Tạp chí Tin học và Điều khiển học, T. 19, S. 4 (2003), 359–365

MỞ RỘNG MỘT SỐ TOÁN TỬ QUAN HỆ LÊN CÁC CƠ SỞ DỮ LIỆU THIẾU THÔNG TIN

HỒ THUẦN[1], HOÀNG THỊ LAN GIAO[2]

[1] *Viện Công nghệ thông tin*
[2]*Khoa CNTT, Trường Đại học Khoa học Huế*

Abstract. The aim of the paper is to extend some relational operators to incomplete databases basing on possibility functions (P. F. s). We discuss which properties a P. F. should posses and what constitutes a reasonable extension of an operator relative to a given P. F. We shall limit our attention to four operators: join, union, project and select. Also, various properties of the extended operators will be considered and proved in detail.

Tóm tắt. Bài báo nhằm mở rộng một số toán tử quan hệ lên cơ sở dữ liệu thiếu thông tin dựa vào các hàm khả năng, tiếp đó là khảo sát các tính chất cần thiết của một hàm khả năng cũng như các tính chất thể hiện tính hợp lý của một toán tử quan hệ mở rộng. Bài báo chỉ tập trung nghiên cứu các tính chất mở rộng của bốn toán tử quen thuộc: phép nối, phép hợp, phép chiếu và phép chọn đối với hai hàm khả năng đặc biệt.

1. MỞ ĐẦU

Cơ sở dữ liệu với thông tin không đầy đủ đang thu hút sự quan tâm của nhiều độc giả, trong khuôn khổ bài báo này chúng tôi sẽ khai thác các cơ sở dữ liệu thiếu thông tin thông qua các hàm khả năng và tìm cách mở rộng một số toán tử quan hệ lên các cơ sở dữ liệu thiếu thông tin. Ở đây, hàm khả năng được hiểu như là một thu hẹp của lớp các thể hiện đầy đủ của một quan hệ thiếu thông tin. Thật ra, ở đây chúng ta chỉ tập trung xem xét các mở rộng hoàn toàn tự nhiên của bốn toán tử quan hệ quen thuộc là phép hợp, phép kết nối, phép chiếu và phép chọn, đồng thời khảo sát tính hợp lý của các mở rộng này theo các nghĩa khác nhau trong tương quan với hai hàm khả năng đặc biệt. Các kết quả trong bài báo này hầu hết được gợi ý từ [1] bởi D. Maiez. Chúng tôi đã hệ thống lại trong một bức tranh chung và đưa ra các chứng minh chi tiết.

2. CÁC KHÁI NIỆM CƠ BẢN

Thuật ngữ chính được sử dụng trong các cơ sở dữ liệu với thông tin không đầy đủ là NULL. Nhiều người sử dụng thuật ngữ này với những ý nghĩa khác nhau. Ở đây chúng ta chủ yếu xét những NULL tồn tại nhưng chưa biết ([1, 2, 6]). Những NULL này được dùng để chỉ một giá trị chắc chắn tồn tại nhưng chưa xác định được. Chúng ta dùng chung ký hiệu $\perp$ trong bảng biểu diễn quan hệ để nói rằng tại đó xuất hiện giá trị NULL như thế.

Cho R là một lược đồ quan hệ. Một bộ t trên R không chứa giá trị NULL nào cả được gọi là bộ toàn phần, còn một bộ nói chung thì được gọi là một phần. Giá trị thuộc tính A của bộ t được gọi là xác định và được ký hiệu là $t(A)\downarrow$ nếu giá trị đó khác NULL. Tương tự, nếu X là tập thuộc tính, ta viết $t(X)\downarrow$ mỗi khi $t(A)\downarrow$ với mọi $A \in X$. Một quan hệ r được gọi là toàn phần và được ký hiệu là $r\downarrow$ nếu mọi bộ của r là toàn phần, một quan hệ nói chung được gọi là một phần.

Cho hai bộ t và u trên cùng một lược đồ R, t được gọi là có thông tin hơn u nếu với mọi $A \in R$, $u(A)\downarrow$ kéo theo $t(A)\downarrow$ và $t(A) = u(A)$, lúc đó ta viết $t \geq u$. Trong trường hợp ngược lại, ta viết $t \ngeq u$.

Với mỗi lược đồ quan hệ R, ta ký hiệu $\mathrm{Rel}(R)$ là tập tất cả các quan hệ toàn phần và $\mathrm{Rel}\uparrow(R)$ là tập tất cả các quan hệ một phần trên R. Ta sẽ viết là Rel hoặc Rel$\uparrow$ nếu không muốn chỉ rõ trên lược đồ quan hệ nào. Cho $r, s \in \mathrm{Rel}\uparrow(R)$, r được gọi là có thông tin nhiều hơn s nếu với mọi $t_s \in s$ tồn tại $t_r \in r$ sao cho $t_r \geq t_s$. Ngoài ra nếu r là toàn phần thì r được gọi là một mở rộng của s và ký hiệu $r\downarrow\geq s$. r được gọi là một mở rộng cực tiểu của s nếu $r\downarrow\geq s$ và với mọi w mà $w\downarrow\geq s$ ta có $w \supset r$. Nếu r là một mở rộng của s và với mọi $t_r \in r$ tồn tại $t_s \in s$ sao cho $t_r \geq t_s$ thì r được gọi là mở rộng đóng của s. Tập tất cả các mở rộng của s được ký hiệu là $E(s)$ còn tập các mở rộng cực tiểu của s được ký hiệu là $ME(s)$. Với mỗi quan hệ $s, E(s)$ luôn khác rỗng nhưng $ME(s)$ thì không nhất thiết như vậy ([1]).

Để mở rộng các toán tử thông thường lên các quan hệ thiếu thông tin, chúng ta sẽ xem một quan hệ $r \in \mathrm{Rel}\uparrow(R)$ như một tập các quan hệ trong $\mathrm{Rel}(R)$ (thường là các mở rộng của r) thể hiện tập khả năng của r. Nói cách khác, tồn tại một ánh xạ đa trị POSS từ $\mathrm{Rel}\uparrow(R)$ vào $\mathrm{Rel}(R)$ mà ta gọi là hàm khả năng. Hàm khả năng POSS sẽ được gọi là hợp lý nếu hai điều kiện sau thỏa mãn:

i) $\forall r \in \mathrm{Rel}\uparrow$: $ME(r) \subseteq POSS(r) \subseteq E(r)$;

ii) $\forall r, s \in \mathrm{Rel}\uparrow$: $s \in POSS(r) \Longleftrightarrow (s\downarrow) \wedge (POSS(r) \subseteq POSS(s))$.

Bây giờ giả sử γ là một toán tử hai ngôi trên Rel và γ' là một mở rộng của nó lên Rel$\uparrow$.

• γ' được gọi là một mở rộng chính xác của γ đối với hàm POSS nếu: $\forall r, s \in \mathrm{Rel}\uparrow$

$$POSS(r\gamma' s) = POSS(r)\gamma POSS(s) := \{q_1\gamma q_2 \,|\, q_1 \in POSS(r),\ q_2 \in POSS(s)\}.$$

• γ' được gọi là một mở rộng phù hợp của γ đối với hàm $POSS$ nếu

$$\forall r, s \in \mathrm{Rel}\uparrow: POSS(r\gamma' s) \supseteq POSS(r)\gamma POSS(s).$$

• γ' được gọi là thu hẹp của γ đối với hàm khả năng POSS nếu với mọi $r, s \in \mathrm{Rel}\uparrow$, không tồn tại $q \in \mathrm{Rel}\uparrow$ sao cho $POSS(r\gamma' s) \supset POSS(q) \supseteq POSS(r)\gamma POSS(s)$.

Nếu γ và γ' là các toán tử một ngôi thì ta có các định nghĩa tương tự:

• γ' được gọi là một mở rộng chính xác của γ đối với hàm POSS nếu $\forall r \in \mathrm{Rel}\uparrow$

$$POSS(\gamma'(r)) = \gamma(POSS(r)) := \{\gamma(q) \,|\, q \in POSS(r)\}$$

• γ' được gọi là một mở rộng phù hợp với γ đối với hàm POSS nếu

$$\forall r \in \mathrm{Rel}\uparrow: POSS(\gamma'(r)) \supseteq \gamma(POSS(r)).$$

• γ' được gọi là thu hẹp của γ đối với hàm khả năng POSS nếu với mọi $r \in \mathrm{Rel}\uparrow$, không tồn tại $q \in \mathrm{Rel}\uparrow$ sao cho $POSS(\gamma'(r)) \supset POSS(q) \supseteq \gamma(POSS(r))$.

Rõ ràng, nếu γ' là một mở rộng chính xác của γ thì γ' cũng phù hợp và thu hẹp của γ.

3. CÁC TOÁN TỬ MỞ RỘNG ĐỐI VỚI MỘT SỐ HÀM KHẢ NĂNG ĐẶC BIỆT

Mục này sẽ xét sự mở rộng và tính chất của các toán tử quan hệ quen thuộc như phép kết nối, phép hợp, phép chọn, phép chiếu đối với các hàm khả năng đặc biệt $POSS_O$ và POSS_{CE} ([1]). Các hàm này được định nghĩa như sau:

$$POSS_O(s) := \{r\!\downarrow \mid r \text{ là một mở rộng của } s\},\ \forall s \in \mathrm{Rel}\!\uparrow;$$

$$POSS_{CE}(s) := \{r\!\downarrow \mid r \text{ là một mở rộng đóng của } s\},\ \forall s \in \mathrm{Rel}\!\uparrow.$$

Từ định nghĩa ta dễ dàng đi đến khẳng định sau:

Mệnh đề 1. *$POSS_O$ và $POSS_{CE}$ là các hàm khả năng hợp lý.*

3.1. Mở rộng các toán tử quan hệ

a) Mở rộng phép nối. Cho $r \in \mathrm{Rel}\!\uparrow\!(R),\ s \in \mathrm{Rel}\!\uparrow\!(S)$ và $R \cap S = X$. Ta định nghĩa

$$r \bowtie' s := \{t(RS) \mid \exists t_r \in r,\ t_s \in s:\ (t_r(X)\!\downarrow = t_s(X)\!\downarrow) \wedge (t(R) = t_r) \wedge (t(S) = t_s)\}$$

b) Mở rộng phép hợp. Cho $r, s \in \mathrm{Rel}\!\uparrow\!(R)$, phép hợp mở rộng được định nghĩa như sau:

$$r \cup' s := \{t \,|\, t \in r \ \text{ hoặc } t \in s\}.$$

c) Mở rộng phép chiếu. Cho $r \in \mathrm{Rel}\!\uparrow(R)$ và $X \subseteq R$. Phép chiếu mở rộng được định nghĩa bởi

$$\Pi'_X(r) := \{t(X) \,|\, t \in r\}.$$

d) Mở rộng phép chọn. Cho $r \in \mathrm{Rel}\!\uparrow(R),\ B, A \subseteq R$ và a là một giá trị bất kỳ, các phép chọn mở rộng được định nghĩa như sau:

$$\sigma'_{A=a}(r) := \{t(R) \,|\, t \in r \ \text{ và } \ t(A) = a\}$$

$$\sigma'_{A=B}(r) := \{t(R) \,|\, t \in r,\ t(A)\!\downarrow = t(B)\!\downarrow\}$$

3.2. Tính chất của các toán tử mở rộng đối với hàm khả năng $POSS_O$

Mệnh đề 2. *$\bowtie'$ là mở rộng phù hợp và thu hẹp nhưng không chính xác của $\bowtie$ đối với $POSS_O$.*

Chứng minh.

i) Phù hợp: Đặt $q = r \bowtie' s$, ta chứng minh $POSS_O(r) \bowtie POSS_O(s) \subseteq POSS_O(q)$.

Với mọi $r' \in POSS_O(r),\ s' \in POSS_O(s)$, đặt $q' = r' \bowtie s'$. Lúc đó, với mọi $t_q \in q$, theo định nghĩa, $\exists t_r \in r,\ t_s \in s$ sao cho $t_r(X)\!\downarrow = t_s(X)\!\downarrow$ và $t_q(R) = t_r,\ t_q(S) = t_s$. Do $r'\!\downarrow \geq r$ và $s'\!\downarrow \geq s$, tồn tại $t_{r'} \in r',\ t_{s'} \in s'$ sao cho $t_{r'} \geq t_r$ và $t_{s'} \geq t_s$, vì vậy $t_{r'}(X)\!\downarrow = t_r(X)\!\downarrow = t_s(X)\!\downarrow = t_{s'}(X)\!\downarrow$. Đặt $t_{q'}$ là bộ sao cho $t_{q'}(R) = t'_r$ và $t_{q'}(S) = t_{s'}$. Lúc đó $t_{q'} \in q'$ và $t_{q'} \geq t_q$. Vậy ta có $q' \in POSS(q)$.

ii) Thu hẹp: Lấy tuỳ ý $q \in$Rel$\uparrow$ sao cho $POSS_O(r) \bowtie POSS_O(s) \subseteq POSS_O(q)$, ta chứng minh $POSS_O(q) \supseteq POSS_O(r \bowtie' s)$. Thật vậy, nếu ngược lại thì $r \bowtie' s \not\geq q$, nghĩa là tồn tại $t \in q$ mà với mọi $u \in r \bowtie' s,\ u \not\geq t$.

Đặt $r_1 := \{t_r \in r \,|\, \exists t_s \in s \text{ sao cho } t_r(X)\!\downarrow = t_s(X)\!\downarrow\};\ r_2 = r - r_1;$

$s_1 := \{t_s \in s \,|\, \exists t_r \in r \text{ sao cho } t_r(X)\!\downarrow = t_s(X)\!\downarrow\};\ s_2 = s - s_1.$

Rõ ràng $r \bowtie' s = r_1 \bowtie' s_1$ và $r_1 \bowtie' s_2 = r_2 \bowtie' s_1 = r_2 \bowtie' s_2 = \varnothing$.

Gọi T là tập các giá trị ($T \neq$ NULL) đã xuất hiện trong $r \cup s \cup \{t\}$. Thiết lập r' và s' là các đầy đủ hóa của r và s bằng cách thay các giá trị NULL trong r, s bởi các giá trị xác định khác nhau và không nằm trong T. Lúc đó $r' \in POSS_O(r)$, $s' \in POSS_O(s)$ và $r \bowtie s' = r'_1 \bowtie s'_1$($r'_1$ và s'_1 tương ứng là tập những bộ trong r' và s' mà làm đầy của một bộ tương ứng trong r_1, s_1). Mặt khác, với mọi $t' \in r'_1 \bowtie s'_1$ tồn tại $u \in r \bowtie' s$, sao cho t' làm đầy u bởi các giá trị không thuộc T. Theo trên $u \not\geq t$, nên $t' \not\geq t$. Điều này đúng với mọi $t' \in r'_1 \bowtie s'_1$ nên $r' \bowtie s' = r'_1 \bowtie s'_1 \notin POSS_O(q)$. Mâu thuẫn với giả thiết.

iii) Không chính xác: Cho các quan hệ

$r(A$	$B)$		$s(B$	$C)$
$\perp$	1		1	a
			1	b

Lúc đó,

$r \bowtie' s$	$(A$	B	$C)$
	$\perp$	1	a
	$\perp$	1	b

Do đó, nếu chọn

$q(A$	B	$C)$
2	1	a
3	1	b

thì $q{\downarrow}\geq r \bowtie' s$ nên $q \in POSS_O(r \bowtie' s)$. Nhưng $q \notin POSS_O(r) \bowtie POSS_O(s)$.

Mệnh đề 3. *$\cup'$ là một mở rộng chính xác của $\cup$ đối với $POSS_O$.*

Chứng minh. Đặt $q = r \cup' s$, ta chứng minh $POSS_O(r) \cup POSS_O(s) = POSS_O(q)$.

Với mọi $q' \in POSS_O(r) \cup POSS_O(s)$, $q' = r' \cup s'$, với $r'{\downarrow}\geq r$ và $s'{\downarrow}\geq s$. Lấy bất kỳ $t \in q$, lúc đó $t \in r$ hoặc $t \in s$. Giả sử $t \in r$ (trường hợp $t \in s$ chứng minh tương tự). Lúc đó tồn tại $t' \in r' \subseteq q'$ sao cho $t' \geq t$. Vậy $q' \geq q$ hay $q' \in POSS_O(q)$.

Ngược lại, với mọi $q'' \in POSS_O(q)$, $q''{\downarrow}\geq q \geq r$ và $q''{\downarrow}\geq q \geq s$ hay $q'' \in POSS_O(r)$ và $q'' \in POSS_O(s)$. Do đó $q'' = q'' \cup q'' \in POSS_O(r) \cup POSS_O(s)$. Mệnh đề được chứng minh. ∎

Mệnh đề 4. *Π'_X là một mở rộng chính xác của Π_X đối với $POSS_O$.*

Chứng minh. Đặt $s = \Pi'_X(r)$, ta chứng minh $POSS_O(s) = \Pi_X(POSS_O(r))$

- $POSS_O(s) \subseteq \Pi'_X(POSS_O(r))$

Lấy tuỳ ý $q \in POSS_O(s)$, ta có $q{\downarrow}\geq s$. Vậy, với mọi $t_s \in s$ tồn tại $t_q \in q$ sao cho $t_q{\downarrow}\geq t_s$. Mặt khác theo định nghĩa của phép chiếu Π'_X, $t_s = t_r(X)$ với $t_r \in r$ nên ta có $t_q{\downarrow}\geq t_r(X)$. Gọi r' là bộ đầy đủ hóa của r mà các giá trị NULL trên X được thay bởi các giá trị của bộ t_q tương ứng trong q. Khi đó $r' \in POSS_O(r)$ và $q = \Pi_X(r') \in \Pi_X(POSS_O(r))$.

- $\Pi_X(POSS_O(r)) \subseteq POSS_O(s)$

Với mọi $r' \in \Pi_X(POSS_O(r))$ ta có $r' = \Pi_X(q)$ với $q{\downarrow}\geq r$ do đó $\Pi_X(q){\downarrow}\geq \Pi'_X(r) = s$. Vậy $\Pi_X(q) \in POSS_O(s)$ hay $r' \in POSS_O(s)$. Mệnh đề được chứng minh. ∎

Mệnh đề 5. *$\sigma'_{A=a}$ ($\sigma'_{A=B}$) là một mở rộng phù hợp và thu hẹp nhưng không chính xác của $\sigma_{A=a}(\sigma_{A=B})$ đối với $POSS_O$.*

Chứng minh. Ta chỉ chứng minh với trường hợp $\sigma'_{A=a}$, trường hợp còn lại tương tự.

i) Phù hợp: Với mọi $q \in \sigma_{A=a}(POSS_O(r))$ tồn tại $s \in POSS_O(r)$ sao cho $q \downarrow= \sigma_{A=a}(s)$. Vì $s \in POSS_O(r)$ nên $s \downarrow\geq r$, suy ra $q \downarrow= \sigma_{A=a}(s) \downarrow\geq \sigma'_{A=a}(r)$. Tức là $q \in POSS_O(\sigma_{A=a}(r))$. Do q được chọn tuỳ ý nên $\sigma_{A=a}(POSS_O(r)) \subseteq POSS_O(\sigma'_{A=a}(r))$.

ii) Thu hẹp: Giả sử tồn tại q sao cho $POSS_O(\sigma'_{A=a}(r) \supset POSS_O(q) \supseteq \sigma_{A=a}(POSS_O(r))$. Vậy có $q' \in POSS_O(\sigma'_{A=a}(r))$ và $q' \notin POSS_O(q)$. Do $q' \in POSS_O(\sigma'_{A=a}(r))$ nên $q' \downarrow\geq \sigma'_{A=a}(r)$. Gọi r' là bộ đầy đủ hóa của r được tạo như sau: với mỗi bộ $t_r \in r$, nếu $t_r \in \sigma'_{A=a}(r)$ ta làm đầy t bằng bộ $t_{q'} \in q'$ sao cho $t_{q'} \geq t_r$, ngược lại, nếu $t_r \notin \sigma'_{A=a}(r)$ thay các giá trị NULL trong t_r bởi các giá trị xác định sao cho $t_r(A) \neq a$. Rõ ràng $r' \downarrow\geq r$ và $q' = \sigma_{A=a}(r')$ do đó $q' \in \sigma_{A=a}(POSS_O(r)) \subseteq POSS_O(q)$, mâu thuẫn với giả thiết. Vậy $\sigma'_{A=a}$ là thu hẹp.

iii) Không chính xác: Cho quan hệ

$r(A$	B	$C)$
1	2	$\perp$
1	2	3
$\perp$	2	0

Ta có

$\sigma'_{A=1}(r)$(A	B	C)
1	2	$\perp$
1	2	3

Chọn

$q(A$	B	$C)$
1	2	3
4	1	0

thì $q \downarrow\geq \sigma'_{A=1}(r)$ tức là $q \in POSS_O(\sigma'_{A=1}(r))$. Nhưng rõ ràng $q \notin \sigma_{A=1}(POSS_O(r))$ (vì nếu $s \in \sigma'_{A=1}(POSS_O(r))$ thì mọi bộ $t \in s$ đều có $t(A) = 1$).

3.3. Tính chất của các toán tử mở rộng đối với hàm khả năng $POSS_{CE}$

Mệnh đề 6. *$\bowtie'$ là mở rộng không phù hợp của $\bowtie$ đối với $POSS_{CE}$.*

Chứng minh. Thật vậy, xét các quan hệ

$r(A$	$B)$	$r'(A$	$B)$
$\perp$	1	3	1
2	$\perp$	2	4

$s(B$	$C)$	$s'(B$	$C)$
1	2	1	2
4	$\perp$	4	0

Rõ ràng r' và s' lần lượt là các mở rộng đóng của r và s và ta có:

$r \bowtie' s(A$	B	$C)$
$\perp$	1	2

$r' \bowtie s'(A$	B	$C)$
3	1	2
2	4	0

Khi đó $r' \bowtie s' \in POSS_{CE}(r) \bowtie POSS_{CE}(s)$ nhưng $r' \bowtie s' \notin POSS_{CE}(r \bowtie' s)$. ∎

Mệnh đề 7. *$\cup'$ là một mở rộng chính xác của $\cup$ đối với $POSS_{CE}$.*

Chứng minh. Ta chứng minh $POSS_{CE}(r \cup' s) = POSS_{CE}(r) \cup POSS_{CE}(s)$ với mọi $r,\ s \in \mathrm{Rel}\!\uparrow\!(R)$.

Với mọi $q \in POSS_{CE}(r \cup' s)$, q là một mở rộng đóng của $r \cup' s$. Đặt

$$q_1 = \{t \in q \,|\, \exists t_r \in r \ \text{ và } \ t \geq t_r\} \text{ và } \ q_2 = \{t \in q \,|\, \exists t_s \in s \ \text{ và } \ t \geq t_s\}.$$

Khi đó, dễ thấy q_1 và q_2 lần lượt là các mở rộng đóng của r và s. Hơn nữa, $q = q_1 \cup q_2$. Tức là $q \in POSS_{CE}(r) \cup POSS_{CE}(s)$. Vậy $POSS_{CE}(r \cup' s) \subseteq POSS_{CE}(r) \cup POSS_{CE}(s)$.

Với mọi $q' \in POSS_{CE}(r) \cup POSS_{CE}(s)$ ta có $q' = r' \cup s'$ với r' và s' lần lượt là các mở rộng đóng của r và s. Do $q' \!\downarrow\geq r' \!\downarrow\geq r$ và $q' \!\downarrow\geq s' \!\downarrow\geq s$ nên q' là một mở rộng của $r \cup' s$. Để chứng minh q' là mở rộng đóng của $r \cup' s$ ta lấy bất kỳ $t' \in q'$. Lúc đó, $t' \in r'$ hoặc $t' \in s'$. Nếu $t' \in r'$ thì tồn tại $t \in r \subseteq r \cup' s$ sao cho $t' \geq t$. Tương tự cho trường hợp $t' \in s'$. Tóm lại $q' \in POSS_{CE}(r \cup' s)$. Vậy, $POSS_{CE}(r) \cup POSS_{CE}(s) \subseteq POSS_{CE}(r \cup' s)$. ∎

Mệnh đề 8. *Π'_X là một mở rộng chính xác của Π_X đối với $POSS_{CE}$.*

Chứng minh. Ta chứng minh $\Pi_X(POSS_{CE}(r)) = POSS_{CE}(\Pi'_X(r))$ với mọi $r \in \mathrm{Rel}\!\uparrow(R)$.

• Lấy bất kỳ $q \in \Pi_X(POSS_{CE}(r))$. Tồn tại một mở rộng đóng r' của r sao cho $q = \Pi_X(r')$. Ta sẽ chứng minh q là một mở rộng đóng của $\Pi'_X(r)$.

+ Với mọi $t \in \Pi'_X(r)$, tồn tại bộ $t_r \in r$ sao cho $t_r(X) = t$. Vì $t_r \in r$ và $r' \!\downarrow\geq r$ nên tồn tại bộ $t' \in r'$ sao cho $t' \geq t_r$. Suy ra $t'(X) \!\downarrow\geq t$. Vì $q = \Pi_X(r')$ nên $t'(X) \in q$ và $q \!\downarrow\geq \Pi'_X(r)$.

+ Với mọi $t' \in q = \Pi_X(r')$, tồn tại bộ $t_{r'} \in r'$ sao cho $t' \!\downarrow= t_{r'}(X)$. Vì $t_{r'} \in r'$ nên lại có bộ $t_r \in r$ và $t_{r'} \geq t_r$, từ đây suy ra $t' \!\downarrow\geq t_r(X) \in \Pi'_X(r)$.

Tóm lại, $q \in POSS_{CE}(\Pi'_X(r))$ và như vậy $\Pi_X(POSS_{CE}(r)) \subseteq POSS_{CE}(\Pi'_X(r))$.

• Ngược lại, lấy bất kỳ $q \in POSS_{CE}(\Pi'_X(r))$. Vì $q \!\downarrow\geq \Pi'_X(r)$ nên với mọi bộ $t_r \in r$, tồn tại $t_q \in q$ sao cho $t_q \!\downarrow\geq t_r(X)$.

Gọi r' là một đầy đủ hóa của r tạo ra như sau: Với mỗi bộ $t_r \in r$ (theo trên thì tồn tại bộ $t_q \in q$ và $t_q \geq t_r(X)$) ta thay các giá trị NULL trên X bởi các giá trị của t_q tương ứng. Rõ ràng r' là một mở rộng đóng của r và $q = \Pi_X(r')$. Nói cách khác, $q \in \Pi_X(POSS_{CE}(r))$ và như vậy $\Pi_X(POSS_{CE}(r)) \supseteq POSS_{CE}(\Pi'_X(r))$. Mệnh đề được chứng minh. ∎

Mệnh đề 9. *$\sigma'_{A=a}$ ($\sigma'_{A=B}$) là một mở rộng không phù hợp của $\sigma_{A=a}$ ($\sigma_{A=B}$) đối với $POSS_{CE}$.*

Chứng minh. Hãy xét các quan hệ

$$\begin{array}{cc} r(A & B) \\ \perp & 5 \end{array} \quad \begin{array}{cc} s(A & B) \\ 5 & 5 \end{array}$$

Khi đó $\sigma'_{A=5}(r)$ và $\sigma'_{A=B}(r)$ là các quan hệ rỗng trong khi s là một mở rộng đóng của r và $\sigma_{A=5}(s) = \sigma_{A=B}(s) = s$. Tức là $s \in \sigma_{A=5}(POSS_{CE}(r))$ và $s \in \sigma_{A=B}(POSS_{CE}(r))$. Nhưng $s \notin POSS_{CE}(\sigma'_{A=5}(r))$ và $s \notin POSS_{CE}(\sigma'_{A=5}(r))$. Vậy $\sigma'_{A=a}$ và $\sigma'_{A=B}$ là các mở rộng không phù hợp. ∎

4. KẾT LUẬN

Có nhiều cách tiếp cận khác nhau đối với cơ sở dữ liệu thiếu thông tin như sử dụng lý thuyết tập mờ, lý thuyết tập thô, lôgic đa trị... ([3, 4, 5]). Phương pháp dùng hàm khả năng là một cách tiếp cận khá tự nhiên. Ngoài hai hàm khả năng đã được đề cập, người ta còn đưa ra một số hàm khả năng khác ([1]).

$$POSS_C(r) = \{s \mid s\!\downarrow\geq r\};\ POSS_E(r) = \{s \mid s\!\downarrow\geq r \ \text{ và } \ s \text{ có số bộ chẵn}\};$$

$$POSS_{ME}(r) = \{s \mid s \text{ là một mở rộng cực tiểu của } r\}.$$

Cũng có thể có nhiều cách mở rộng khác nhau của các toán tử quan hệ. Chẳng hạn, LaCroix và Pitrotte ([1]) đề nghị một toán tử nối tự nhiên mở rộng, bảo đảm tất cả các bộ đều được nối. Những bộ không được nối với bộ khác sẽ được thêm những giá trị NULL và bổ sung vào kết quả. Zaniolo ([6]) cũng đã xây dựng một phép nối tổng quát, ở đó, những bộ được nối là những bộ đồng nhất trên những thuộc tính chung hoặc có đúng một giá trị trên thuộc tính chung đó là NULL. Nếu những bộ có giá trị NULL trên một thuộc tính thì lấy giá trị thuộc tính tương ứng trên bộ khác nối với nó. Nếu một bộ không nối với bộ nào cả thì bổ sung các giá trị NULL vào những bộ này và đưa vào kết quả. Các toán tử mở rộng này đều không có tính kết hợp.

Tuy nhiên, do khuôn khổ bài báo có hạn, chúng tôi chỉ xét tính chất của bốn toán tử mở rộng quen thuộc, tương ứng với hai hàm khả năng $POSS_O$ và $POSS_{CE}$ mà không đưa ra các khảo sát rộng hơn.

TÀI LIỆU THAM KHẢO

[1] D. Maier, *The theory of relational databases,* Computer Science Press, 1982.

[2] J. Paredaens, P. DeBra, M. Gyssens and D. Van Gucht, "The structure of the relational databases containing null values", *Prep. of Inter. Conf. on Database Theory, Roma,* 1986.

[3] T. Đ. Hùng, L. T. Vương, Về một cách tiếp cận cơ sở dữ liệu thiếu thông tin, *Tạp chí Tin học và Điều khiển học* **12** (4) (1996) 125–136.

[4] H. Thuần, H. C. Hà, Đại số quan hệ và quan điểm sử dụng null value trên một mô hình cơ sở dữ liệu mờ, *Tạp chí Tin học và Điều khiển học* **17** (4) (2001) 1–10.

[5] L. T. Vương, H. Thuần, A Relational Database extended by Application of Fuzzy Set Theory and Linguistic Variables, *Computer and Intelligence* **8** (2) (1989).

[6] C. Zaniolo, Database Relations with Null Values, *J. of Computer and Syst, Science* **28** (1984) 142–166.

Nhận bài ngày 20 - 12 - 2002

Nhận lại sau sửa ngày 29 - 7 -2003