Tạp chí Tin học và Điều khiển học, T. 15, S. 2 (1999), 47-53

MỞ RỘNG CÁC PHÉP TOÁN QUAN HỆ CHO NULL PHỤ THUỘC NGỮ CẢNH

BÙI THỊ THÚY HIỀN

Abstract. In this paper, we apply the concept of model introduced in [2] to define extended relational algebra operators. It is proved that these extensions have properties which are formally stated as being adequate and precise.

1. MỞ ĐẦU

Khi nghiên cứu về quan hệ không chứa giá trị null, chúng ta đã làm quen với đại số quan hệ. Để áp dụng các phép toán đại số quan hệ thông thường trên các quan hệ chứa null, cần định nghĩa lại chúng trên các miền thuộc tính mở rộng. Việc định nghĩa lại đó được gọi là mở rộng các phép toán. Để có ý nghĩa, các phép toán được mở rộng phải thỏa mãn một số tiêu chuẩn được đề ra. Ví dụ, việc định nghĩa lại phép chọn theo cách mà kết quả luôn là quan hệ rỗng mỗi khi nó được áp dụng cho một quan hệ có chứa null không phải là một giải pháp mong muốn và ràng buộc trước tiên mà người ta muốn là các phép toán được mở rộng phải cho kết quả như ban đầu nếu được áp dụng trên những quan hệ không chứa null.

Trong bài báo này, chúng tôi mở rộng một vài phép toán đại số quan hệ trên những quan hệ null phụ thuộc ngữ cảnh, chứng minh các phép toán mở rộng đó thỏa mãn một số tiêu chuẩn được đề ra trong [10].

2. NỘI DUNG THÔNG TIN

2.1. Những kí hiệu và khái niệm cơ sở

Cho $R(A_1, \ldots, A_n)$ là một lược đồ quan hệ được xác định trên một tập thuộc tính $A_1, \ldots, A_n$.

Với mỗi thuộc tính A_i, ta kí hiệu miền giá trị tương ứng là $\mathrm{Dom}(A_i)$. Miền của R là tích Đề các $\mathrm{Dom}(A_1) \times \mathrm{Dom}(A_2) \times \cdots \times \mathrm{Dom}(A_n)$ và kí hiệu là $\mathrm{Dom}(R)$.

Chúng ta mở rộng mỗi miền $\mathrm{Dom}(A_i)$ thành $\mathrm{Dom}^*(A_i)$ bằng cách thêm vào một tập hữu hạn các kí hiệu null: $\mathrm{Dom}^*(A_i) = \mathrm{Dom}(A_i) \cup \Delta_{i_1} \cup \Delta_{i_2} \cup \{\mathrm{dne}\}$, trong đó.

- Δ_{i_1} là tập các null phụ thuộc ngữ cảnh chưa biết và được kí hiệu bởi các kí tự Hy lạp như δ_1, δ_2,...
- Δ_{i_2} là tập các null phụ thuộc ngữ cảnh mở và được kí hiệu bởi β_1, β_2,...
- dne là kí hiệu cho null không tồn tại,
- $\mathrm{Dom}(A_i)$, Δ_{i_1}, Δ_{i_2}, {dne} là các tập không giao nhau.

Tương tự một sự mở rộng của $\mathrm{Dom}(R)$ là $\mathrm{Dom}^*(R) = \mathrm{Dom}^*(A_1) \times \cdots \times \mathrm{Dom}^*(A_n)$.

Một thể hiện (hay một quan hệ) có chứa null phụ thuộc ngữ cảnh của một lược đồ R là một tập con của $\mathrm{Dom}^*(R)$. Những quan hệ như vậy được kí hiệu bởi các kí tự thường như $r, r_1, \ldots$ và được gọi là các quan hệ một phần. Tập tất cả các quan hệ một phần trên R được kí hiệu là $\mathrm{Rel}_{\dagger}(R)$.

Các quan hệ không chứa null được gọi là quan hệ toàn phần, tập tất cả các quan hệ toàn phần trên lược đồ R được ký hiệu là $\mathrm{Rel}(R)$.

Một bộ của quan hệ r là một phần tử của r. Chúng ta kí hiệu những bộ của r bởi những kí tự như t, t', s, s',.... Nếu t là bộ của một quan hệ r thì $t[A_i]$ là giá trị của t tại thuộc tính A_i. Nếu $t[A_i]$ khác null ta viết $t[A_i]!$.

Kí hiệu open để chỉ một null phụ thuộc ngữ cảnh mở và unk để chỉ một null phụ thuộc ngữ cảnh chưa biết.

2.2. Nội dung thông tin của quan hệ một phần

Một câu hỏi được đặt ra: làm thế nào để định nghĩa nội dung thông tin của quan hệ một phần, so sánh nội dung thông tin giữa chúng với nhau.

Do quan hệ một phần sẽ trở thành toàn phần khi tất cả các thông tin trong quan hệ đều đầy đủ nên người ta có thể coi quan hệ một phần là biểu thị cho một quan hệ toàn phần nào đó. Nhưng vì thông tin trong quan hệ một phần là chưa đầy đủ nên không thể biết chính xác nó biểu thị cho quan hệ toàn phần nào. Vì vậy, với mỗi quan hệ một phần r, người ta xem xét tập các quan hệ toàn phần mà có khả năng biểu thị. Tập các quan hệ toàn phần đó được gọi là tập *khả năng* hay tập *mô hình* của quan hệ r và nội dung thông tin của một quan hệ r có thể được coi là tập khả năng mà nó biểu thị.

Mối liên hệ giữa một quan hệ một phần và tập các khả năng của nó có thể được mô tả bởi một ánh xạ biểu thị. Maier [10] gọi ánh xạ biểu thị là POSS, Lipski và đồng tác giả [8] gọi là Rep, Gottlob-Zicari [6] gọi là MODELS. Bài báo này sử dụng cách gọi của Maier.

Chúng ta phát biểu lại định nghĩa của Maier về hàm khả năng POSS.

Định nghĩa 1 (Hàm khả năng [10]). Một hàm khả năng POSS là một ánh xạ: $\mathrm{Rel}_\dagger \to 2^{\mathrm{Rel}}$, sao cho $\forall r \in \mathrm{Rel}_\dagger(R)$ thì $\mathrm{POSS}(r) \subseteq \mathrm{Rel}(R)$.

Để so sánh nội dung thông tin giữa các quan hệ một phần, người ta có nhận xét là nội dung thông tin của một giá trị không xác định luôn ít hơn một giá trị xác định. Ví dụ, một giá trị null "chưa biết" chỉ cung cấp một thông tin là có tồn tại một giá trị nào đó mà sau này có thể được đưa vào trong quan hệ, trong khi một giá trị xác định chỉ ra ngay nó là giá trị gì. Như vậy, nội dung thông tin của một quan hệ r là nhiều hơn s nếu như số lượng mô hình của r là bằng hoặc ít hơn số lượng mô hình của s.

Định nghĩa 2. Cho một hàm khả năng POSS, r và s là hai quan hệ thuộc $\mathrm{Rel}_\dagger(R)$, r được gọi là có nội dung thông tin nhiều hơn s (viết $r \succeq s$ hoặc $s \preceq r$) nếu và chỉ nếu $\mathrm{POSS}(r) \subseteq \mathrm{POSS}(s)$.

Nếu $\mathrm{POSS}(r) \subseteq \mathrm{POSS}(s)$ và $\mathrm{POSS}(s) \subseteq \mathrm{POSS}(r)$ thì kí hiệu $r \equiv_{\mathrm{POSS}} s$.

Nếu $\mathrm{POSS}(r) \subseteq \mathrm{POSS}(s)$ và $\mathrm{POSS}(s) \neq \mathrm{POSS}(r)$ ta kí hiệu $r \succ s$ hay $s \prec r$.

3. MỞ RỘNG CÁC PHÉP TOÁN QUAN HỆ CHO NULL PHỤ THUỘC NGỮ CẢNH

3.1. Những định nghĩa cơ sở

Trong [10] Maier đã đưa ra một số tiêu chuẩn nên đạt được cho các phép toán quan hệ mở rộng.

Định nghĩa 3. Cho γ là một phép toán trên Rel, γ' là một phép toán trên $\mathrm{Rel}_\dagger$; γ' được gọi là một mở rộng chính xác (precise) của γ tương ứng với hàm khả năng POSS nếu như

1. γ là γ' là các phép toán một ngôi thì $\mathrm{POSS}(\gamma'(r)) = \gamma(\mathrm{POSS}(r))\ \forall r \in \mathrm{Rel}_\dagger$, hoặc
2. γ và γ' là các phép toán hai ngôi thì $\mathrm{POSS}(r\gamma' s) = \mathrm{POSS}(r)\,\gamma\,(\mathrm{POSS})(s)\ \forall r, s \in \mathrm{Rel}_\dagger$.

Trong đó, với hai tập quan hệ toàn phần P_1 và P_2 thì

$\gamma(P_1) = \{\gamma(q) \mid q \in P_1\}$ và

$P_1 \gamma P_2 = \{q_1 \gamma q_2 \mid q_1 \in P_1, q_2 \in P_2\}$.

Tiêu chuẩn *chính xác* thường khó đạt được đối với một hàm POSS cho trước vì vậy Maier đã đưa ra hai tiêu chuẩn ít ngặt nghèo hơn.

Định nghĩa 4. Cho γ là một phép toán trên Rel, γ' là một phép toán trên $\mathrm{Rel}_\dagger$, γ' được gọi là một mở rộng thỏa đáng (adequate) của γ tương ứng với hàm khả năng POSS nếu như

1. γ và γ' là các phép toán một ngôi thì $\mathrm{POSS}(\gamma'(r)) \supseteq \gamma(\mathrm{POSS}(r))\ \forall r \in \mathrm{Rel}_\dagger$, hoặc
2. γ và γ' là các phép toán hai ngôi thì $\mathrm{POSS}(r\gamma' s) \supseteq \mathrm{POSS}(r)\gamma(\mathrm{POSS}(s)\ \forall r, s \in \mathrm{Rel}_\dagger$.

Định nghĩa 5. Cho γ là một phép toán trên Rel, γ' là một phép toán trên $\text{Rel}_\dagger$, γ' được gọi là một mở rộng có giới hạn (restrited) của γ tương ứng với hàm khả năng POSS nếu như

1. γ và γ' là các phép toán một ngôi thì $\forall r \in \text{Rel}_\dagger$, $\nexists q \in \text{Rel}_\dagger$ sao cho $\text{POSS}(\gamma'(r)) \supset \text{POSS}(q) \supseteq \gamma(\text{POSS}(r))$, hoặc

2. γ và γ' là các phép toán hai ngôi thì $\forall r, s \in \text{Rel}_\dagger$, $\nexists q \in \text{Rel}_\dagger$ sao cho $\text{POSS}(r\gamma' s) \supset \text{POSS}(q) \supseteq \text{POSS}(r)\gamma(\text{POSS}(s))$.

3.2. Quan hệ một phần được phân hoạch

Trong [4] Biskup đã phân các bộ trong một quan hệ chứa null gồm hai loại, những *bộ chắc chắn* và những *bộ có thể*. Những quan hệ được xem xét theo kiểu của Biskup được Maier [10] gọi là những quan hệ được phân hoạch. Một quan hệ được phân hoạch có thể xem như một cặp được sắp của hai quan hệ một phần không giao nhau.

Định nghĩa 6. Cho r_1 và r_2 là hai quan hệ trên cùng một lược đồ R. Chúng ta gọi cặp được sắp (r_1, r_2) là một quan hệ được phân hoạch r nếu $r_1 \cap r_2 = \emptyset$ và $\text{SURE}(r) = r_1$, $\text{MAYBE}(r) = r_2$. Trong đó $\text{SURE}(r)$ được gọi là tập các bộ chắc chắn của r và $\text{MAYBE}(r)$ là tập các bộ có thể của r.

Định nghĩa 7. Cho r là một quan hệ được phân hoạch trên lược đồ R, s là quan hệ một phần trên R. Ta nói s xấp xỉ r (kí hiệu $s \triangleright r$) nếu và chỉ nếu $\text{SURE}(r) \cup \text{MAYBE}(r) \supseteq s \supseteq \text{SURE}(r)$.

Khi biểu diễn một quan hệ được phân hoạch, theo Maier chúng ta cũng dùng một đường đứt nét để mô tả những bộ chắc chắn ở trên và những bộ có thể ở dưới.

Ví dụ 1. Cho r là quan hệ trong bảng 1, s_1 và s_2 là các quan hệ trong bảng 2 và bảng 3. Khi đó $s_1 \triangleright r$ và $s_2 \triangleright r$.

r	A	B	C
	1	δ_1	4
	β_1	5	6
	- - -	- - -	- - -
	2	dne	4
	1	3	6

bảng 1

s_1	A	B	C
	1	δ_1	4
	β_1	5	6
	1	3	6

bảng 2

s_2	A	B	C
	1	δ_1	4
	β_1	5	6

bảng 3

3.3. Hàm khả năng POSS_{NC}

Trong bài báo này chúng ta xem xét quan hệ chứa null phụ thuộc ngữ cảnh dưới dạng được phân hoạch thành các bộ chắn chắn và các bộ có thể và định nghĩa hàm khả năng $\text{POSS}_{\text{NC}}(r) = \{\bar{s} \mid \bar{s}$ là mô hình yếu của s nào đó mà $s \triangleright r\}$.

Mệnh đề 1. *Với quan hệ được phân hoạch r và hàm khả năng $\text{POSS}_{\text{NC}}(r) = \{\bar{s} \mid \bar{s}$ là mô hình yếu của s nào đó mà $s \triangleright r\}$ thì $\text{POSS}_{\text{NC}}(r)$ là xác định duy nhất.*

Chứng minh. Rõ ràng, theo định nghĩa hàm khả năng POSS_{NC}, với mọi q xấp xỉ r ($q \triangleright r$) thì $\bar{q} \in \text{POSS}_{\text{NC}}(r)$. Điều này chứng tỏ $\text{POSS}_{\text{NC}}(r)$ chỉ phụ thuộc vào r mà không phụ thuộc vào bất kì một quan hệ s nào mà s xấp xỉ r. Như vậy $\text{POSS}_{\text{NC}}(r)$ là xác định duy nhất.

Mệnh đề 2. *Cho r và s là các quan hệ được phân hoạch trên lược đồ R và hàm khả năng POSS_{NC}, nếu tồn tại các null phụ thuộc ngữ cảnh $a_1, a_2, \ldots, a_k$ trong s và tồn tại các phép thế có thể V_i của a_i $\forall 1 \le i \le k$ sao cho*

i) $\bar{s} = S^{V_1V_2\ldots V_k}_{a_1a_2\ldots a_k}(s)$,

ii) $\forall t_{\bar{s}} \in \text{SURE}(\bar{s})\ \exists t_r \in \text{SURE}(r) : t_r = t_{\bar{s}}$, *iii)* $\forall t_r \in r\ \exists t_{\bar{s}} \in \bar{s} : t_r = t_{\bar{s}}$,

thì $r \succeq s$.

Chứng minh. Để chứng minh $\mathrm{POSS_{NC}}(r) \subseteq \mathrm{POSS_{NC}}(s)$ ta sẽ chỉ ra nếu ba điều kiện trên thỏa mãn thì (*) $\mathrm{POSS_{NC}}(r) \subseteq \mathrm{POSS_{NC}}(\bar{s})$ và (**) $\mathrm{POSS_{NC}}(\bar{s}) \subseteq \mathrm{POSS_{NC}}(s)$. Đầu tiên, từ các tính chất (ii) và (iii) ta suy ra $\mathrm{SURE}(\bar{s}) \subseteq \mathrm{SURE}(r)$ và $r \subseteq \bar{s}$.

(*) Gọi $\hat{r}$ là một quan hệ trong $\mathrm{POSS_{NC}}(r)$, khi đó $\hat{r}$ phải là mô hình yếu của một quan hệ r' nào đó sao cho $r' \triangleright r$. Do $r' \triangleright r$ nên $r' = \{\mathrm{SURE}(r) \cup \mathrm{MB_1}(r)\}$ với $\mathrm{MB_1}(r) \subseteq \mathrm{MAYBE}(r)$. Vì $r' \subseteq r$, $r \subseteq \bar{s}$ và $\mathrm{SURE}(\bar{s}) \subseteq \mathrm{SURE}(r)$ nên $r' = \{\mathrm{SURE}(\bar{s}) \cup \mathrm{MB_2}(\bar{s})\}$ với $\mathrm{MB_2}(\bar{s}) \subseteq \mathrm{MAYBE}(\bar{s})$. Vậy $r' \triangleright \bar{s}$. Do đó $\hat{r} \in \mathrm{POSS_{NC}}(\bar{s})$.

(**) Gọi $\hat{r}$ là một quan hệ trong $\mathrm{POSS_{NC}}(\bar{s})$. Khi đó $\hat{r}$ phải là mô hình yếu của một quan hệ r' nào đó mà sao cho $r' \triangleright \bar{s}$. Do $r' \triangleright \bar{s}$ nên $r' = \{\mathrm{SURE}(\bar{s}) \cup \mathrm{MB}(\bar{s})\}$ với $\mathrm{MB}(\bar{s}) \subseteq \mathrm{MAYBE}(\bar{s})$. Do $\bar{s} = S^{V_1V_2\ldots V_k}_{a_1a_2\ldots a_k}(s)$ nên $r' = \{\mathrm{SURE}(S^{V_1V_2\ldots V_k}_{a_1a_2\ldots a_k}(s)) \cup \mathrm{MB}(S^{V_1V_2\ldots V_k}_{a_1a_2\ldots a_k}(s))\} = \{S^{V_1V_2\ldots V_k}_{a_1a_2\ldots a_k}(\mathrm{SURE}(s)) \cup S^{V_1V_2\ldots V_k}_{a_1a_2\ldots a_k}(\mathrm{MB}(s))\} = S^{V_1V_2\ldots V_k}_{a_1a_2\ldots a_k}(\mathrm{SURE}(s) \cup \mathrm{MB}(s)) = S^{V_1V_2\ldots V_k}_{a_1a_2\ldots a_k}(s')$ với $s' \triangleright s$. Theo định nghĩa của mô hình yếu thì tồn tại $b_1, \ldots, b_m$ và $Y_1, \ldots Y_m$ sao cho $\hat{r} = S^{Y_1\ldots Y_m}_{b_1\ldots b_m}(r')$. Vậy $\hat{r} = S^{Y_1\ldots Y_m}_{b_1\ldots b_m} S^{V_1V_2\ldots V_k}_{a_1a_2\ldots a_k}(s')$. Do $s' \triangleright s$ nên $\hat{r} \in \mathrm{POSS_{NC}}(s)$.

Kết hợp (*) và (**) ta có $r \succeq s$.

Mệnh đề 3. *Cho* r *và* s *là các quan hệ được phân hoạch trên lược đồ* R *và hàm khả năng* $\mathrm{POSS_{NC}}$, *nếu* $\mathrm{POSS_{NC}}(r) \subseteq \mathrm{POSS_{NC}}(s)$ *thì*

i) $\forall t_s \in \mathrm{SURE}(s)\ \exists t_r \in \mathrm{SURE}(r) : t_s \le t_r$,

ii) $\forall t_r \in r\ \exists t_s \in s : t_s \le t_r$.

Chứng minh.

i) Giả sử ngược lại, $\exists t_s \in \mathrm{SURE}(s)\ \forall t_r \in \mathrm{SURE}(r) : t_s \not\le t_r$, ta xây dựng mô hình $\hat{r}$ của $\mathrm{SURE}(r)$ sao cho $\forall t_{\hat{r}} \in \hat{r}$, $t_s \not\le t_{\hat{r}}$ khi đó $\hat{r}$ không thể là mô hình của quan hệ s' nào mà có chứa t_s, vậy $\hat{r} \notin \mathrm{POSS_{NC}}(s)$ nhưng $\hat{r} \in \mathrm{POSS_{NC}}(r)$, mâu thuẫn với giả thiết. Ta có (i).

ii) Ta cũng giả sử ngược lại, $\exists t_r \in r\ \forall t_s \in s : t_s \not\le t_r$. Ta xây dựng mô hình $\hat{r}$ của r sao cho $\forall t_s \in s : t_s \not\le t_{\hat{r}}$, rõ ràng $\hat{r}$ không thể là mô hình của bất kì quan hệ $s' \triangleright s$, tức là $\hat{r} \notin \mathrm{POSS_{NC}}(s)$ nhưng $\hat{r} \in \mathrm{POSS_{NC}}(r)$, mâu thuẫn với giả thiết. Vậy ta có (ii).

3.4. Mở rộng các phép toán quan hệ cho null phụ thuộc ngữ cảnh

Để phân biệt ta kí hiệu phép toán mở rộng trùng kí hiệu với các phép toán gốc nhưng có thêm chỉ số trên NC (ngữ cảnh).

3.4.1. Mở rộng phép toán chọn

Phép toán chọn được áp dụng trên một quan hệ. Kết quả của phép toán chọn cho một quan hệ mà tất cả các bộ của quan hệ đó phải thỏa mãn một điều kiện xác định. Trong bài báo này chúng ta sẽ xét hai dạng của phép chọn: "A = a" hoặc "A = B", trong đó A và B là tên các thuộc tính và a là một giá trị xác định. Cho r là một quan hệ được phân hoạch trên lược đồ R, $A \in R$. Chúng ta định nghĩa

$$\sigma^{\mathrm{NC}}_{A=a}(r) = s(R),$$

trong đó

$$\mathrm{SURE}(s) = \{t \mid t \in \mathrm{SURE}(r) \text{ và } t[A] = a\}$$
$$\mathrm{MAYBE}(s) = \{t \mid t \in \mathrm{MAYBE}(r) \text{ và } t[A] = a\} \cup \{t' \mid \exists t \in r \text{ mà } t[A] \in \Delta_{i_1} \cup \Delta_{i_2} \text{ và } t'[A] = a,\ t'[B] = t[B] \text{ với mọi } B \ne A\}$$

$$\sigma^{\mathrm{NC}}_{A=B}(r) = s(R),$$

trong đó

$\text{SURE}(s) = \{t \mid t \in \text{SURE}(r) \text{ và } t[A] = t[B]\}$

$\text{MAYBE}(r) = \{t \mid t \in \text{MAYBE}(r) \text{ và } t[A] = t[B]\} \cup$

$\{t' \mid \exists t \in r \text{ mà } t[A] \in \Delta_{i_1} \cup \Delta_{i_2} \text{ và } t[B]!,\ t'[A] = t[B],\ t'[C] = t[C] \text{ với mọi } C \neq A\} \cup$

$\{t' \mid \exists t \in r \text{ mà } t[B] \in \Delta_{i_1} \cup \Delta_{i_2} \text{ và } t[A]!,\ t'[B] = t[A],\ t'[C] = t[C] \text{ với mọi } C \neq B\}.$

Ví dụ 2. Với quan hệ $r(A, B, C)$ của Ví dụ 1 ta có $\sigma^{NC}_{A=1}(r)$ là quan hệ trong bảng 1 và $\sigma^{NC}_{A=B}(r)$ là quan hệ trong bảng 2.

s_1	A	B	C
	1	δ_1	4
	- - -	- - -	- - -
	1	5	6
	1	3	6

bảng 1

s_1	A	B	C
	- - -	- - -	- - -
	1	1	4
	5	5	6

bảng 2

Mệnh đề 4. *$\sigma^{NC}_{A=a}$ là một mở rộng thỏa đáng của $\sigma_{A=a}$ tương ứng với hàm khả năng* POSS_{NC}.

Chứng minh. Để chứng minh ta phải chỉ ra được $\sigma_{A=a}(\text{POSS}_{NC}(r)) \subseteq \text{POSS}_{NC}(\sigma^{NC}_{A=a}(r))$. Đặt $s = \sigma^{NC}_{A=a}(r)$. Lấy q là một quan hệ bất kỳ trong $\sigma_{A=a}(\text{POSS}_{NC}(r))$. Gọi $\hat{r}$ là quan hệ trong $\text{POSS}_{NC}(r)$ sao cho $q = \sigma_{A=a}(\hat{r})$. Ta sẽ chỉ ra $q \in \text{POSS}_{NC}(s)$. Do $\hat{r} = \text{POSS}_{NC}(r)$ nên $\hat{r}$ là mô hình của $r' \triangleright r$, tức là $\forall a_1, a_2, \ldots, a_k$ là các null xuất hiện trong r', tồn tại các thay thế V_i của a_i $\forall 1 \le i \le k$ sao cho $\hat{r} = S^{V_1V_2\ldots V_k}_{a_1a_2\ldots a_k}(r')$. Để ngắn gọn ta viết là $S_1(r')$ thay cho $S^{V_1V_2\ldots V_k}_{a_1a_2\ldots a_k}(r')$. Vậy $q = \sigma_{A=a}(S_1(r'))$. Ta có $r' = T_1 \cup T_2 \cup T_3 \cup T_4$ với $T_1 = \{t \mid t \in \text{SURE}(r') \text{ và } t[A] = a\}$, $T_2 = \{t \mid t \in \text{MAYBE}(r') \text{ và } t[A] = a\}$, $T_3 = \{t \in r' \text{ mà } t[A] \in \Delta_{i_1} \cup \Delta_{i_2}\}$, $T_4 = r' - T_1 \cup T_2 \cup T_3$. Vậy $q = \sigma_{A=a}(S_1(T_1 \cup T_2 \cup T_3 \cup T_4)) = \sigma_{A=a}(S_1(T_1)) \cup \sigma_{A=a}(S_1(T_2)) \cup \sigma_{A=a}(S_1(T_3)) \cup \sigma_{A=a}(S_1(T_4)) = S_1(T_1) \cup S_1(T_2) \cup S_1(T_3')$ với $T_3' \subseteq T_3$. Vậy $q = S_1(T_1 \cup T_2 \cup T_3') = S_1(s'')$. Do $\text{SURE}(s) = T_1 \subseteq s''$ nên $s'' \triangleright s$ do đó ta có q là mô hình yếu của s'' với $s'' \triangleright s$. Vậy $q \in \text{POSS}_{NC}(s)$.

Mệnh đề 5. *Nếu $\nexists t, t' \in \text{SURE}(r)$ sao cho $t[A] \in \Delta_{i_1} \cup \Delta_{i_2}$ và $t[A] = t'[A]$ thì $\sigma^{NC}_{A=a}$ là một mở rộng chính xác của $\sigma_{A=a}$ tương ứng với hàm khả năng* POSS_{NC}.

Chứng minh. Theo Mệnh đề 4 ta đã có $\sigma_{A=a}(\text{POSS}_{NC}(r) \subseteq \text{POSS}_{NC}(\sigma^{NC}_{A=a}(r))$ nên chỉ cần chứng minh $\text{POSS}_{NC}(s) \subseteq \sigma_{A=a}(\text{POSS}_{NC}(r))$ là đủ. Cho $q \in \text{POSS}_{NC}(s)$. Chúng ta phải chỉ ra tồn tại $\hat{r} \in \text{POSS}_{NC}(r))$ để $q = \sigma_{A=a}(\hat{r})$. Do $q \in \text{POSS}_{NC}(s)$ nên q là mô hình yếu của s' với $s' \triangleright s$, giả sử $q = S_1(s')$. Ta có $s = T_1 \cup T_2 \cup T_3 \cup T_4$ trong đó $T_1 = \{t \mid t \in \text{SURE}(r) \text{ và } t[A] = a\}$, $T_2 = \{t \mid t \in \text{MAYBE}(r) \text{ và } t[A] = a\}$, $T_3 = \{t' \mid \exists t \in \text{SURE}(r) \text{ mà } t[A] \in \Delta_{i_1} \cup \Delta_{i_2} \text{ và } t'[A] = a,\ t'[B] = t[B] \text{ với mọi } B \neq A\}$, $T_4 = \{t' \mid \exists t \in \text{MAYBE}(r) \text{ mà } t[A] \in \Delta_{i_1} \cup \Delta_{i_2} \text{ và } t'[A] = a,\ t'[B] = t[B] \text{ với mọi } B \neq A\}$. Gọi các null phụ thuộc ngữ cảnh chưa biết và null phụ thuộc ngữ cảnh mở xuất hiện trong cột thuộc tính A là $b_1, b_2, \ldots, b_m$. Ta có $T_3 = S^{aa\ldots a}_{b_1b_2\ldots b_k}(T_5) = S_2(T_5)$ (*) với $T_5 = \{t \in \text{SURE}(r) \text{ mà } t[A] \in \Delta_{i_1} \cup \Delta_{i_2}\}$ và $T_4 = S^{aa\ldots a}_{b_1b_2\ldots b_k}(T_6) = S_2(T_6)$ (**) với $T_6 = \{t \in \text{MAYBE}(r) \text{ mà } t[A] \in \Delta_{i_1} \cup \Delta_{i_2}\}$. Do $s' \triangleright s$ nên $s' = T_1 \cup T_2' \cup T_3' \cup T_4'$ với $T_2' \subseteq T_2$, $T_3' \subseteq T_3$, $T_4' \subseteq T_4$. Vậy $q = S_1(T_1 \cup T_2' \cup T_3' \cup T_4') = \sigma_{A=a}(S_1(T_1 \cup T_2' \cup T_3' \cup T_4'))$. Theo (*) và (**) ta có $T_3' = S_2(T_5')$ và $T_4' = (S_2(T_6')$ với $T_5' \subseteq T_5$, $T_6' \subseteq T_6$. Đặt $T_7 = \text{SURE}(r) \setminus T_1 \cup T_5'$, theo giả thiết các giá trị null xuất hiện tại cột thuộc tính A trong T_7 sẽ không xuất hiện trong T_5', do đó ta có thể chọn được một thay thế S_3 cho các bộ trong T_7 mà sao cho các phép thế có thể của các giá trị null xuất hiện tại cột thuộc tính A trong T_7 không trùng với giá trị a. Khi đó $q = \sigma_{A=a}(S_1(T_1 \cup T_2') \cup S_2(T_5') \cup S_2(T_6')) = \sigma_{A=a}(S_1S_2(T_1 \cup T_2' \cup T_5' \cup T_6')) = \sigma_{A=a}(S_1S_2S_3(T_1 \cup T_5' \cup T_7 \cup T_2' \cup T_6'))$. Đặt $r' = T_1 \cup T_5' \cup T_7 \cup T_2' \cup T_6'$, ta có $\text{SURE}(r') = T_1 \cup T_5' \cup T_7 = \text{SURE}(r)$ và $r' \subseteq r$ nên $r' \triangleright r$. Vậy $q = \sigma_{A=a}(S_1S_2S_3(r'))$ với $r' \triangleright r$. Đặt $\hat{r} = S_1S_2S_3(r')$ rõ ràng $\hat{r} \in \text{POSS}_{NC}(r)$ vậy tồn tại $q = \sigma_{A=a}(\hat{r})$.

Mệnh đề 6. *$\sigma^{NC}_{A=B}$ là một mở rộng thỏa đáng của $\sigma_{A=B}$ tương ứng với hàm khả năng* POSS_{NC}.

Chứng minh. Cách chứng minh tương tự như đã làm cho Mệnh đề 4.

Mệnh đề 7. *Nếu* $\nexists t, t' \in \mathrm{SURE}(r)$ *sao cho* $t[A] \in \Delta_{i_1} \cup \Delta_{i_2}$, $t[A] = t'[A]$ *và* $t[B] \in \Delta_{i_1} \cup \Delta_{i_2}$, $t[B] = t'[B]$ *thì* $\sigma^{\mathrm{NC}}_{A=B}$ *là một mở rộng chính xác của* $\sigma_{A=B}$ *tương ứng với hàm khả năng* $\mathrm{POSS_{NC}}$.

Chứng minh. Cách chứng minh tương tự như đã làm cho Mệnh đề 5.

3.4.2. Mở rộng phép kết nối

Cho $r(R)$ và $s(S)$ là các quan hệ được phân hoạch, trong đó $R \cap S = X$. Bộ $t_r \in r$ và $t_s \in s$ được gọi là tương hợp trên X nếu $\forall A \in X$ hoặc $t_r(A)!$ và $t_r(A) = t_s(A)$ hoặc $t_r(A) \in \Delta_{i_1} \cup \Delta_{i_2}$ hoặc $t_s(A) \in \Delta_{i_1} \cup \Delta_{i_2}$. Chúng ta định nghĩa

$$r \rhd\lhd^{\mathrm{NC}} s = q(RS),$$

trong đó

$$\mathrm{SURE}(q) = \{t(RS) \mid \exists t_r \in \mathrm{SURE}(r),\ \exists t_s \in \mathrm{SURE}(s) \text{ sao cho } t(R) = t_r \text{ và } t(S) = t_s\}$$

và $\mathrm{MAYBE}(q) = \{t(RS) \mid$ tồn tại các bộ tương hợp $t_r \in r$ và $t_s \in s$ sao cho $t(R - X) = t_r(R - X)$, $t(S - X) = t_s(S - X)$ và $\forall A \in X$, nếu $t_r(A)!$ thì $t(A) = t_r(A)$ ngược lại $t(A) = t_s(A)\}$.

Ví dụ 8. Giả sử r và s là hai quan hệ trong bảng 1 và bảng 2 thì $r \rhd\lhd^{\mathrm{NC}} s$ là quan hệ trong bảng 3.

r	A	B
	δ_1	1
	3	δ_2
	- - -	- - -
	4	2

bảng 1

s	B	C
	1	2
	δ_2	4
	- - -	- - -
	δ_3	6

bảng 2

$r \rhd\lhd^{\mathrm{NC}} s$	A	B	C
	δ_1	1	2
	3	δ_2	4
	- - -	- - -	- - -
	δ_1	1	4
	δ_1	1	6
	3	1	2
	3	δ_3	6
	4	2	4
	4	2	6

bảng 3

Định nghĩa 8. Cho r và s là hai quan hệ null phụ thuộc ngữ cảnh được phân hoạch, $\hat{r}$ là mô hình yếu của r' với r' *triangleright* r, $\hat{s}$ là mô hình yếu của s' với s' *triangleright* s. Khi đó $\hat{r}$ và $\hat{s}$ được gọi là tương hợp nếu như trong quá trình thay thế các null phụ thuộc ngữ cảnh từ r' thành $\hat{r}$ và s' thành $\hat{s}$ các giá trị null mà cùng xuất hiện trong r' và s' đều được thay thế như nhau.

Định nghĩa 9. Cho r và s là hai quan hệ null phụ thuộc ngữ cảnh được phân hoạch. Ta định nghĩa $\mathrm{POSS}(r) \rhd\lhd \mathrm{POSS}(s) = \{\hat{r} \rhd\lhd \hat{s} \mid \hat{r} \in \mathrm{POSS_{NC}}(r),\ \hat{s} \in \mathrm{POSS_{NC}}(s)$ và $\hat{r}, \hat{s}$ là tương hợp$\}$.

Mệnh đề 8. *Phép kết nối* $\rhd\lhd^{\mathrm{NC}}$ *là một mở rộng thỏa đáng của* $\rhd\lhd$ *tương ứng với hàm khả năng* $\mathrm{POSS_{NC}}$.

Chứng minh. Đặt $q = r \rhd\lhd^{\mathrm{NC}} s$, giả sử $\hat{r} \in \mathrm{POSS_{NC}}(r)$ và $\hat{s} \in \mathrm{POSS_{NC}}(s)$ là tương hợp. Đặt $\hat{q} = \hat{r} \rhd\lhd \hat{s}$. Ta phải chứng minh $\hat{q} \in \mathrm{POSS_{NC}}(q)$. Giả sử $\hat{r} = S_1S_2(r')$ và $\hat{s} = S_3S_2(s')$ với $r' \rhd r$, $s' \rhd s$ và S_2 là một phép thế có thể của các null phụ thuộc ngữ cảnh cùng xuất hiện trong r' và s', như vậy $\hat{q} = S_1S_2(r') \rhd\lhd S_3S_2(s')$. Ta có $r' = T_1 \cup T_2$ với $T_1 = \{t_r \in \mathrm{SURE}(r) \mid \exists t_s \in \mathrm{SURE}(s) : \forall A \in X,\ t_r(A) = t_s(A)\}$, $T_2 = r' - T_1$ và $s' = S_1 \cup S_2$ với $S_1 = \{t_s \in \mathrm{SURE}(s) \mid \exists t_r \in \mathrm{SURE}(r) : \forall A \in X,\ t_r(A) = t_s(A)\}$, $S_2 = s' - S_1$. Vậy $\hat{q} = S_1S_2(T_1 \cup T_2) \rhd\lhd S_3S_2(S_1 \cup S_2)$. Ta có

$S_1S_2(T_1) \bowtie S_3S_2(S_1) = S_1S_2S_3(T_1 \bowtie^{\mathrm{NC}} S_1) = S_1S_2S_3(\mathrm{SURE}(q))$, hơn nữa $\forall t_r \in r'$, $t_s \in s'$ nếu t_r và t_s không tương hợp thì $S_1S_2(t_r) \bowtie S_3S_2(t_s) = \emptyset$ (*), thật vậy do t_r và t_s không tương hợp nên $t_r(A)!$, $t_s(A)!$ và $t_r(A) \neq t_s(A)$, rõ ràng ta có (*), từ đó suy ra $S_1S_2(T_i) \bowtie S_3S_2(S_j) \subseteq S_1S_2S_3(T_i \bowtie^{\mathrm{NC}} S_j) \subseteq S_1S_2S_3(\mathrm{MAYBE}(q))$ (với $i \neq j$). Vậy $\hat{q} = S_1S_2S_3(\mathrm{SURE}(q) \cup q_1)$ với $q_1 \subseteq \mathrm{MAYBE}(q)$. Đặt $q' = \mathrm{SURE}(q) \cup q_1$, ta có $q' \rhd q$ và $\hat{q}$ là mô hình yếu của q', tức là $\hat{q} \in \mathrm{POSS}_{\mathrm{NC}}(q)$.

4. KẾT LUẬN

Trong mô hình dữ liệu quan hệ, đại số quan hệ là cơ sở trung tâm của các ngôn ngữ hỏi đáp vì vậy việc mở rộng đại số quan hệ cho các giá trị null là việc làm quan trọng. Trong bài báo này, mặc dù chúng tôi đã không đưa ra một đại số quan hệ mở rộng đầy đủ, nhưng đã trình bày được phương pháp giải quyết vấn đề khi cố gắng mở rộng một vài phép toán chọn lọc. Các vấn đề có thể tiếp tục là đi sâu nghiên cứu nội dung thông tin của quan hệ một phần và mở rộng nốt các phép toán còn lại.

TÀI LIỆU THAM KHẢO

[1] Bùi Thị Thúy Hiền, Cơ sở dữ liệu quan hệ với giá trị null phụ thuộc vào ngữ cảnh, *Tạp chí Tin học và Điều khiển học* **15** (1) (1999) (26-36).

[2] Bùi Thị Thúy Hiền - Ngữ nghĩa dữ liệu trong cơ sở dữ liệu null ngữ cảnh, sẽ đăng trong *Tạp chí Tin học và Điều khiển học* **15** (4) (1999).

[3] N. C. Hồ và L. T. Thăng, Ngữ nghĩa của dữ liệu trong cơ sở dữ liệu với thông tin không đầy đủ, *Tạp chí Tin học và Điều khiển học* **11** (2) (1995) 7-15.

[4] J. Biskup, A formal approach to null values in database relations, in Hervé Gallaire, Jack Minker, and Jean Marie Nicolas editors, *Advances in Database Theory, Vol. I. Plenumpress*, NewYork, 1981.

[5] E. F. Codd, Extending the database relational model to capture more meaning, *ACM TODS* **4** (4)(1979) 397-434.

[6] G. Gottlob and R. Zicari, Closed world databases opened through null values, in *Proc. 14th. Conf. on very Large Data Bases*, pages 50-61, 1988.

[7] N. C. Ho, A relational model of databases with context dependent null values, *Bull. Pol. Ac. Tech.* **36** (1-2) (1988) 91-105.

[8] T. Imielinski and W. Lipski Jr., Incomplete information in relational databases, *Journal of the ACM* **31** (4) (1984) 761-791.

[9] W. Lipski Jr., On semantic issues connected with incomplete information databases, *Journal of the ACM* **4** (3) (1979) 262-296.

[10] D. Maier, *The Theory of Relational Databases*, Computer Science Press, 1983.

[11] J. Paredaens et al., *In the Structure of the Relational Database Model*, Chapter 6, pages 157-176, Springer-Verlag Berlin Heidelberg, 1989.

[12] C. Zaniolo, Database relations with null values, *Journal of Computer and System Sciences* **28** (1) (1984) 142-166.

Nhận bài ngày 14 - 12 - 1998
Nhận lại sau khi sửa ngày 24 - 5 - 1999

Khoa Toán - Cơ - Tin học
Trường Đại học Khoa học tự nhiên - ĐHQG Hà Nội.