyết tập thô

Trong phần này sẽ trình bày một số khái niệm cơ bản trong lý thuyết tập thô [9].

Bảng quyết định là một bộ tứ $DT = (U, C \cup D, V, f)$ trong đó $U = \{u_1, u_2, ..., u_n\}$ là tập khác rỗng, hữu hạn các đối tượng; $C = \{c_1, c_2, ..., c_m\}$ là tập các thuộc tính điều kiện; D là tập các thuộc tính quyết định với $C \cap D = \oslash$. $V = \prod\limits_{a \in C \cup D} V_a$ với V_a là tập giá trị của thuộc tính $a \in A$; $f : U \times (C \cup D) \to V$ là hàm thông tin, với $\forall a \in C \cup D, u \in U$ hàm f cho giá trị $f\,(u, a) \in V_a$. Không mất tính chất tổng quát giả thiết D chỉ gồm một thuộc tính quyết định d duy nhất (trường hợp D có nhiều thuộc tính thì bằng một phép mã hóa có thể quy về một thuộc tính [8]). Do đó, từ nay về sau ta xét bảng quyết định $DT = (U, C \cup d, V, f)$, trong đó $\{d\} \notin C$.

Mỗi tập con $P \subseteq C \cup \{d\}$ xác định một quan hệ không phân biệt được, gọi là quan hệ tương đương:

$$IND\,(P) = \{(x, y) \in U \times U\,|\forall a \in P, f\,(x, a) = f\,(y, a)\}.$$

$IND(P)$ xác định một phân hoạch của U, ký hiệu là $U/P = \{P_1, P_2, ..., P_m\}$. Một phần tử trong U/P gọi là một lớp tương đương.

Với $B \subseteq C$ và $X \subseteq U$, $B-$xấp xỉ trên của X là tập $\overline{B}X = \{u \in U\,|[u]_B \cap X \neq \emptyset\}$, $B-$xấp xỉ dưới của X là tập $\underline{B}X = \{u \in U\,|[u]_B \subseteq X\}$, $B-$miền biên của X là tập $BN_B\,(X) = \overline{B}X \backslash \underline{B}X$ và $B-$miền dương của $\{d\}$ là tập $POS_B\,(\{d\}) = \bigcup\limits_{X \in U/D} (\underline{B}X)$. Bảng quyết định DT được gọi là nhất quán khi và chỉ khi $POSC(d) = U$, hay phụ thuộc hàm $C \to d$ đúng, ngược lại DT là không nhất quán. Trong trường hợp DT không nhất quán thì $POS_C(\{d\})$ chính là tập con cực đại của U sao cho phụ thuộc hàm $C \to d$ đúng.

Trong lý thuyết tập thô [9], Pawlak đưa ra khái niệm tập rút gọn của bảng quyết định, còn gọi là tập rút gọn dựa trên miền dương.

Định nghĩa 2.3. Cho bảng quyết định $DT = (U, C \cup d, V, f)$. Nếu $B \subseteq C$ thỏa mãn:

(1) $POS_B(\{d\}) = POS_C(\{d\})$

(2) $\forall B' \subset B \left(POS_{B'}(\{d\}) \neq POS_C(\{d\})\right)$

thì B được gọi là một tập rút gọn của C.

Trường hợp DT nhất quán, định nghĩa trên cho thấy B là một tập rút gọn nếu B thỏa mãn $B \to d$ và $\forall B' \subset B, B' \not\to \{d\}$, ký hiệu R_d là tập tất cả các rút gọn của DT.

3. MỘT SỐ THUẬT TOÁN CƠ BẢN TRONG CƠ SỞ DỮ LIỆU QUAN HỆ

3.1. Thuật toán tìm tập phản khóa

Thuật toán 3.1. [10] Tìm tập phản khóa K^{-1}

Đầu vào: $K = \{B_1, ..., B_n\}$ là hệ Sperner trên R.

Đầu ra: K^{-1}

Bước 1: Đặt $K_1 = \{R - \{a\} : a \in B_1\}$. Hiển nhiên $K_1 = \{B_1\}^{-1}$

Bước q+1: (q<m). Giả thiết rằng $K_q = F_q \cup \{X_1, ..., X_{tq}\}$, ở đây $X_1, ..., X_{tq}$ chứa $B_q + 1$ và $F_q = \{A \in K_q : B_{q+1} \not\subset A\}$. Đối với mỗi $i\,(i = 1, ..., t_q)$ ta tìm các phản khóa của $B_q + 1$ trên X_i tương tự như K_1. Ký pháp của chúng là $A_1^i, ..., A_{ri}^i$. Đặt:

$K_{q+1} = F_q \cup \left\{A_p^i : A \in F_q \Rightarrow A_p^i \not\subset A, 1 \leq i \leq t_q, 1 \leq p \leq r_i\right\}$

Cuối cùng ta đặt $K^{-1} = K_m$

Độ phức tạp thuật toán

Đặt $I_q\,(1 \leq q \leq m-1)$ là số phần tử của K_q trong thuật toán trên. Theo [10], độ phức tạp thời gian của thuật toán là $O\left(|R|^2 \sum\limits_{q=1}^{m-1} t_q u_q\right)$ với $u_q = I_q - t_q$ nếu $I_q > t_q$ và $u_q = 1$ nếu $I_q = t_q$.

Nhận xét 3.1

1) Trong mỗi bước của thuật toán, K_q là hệ Sperner trên R. Theo [5], kích thước của hệ Sperner bất kỳ trên R không vượt quá $C_n^{[n/2]} \approx 2^{n+1/2} / \left(\prod .n^{1/2}\right)$ với $n = |R|$. Do đó, độ phức tạp tồi nhất của thuật toán là hàm số mũ theo n.

2) Trường hợp $I_q \leq I_m\,(q = 1, ..., m-1)$, độ phức tạp của thuật toán không lớn hơn $O\left(|R|^2 |K| \left|K^{-1}\right|^2\right)$, khi đó độ phức tạp thuật toán là đa thức theo $|R|, |K|$ và $|K|^{-1}$. Nếu số lượng các phần tử của K là nhỏ thì thuật toán rất hiệu quả, đòi hỏi thời gian đa thức theo $|R|$.

3.2. Thuật toán tìm tập khóa tối thiểu từ tập các phản khóa

Thuật toán 3.2. [5] Tìm một khóa tối thiểu từ tập các phản khóa

Đầu vào: Cho K là hệ Sperner đóng vai trò là tập phản khóa, $C = \{b_1, ..., b_m\} \subseteq R$ và H là hệ Sperner đóng vai trò là tập khóa $\left(K = H^{-1}\right)$ sao cho $\exists B \in K : B \subseteq C$.

Đầu ra: $D \in H$

Bước 1: Đặt $T(0) = C$;

Bước i+1: Đặt $T(i+1) = T(i) - b_{i+1}$ nếu $\forall B \in K$ không có $T \subseteq B$; trong trường hợp ngược lại đặt $T(i+1) = T(i)$;

Cuối cùng ta đặt $D = T(m)$;

Thuật toán 3.3. [5] Tìm tập các khóa tối thiểu từ tập các phản khóa

Đầu vào: Cho $K = \{B_1, ..., B_k\}$ là hệ Sperner trên R.

Đầu ra: H mà $H^{-1} = K$.

Bước 1: Bởi Thuật toán 3.2 ta tính A_1, đặt $K_1 = A_1$.

Bước i+1: Nếu có $B \in K_i^{-1}$ sao cho $B \not\subset B_j \, (\forall j : 1 \leq j \leq k)$ thì bởi Thuật toán 3.2 ta tính $A_i + 1$, ở đây $A_{i+1} \in H, A_{i+1} \subseteq B$. Đặt $K_{i+1} = K_i \cup A_{i+1}$. Trong trường hợp ngược lại ta đặt $H = K_i$.

Độ phức tạp thuật toán 3.3

Theo [5], độ phức tạp của Thuật toán 3.3 là $O\left(|R|\left(\sum_{q=1}^{m-1}\left(|K|\, I_q + |R|\, t_q u_q\right) + |K|^2 + |R|\right)\right)$ với I_q, t_q, u_q như trong Thuật toán 3.1. Do đó, độ phức tạp tồi nhất của Thuật toán 3.3 là hàm số mũ theo n với n là số phần tử của R. Trường hợp $I_q \leq |K| \, (q = 1, ..., m-1)$, độ phức tạp của Thuật toán 3.3 là $O\left(|R|^2 |K|^2 \, |H|\right)$, độ phức tạp này là đa thức theo $|R|\,, |K|$ và $|H|$. Nếu $|H|$ là đa thức theo $|R|\,, |K|$ thì thuật toán hiệu quả. Nếu số lượng các phần tử của H là nhỏ thì thuật toán rất hiệu quả.

4. THUẬT TOÁN TÌM TẤT CẢ RÚT GỌN TRONG BẢNG QUYẾT ĐỊNH

Cho bảng quyết định nhất quán $DT = (U, C \cup d, V, f), R \subseteq C$ là tập rút gọn của DT nếu thỏa mãn $POS_R(\{d\}) = POS_C(\{d\}) = U$ hay $R \to d$, và $R' \not\subset R : POS_{R'}(\{d\}) = POS_C(\{d\}) = U$ hay $R' \not\subset R : R' \to \{d\}$. Từ Định nghĩa 2.2 và Định nghĩa 2.3, bài toán tìm tất cả các tập rút gọn của bảng quyết định nhất quán $DT = (U, C \cup d, V, f)$ với trở thành bài toán tìm họ các tập tối thiểu của thuộc tính $\{d\}$ mà không chứa d đối với quan hệ $r = \{u_1, u_2, ..., u_m\}$ trên tập thuộc tính $R = C \cup d$. Ký hiệu R_d là tập tất cả các rút gọn của DT, khi đó $R_d = K_d^r - \{d\}$ với K_d^r là họ các tập tối thiểu của thuộc tính $\{d\}$ trên quan hệ r.

Thuật toán 4.1. Thuật toán tìm tất cả các tập rút gọn trên bảng quyết định.

Đầu vào: Bảng quyết định $DT = (U, C \cup d, V, f)$ với $POS_C(\{d\}) = U, C = \{c_1, c_2, ..., c_k\}$, $U = \{u_1, u_2, ..., u_m\}$.

Đầu ra: R_d.

Xem bảng quyết định DT là quan hệ $r = \{u_1, u_2, ..., u_m\}$ trên tập thuộc tính $R = C \cup d$.

Bảng 1. Bảng quyết định

U	a	b	c	d
u_1	6	6	0	6
u_2	0	2	2	0
u_3	0	0	0	0
u_4	0	0	3	0
u_5	4	4	0	0
u_6	5	0	5	5
u_7	1	0	0	0

Bước 1: Từ r ta xây dựng hệ bằng nhau $E_r = \{E_{\mathrm{ij}} : 1 \le i < j \le m\}$

với $E_{\mathrm{ij}} = \{a \in R : u_i(a) = u_j(a)\}$.

Bước 2: Từ E_r ta xây dựng tập $M_d = \{A \in E_r : d \notin A, \nexists B \in E_r : d \notin B, A \subset B\}$.

Bước 3: Bởi Thuật toán 3.3, tính tập K từ tập $M_d\left(K^{-1} = M_d\right)$.

Bước 4: Đặt $R_d = K - \{d\}$.

Chứng minh tập K thu được từ Bước 3 là hay họ các tập tối thiểu của thuộc tính d trên quan hệ r.

Thật vậy, theo cách xây dựng M_d tại Bước 2 và theo công thức tính bao đóng của tập thuộc tính trên quan hệ, $\forall A \in N_d$ ta có $A^+ = A$ và A không chứa d nên A^+ không chứa d, suy ra $A \to \{d\} \notin F^+$. Mặt khác, nếu tồn tại B sao cho $A \subset B$ thì xảy ra hai trường hợp: (1) Nếu B không chứa d thì $B^+ = R$; (2) Nếu B chứa d thì hiển nhiên B^+ chứa d. Cả hai trường hợp ta đều có B^+ chứa d hay $B \to \{d\} \in F^+$. Do đó $M_d = MAX\left(F^+, d\right)$ với $MAX\left(F^+, d\right) = \{A \subseteq R : A \to \{d\} \notin F^+, A \subset B \Rightarrow B \to \{d\} \in F^+\}$. Theo [3, 6], $MAX\left(F^+, d\right) = \left(K_d^r\right)^{-1}$ nên $M_d = \left(K_d^r\right)^{-1}$ do đó tại Bước 3, $K = K_d^r$ là họ các tập tối thiểu của thuộc tính d trên quan hệ r. Tại Bước 4, $R_d = K - \{d\}$ thu được là tập tất cả các rút gọn của bảng quyết định.

Phân tích độ phức tạp thuật toán

Dễ thấy, độ phức tạp của Bước 1 và Bước 2 là đa thức theo kích thước của r. Do đó, độ phức tạp của thuật toán là độ phức tạp của Thuật toán 3.3 tính tập khóa tối thiểu từ tập phản khóa tại Bước 3. Do đó, độ phức tạp của thuật toán là:

$$O\left(|R|\left(\sum_{q=1}^{m-1}\left(|M_d|\, I_q + |R|\, t_q u_q\right) + |M_d|^2 + |R|\right)\right)$$

với I_q, t_q, u_q như trong Thuật toán 3.1 và độ phức tạp này trong trường hợp xấu nhất là hàm số mũ theo n với n là số phần tử của R. Trường hợp $I_q \le |M_d|\,(q = 1, ..., m-1)$, độ phức tạp của thuật toán là $O\left(|R|^2 |M_d|^2 \left|K_d^r\right|\right)$, độ phức tạp này là đa thức theo $|R|, |M_d|$ và $|K_d^r|$. Dễ thấy tại bước 2, $|M_d|$ là đa thức theo kích thước của r, do đó nếu $|K_d^r|$ là đa thức theo $|R|$ thì độ phức tạp của thuật toán là đa thức theo kích thước của r. Nếu số lượng các phần tử của K_d^r là nhỏ thì thuật toán rất hiệu quả.

Ví dụ 4.1: Cho bảng quyết định $DT = (U, C \cup d, V, f)$ với $U = \{u_1, u_2, u_3, u_4, u_5, u_6, u_7\}$, $C = \{a, b, c\}$ như sau:

Xem bảng quyết định như quan hệ $r = \{u_1, u_2, u_3, u_4, u_5, u_6, u_7\}$ trên tập thuộc tính $R = \{a, b, c, d\}$. áp dụng Thuật toán 4.1 ta tính được:

$E_r = \{\{a, b, d\}, \{b, c, d\}, \{a, d\}, \{b, d\}, \{c, d\}, \{b\}, \{c\}, \{d\}\}$

$M_d = \{\{b\}, \{c\}\}$

Bởi Thuật toán 3.3, ta tính được $K = K_d^r = \{\{a\}, \{b, c\}, \{d\}\}$

Như vậy, tập tất cả các rút gọn của bảng quyết định DT là $R_d = K_d^r - \{d\} = \{\{a\}, \{b, c\}\}$.

5. KẾT LUẬN

Bài báo đưa ra khái niệm tập tối thiểu của một thuộc tính trên quan hệ dựa trên khái niệm tập tối thiểu của một thuộc tính trên sơ đồ quan hệ trong [6] và xây dựng thuật toán tìm tất cả các tập rút gọn của tập thuộc tính điều kiện trong bảng quyết định dựa vào khái niệm khóa, phản khóa và thuật toán tìm tập tất cả các khóa, phản khóa trong cơ sở dữ liệu quan hệ [5, 10]. Trong trường hợp xấu nhất, độ phức tạp thời gian của thuật toán được xây dựng là hàm mũ theo số thuộc tính điều kiện. Trường hợp lực lượng tập các khóa tối thiểu thu được từ tập các phản khóa cho trước là đa thức theo n, với n là số thuộc tính điều kiện thì độ phức tạp thời gian của thuật toán là đa thức theo số hàng và số cột của bảng quyết định.

TÀI LIỆU THAM KHẢO

[1] W. W. Armstrong, Dependency structures of database relationships, *Information Processing 74*, North-Holland Publ. Co. 1974 (580-583).

[2] J. Demetrovics, On the equivalence of candidate keys with Sperner systems, *Acta Cybernetica* **4** (1979) 247-252.

[3] J. Demetrovics, V.D. Thi, Algorithms for generating Armstrong relation and inferring functional dependencies in the relational datamodel, *Computers and Mathematics with Applications* **26** (4) (1993) 43-55 (Great Britain).

[4] J. Demetrovics, V.D. Thi, Keys, antikeys and prime attributes, *Ann. Univ. Scien. Budapest Sect. Comput.* **8** (1987) 37-54.

[5] J. Demetrovics, V.D. Thi, Relations and minimal keys, *Acta Cybernetica* **8** (3) (1998) 279-285.

[6] J. Demetrovics, V.D. Thi, Some remarks on generating Armstrong and inferring functional dependencies relation, *Acta Cybernetica* **12** (1995) 167-180.

[7] J. Demetrovics, V.D. Thi, Some results about functional dependencies, *Acta Cybernetica* **8** (3) (1988) 273-278.

[8] M. Kryszkiewicz, Rough set approach to incomplete information systems, *Information Science* (1998) 39-49.

[9] Z. Pawlak, *Rough Sets - Theoritical Aspects of Reasoning about Data*, Kluwer Academic Publishers, Dordrecht, 1991.

[10] V.D. Thi, Minimal keys and Antikeys, *Acta Cybernetica* **7** (4) (1986) 361-371.

Ngày nhận bài 8 - 6 - 2011