Tạp chí Tin học và Điều khiển học, T.20, S.2 (2004), 162–170

MỘT SỐ VẤN ĐỀ TRUY VẤN VÀ CẬP NHẬT TRONG CƠ SỞ DỮ LIỆU CÓ TÍNH THỜI GIAN

NGUYỄN ĐÌNH THUÂN

Trường Đại học Thủy sản Nha Trang

Abstract. The developments of database technology have extended the traditional relation database concept in several directions ([1–5]). One of particular interest, to this work is temporal extension. In this paper, we are going to add the temporal dimension to a logic system and define the notion of safe temporal query, temporal rule. After that, we describe some results about effects of updates in temporal database.

Tóm tắt. Sự phát triển về CSDL được phát triển theo nhiều hướng khác nhau ([1–5]). Một trong các hướng đó là phát triển bằng cách thêm vào yếu tố thời gian. Trong bài báo này, chúng tôi bổ sung chiều thời gian vào các khái niệm biến tự do, biến có giới hạn, độc lập trên miền. Từ đó, xác định các công thức truy vấn temporal nào là tin cậy. Ngoài ra, chúng tôi sẽ đề cập đến ngữ nghĩa và hiệu ứng của thao tác cập nhật dữ liệu trong mô hình CSDL có tính thời gian.

1. MỞ ĐẦU

1.1. Hệ lôgic hình thức

Định nghĩa 1. ([3]) Cho L là tập các ký hiệu của ngôn ngữ gọi là công thức, $\vdash$ là hệ suy diễn, Σ là tập các ký hiệu dưới dạng các luật, M là mô hình trên L, ký hiệu $M \models A$ nếu $A \in L$ là True đối với mô hình M và lớp gồm các mô hình của L gọi là K.

Chẳng hạn, trong mô hình lôgic thông thường được đề nghị

$$\Sigma = \{\wedge, \vee, \neg, \rightarrow, (,)\}.$$

Gọi P là tập đếm được các ký hiệu, ta có các luật của L là:

- $P \subset L$.
- Nếu $A \in L$ thì $\neg A \in L$.
- Nếu $A, B \in L$ thì $(A \wedge B), (A \vee B), (A \rightarrow B) \in L$.

Hệ suy diễn $\vdash$ là các phép suy diễn: giao hoán, kết hợp, phân phối, phủ định, Modus Ponens, Modus Tollens.

Mô hình M được định nghĩa như sau:

- Gọi B={True, False}.
- Xét hàm $v : P \rightarrow B$ là hàm lượng giá cho các ký hiệu.
- Mô hình M là cặp $M = (B, v)$.

1.2. Mô hình Temporal Lôgic (TL)

Mô hình TL được nhiều tác giả mô tả khác nhau tuỳ theo lĩnh vực nghiên cứu, trong [5] các tác giả đã đề nghị:

Định nghĩa 2. ([5]) Mệnh đề P_{TL} ngoài các phép toán về lôgic thông thường, còn có các phép toán về lôgic thời gian như sau. Cho A là mệnh đề thì:

PA (P : Past) được phát biểu là trong quá khứ đã có trường hợp là A.

FA (F : Future) được phát biểu là trong tương lai sẽ có trường hợp là A.

HA (Always in the past) được hiểu là A luôn đúng trong quá khứ.

GA (Always in the future) được hiểu là A luôn đúng trong tương lai.

$S(A, B)$ (S : Since) từ khi đã có A thì có trường hợp là B.

$U(A, B)$ (U : Until) đến khi nào có A thì sẽ có trường hợp là B.

Ta gọi tập hợp các mệnh đề cơ sở là $P_T = \{p, q, r...\}$ và gọi tập các biểu thức là $L_T = \{A, B, C, D...\}$.

Định nghĩa 3. [2] (Cú pháp của các mệnh đề P_{TL})

Cho P_T là tập vô hạn các mệnh đề cơ sở, L_T là tập nhỏ nhất là biểu thức thỏa :

- $P_T \subset L_T$.
- Nếu $A \in L_T$ và $B \in L_T$ thì $\neg A$ và $(A \wedge B) \in L_T$.
- Nếu $A \in L_T$ và $B \in L_T$ thì $S(A, B)$ và $U(A, B) \in L_T$.

Định nghĩa 4. ([2]) Ta gọi dòng thời gian là một cặp $F = (T, <)$ trong đó T là không gian các thời điểm một chiều không âm, trên T có xác định quan hệ thứ tự " $<$ ". Các tính chất của F được Burgess [2] đề nghị gồm a), b), c), d) :

a) Không phản xạ: $\neg\exists t \in T$ sao cho $t < t$.

b) Bắc cầu: $\forall s, t, u \in T$, nếu $s < t$ và $t < u$ thì $s < u$.

c) Đầy đủ: $\forall s, t \in T$, hoặc $s < t$ hoặc $s = t$ hoặc $t < s$.

Nếu T là tập Z_+ thì thỏa mãn thêm tính chất e) còn nếu T là Q_+ hoặc R_+ thì thỏa mãn thêm tính chất f)

d) Mở rộng: $\forall t \in T, \exists u, s \in T$ sao cho $u < t < s$.

e) Rời rạc: $\forall t \in T$, nếu $\exists s \in T, t < s$ thì $\exists t' \in T$ sao cho $t < t'$ và $\neg\exists u \in T$ sao cho $t < u < t'$; và nếu $\exists s \in T, s < t$ thì $\exists t" \in T$ sao cho $t" < t$ và $\neg\exists u \in T$ sao cho $t" < u < t$.

f) Trù mật: $\forall s, t \in T$, sao cho $s < t$ thì $\exists u \in T$ sao cho $s < u < t$.

Định nghĩa 5. ([3]) (Ngữ nghĩa của các mệnh đề P_{TL}) Cho hàm theo thời gian $g : T \to P_{TL}$. Ta gọi mô hình M là bộ ba được xác định $M = (T, <, g)$.

Ký hiệu: $M, t \models A$ là công thức A xác định trên mô hình M tại thời điểm t, một vài định nghĩa tương tự như trên:

$M, t \models p$ nếu $p \in P$ thì $p \in g(t)$.

$M, t \models \neg A$ nếu không phải là $M, t \models A$.

$M, t \models A \wedge B$ nếu $M, t \models A$ và$M, t \models B$.

$M, t \models S(A, B)$ nếu $\exists s \in T$ với $s < t$ và $M, s \models A$ và $\forall u \in T$, nếu $s < u < t$ thì $M, u \models B$.

$M, t \models U(A, B)$ nếu $\exists s \in T$ với $t < s$ và $M, s \models A$, $\forall u \in T$, nếu $t < u < s$ thì $M, u \models B$.

Gọi lớp gồm các mô hình của L là K. Công thức A là đúng trên lớp K theo dòng thời gian, ký hiệu là $K \models A$ nếu $\forall$ mô hình M theo dòng thời gian trong K và $\forall t \in T$, ta có $M, t \models A$.

Nếu Σ là tập các công thức, ta viết $K \models \Sigma$ nếu $K \models A, \forall A \in \Sigma$. Như vậy, với các lớp K khác nhau ta có các tập công thức khác nhau.

Dựa vào hệ tiên đề của Hilbert [4] đã mở rộng thành các tiên đề có tính thời gian được phát biểu dưới dạng các luật dẫn như sau:

A1a $G(p \to q) \to (U(p, r) \to S(q, r))$.

A1b $H(p \to q) \to (S(p, r) \to S(q, r))$.

A2a $G(p \to q) \to (U(r, p) \to U(r, q))$.

A2b $H(p \to q) \to (S(r, p) \to S(r, q))$.

A3a $(p \wedge U(q, r)) \to U(q \wedge S(p, r), r)$.

A3b $(p \wedge S(q, r)) \to S(q \wedge U(p, r), r)$.

MP modus ponens: Nếu có A và $A \to B$ thì B.

TG Temporal Generalisation: Nếu có A thì có HA và GA.

A4 $U(q, p) \to U(p, q \wedge U(p, q))$.

A5a $U(q \wedge U(p, q), q) \to U(p, q)$.

A6 $U(q, p) \wedge U(r, s) \to (U(p \wedge r, q \wedge s) \vee U(p \wedge s, q \wedge s) \vee U(p \wedge r, q \wedge s)$.

2. THAO TÁC TRUY VẤN TRONG CSDL CÓ TÍNH THỜI GIAN

2.1. Ngữ nghĩa của truy vấn có tính thời gian

Định nghĩa 6. ([5]) Ta gọi một ký số (Database Signature) của CSDL là một cặp $S = (S_c, S_p)$ với S_c là tập đếm được các ký hiệu hằng và S_p là tập các vị từ. Ký số được gọi là hữu hạn nếu S_p là hữu hạn. Cho V là tập vô hạn các biến. Từ (Term) là thuật ngữ chỉ biến hoặc hằng. Ta gọi một công thức nguyên tử là công thức có dạng $(a_1 = a_2) \wedge p(a_1, a_2 ... a_r)$ với r là số biến, a_i là các từ và p là vị từ. Tập các biến tự do (Free) của công thức A ta gọi là $Fr(A)$. Các tính chất trên các biến tự do của công thức trên được phát biểu:

- Với mỗi công thức nguyên tử thì tập tất cả các biến đều là biến tự do.
- Với A, B thì $\neg A, A \wedge B, A \vee B$ sẽ có các tập biến tự do tương ứng là $Fr(A)$, $Fr(A) \cap Fr(B), Fr(A) \cup Fr(B)$.
- Với A, B thì $U(A, B)$ và $S(A, B)$ có tập các biến tự do là $Fr(A) \cup Fr(B)$.

Ví dụ 1. $S = (S_c, S_p)$, với $S_c = \{$Chuỗi các ký tự$\}$, $S_p = \{$Nhân viên$\}$.

Xét vị từ: Nhân viên (Họ tên, Bậc lương, Phòng). Ta có một số công thức như sau:

Nhân viên (Peter, 2000, Kinh doanh);

Nhân viên (x, 2000, Kinh doanh) $\wedge \neg(x =$ John);

P Nhân viên $(x, y,$ Tài vụ) $\vee$ Nhân viên $(x, y,$ Kinh doanh).

Trong công thức thứ nhất không có biến tự do, công thức thứ hai có một biến và công thức thứ ba có hai biến tự do.

Định nghĩa 7. Ta gọi cấu trúc của ký số $S = (S_c, S_p)$ là cặp (D, I) trong đó D là tập vô hạn đếm được gọi là miền, I là tập các thể hiện của các hằng, $I(c) \in D$ và $I(p) \subseteq D^r$ với r là số biến của p.

Ta gọi một phép gán toàn cục v là một ánh xạ ứng với mỗi biến $x \in V$ đến giá trị $v(x) \in D$, phép gán này là độc lập với thời gian.

Ta ký hiệu $D, v, t \models A$ nghĩa là công thức A là True tại thời điểm t tương ứng với phép gán v. Bằng cách suy diễn từ định nghĩa trên ta có các công thức sau:

$D, v, t \models p(a_1, a_2...a_r)$	nếu $(v(a_1), v(a_2), \cdots, v(a_r)) \in I_t(p)$.
$D, v, t \models \neg A$	nếu không phải là $D, v, t \models A$.
$D, v, t \models A \wedge B$	nếu $D, v, t \models A$ và $D, v, t \models B$.
$D, v, t \models A \vee B$	nếu $D, v, t \models A$ hoặc $D, v, t \models B$.
$D, v, t \models S(A, B)$	nếu $\exists s \in T$ với $s < t$ và $D, v, s \models A$ và $\forall u \in T$, nếu $s < u < t$ thì $D, v, u \models B$.
$D, v, t \models U(A, B)$	nếu $\exists\, s \in T$ với $t < s$ và $D, v, s \models A$ và $\forall u \in T$, nếu $t < u < s$ thì $D, v, t \models B$.

Ta gọi biên của CSDL có tính thời gian là $t_{\min}$ và $t_{\max}$ là các giá trị mà $\forall s < t_{\min}$ và $u > t_{\max}$ thì $g(s) = g(t_{\min})$ và $g(t) = g(t_{\max})$.

Ví dụ 2. Giả sử CSDL được cho trong ví dụ 1. Ta gọi S_c là tập các chuỗi, số, phép nối chuỗi,...; các thời điểm là các số nguyên chỉ các tháng, phép $I_t(c) = c$.

- (Peter, 2000, Kinh doanh) $\in I_t$ (Nhân viên) Nếu $1/1999 \leq t \leq 12/2001$.
- (Peter, 3000, Tài vụ) $\in I_t$ (Nhân viên) Nếu $1/2002 \leq t \leq 4/2002$.
- (John, 2000, Kinh doanh) $\in I_t$ (Nhân viên) Nếu $2/1999 \leq t \leq 12/2001$.
- (Mary, 3000, Hành chính) $\in I_t$ (Nhân viên) Nếu $9/2000 \leq t \leq 4/2003$.
- Không còn gì khác trong $\in I_t$ (Nhân viên).

Ta có $t_{\min} = 12/1998$ và $t_{\max} = 5/2003$ và giả sử rằng phép gán v là $v(x) = Peter$ và $v(y) = 2003$ thì $D, v, 4/2003| \not\models P\exists y\ \exists z$ Nhân viên $(x, y, x) \wedge \neg\ \exists y_1\ \exists z_1$ Nhân viên (x, y_1, z_1) vì Peter là hiện tại là Nhân viên, tuy nhiên $D, v, 4/2003 \models P$ Nhân viên (x, y, Kinh doanh) $\vee$ Nhân viên (x, y, Kinh doanh) vì Peter là Nhân viên Phòng Kinh doanh từ 1/1999 đến 12/2001.

2.2. Tính tin cậy của truy vấn

Định nghĩa 7. ([10]) Ta gọi phần tử $d \in D$ là thích hợp với vị từ $p \in S_p$ nếu thỏa trong $I_t(p)$ với $t \in Z$.

Phần tử d là thích hợp với công thức A nếu d là hằng trên A hoặc d thích hợp với các vị từ nào đó trên A.

Gọi $R_D \subset D$ là tập tất cả các phần tử thích hợp với các vị từ trong CSDL, ta thấy R_D là tập gồm hữu hạn phần tử.

Ta gọi một công thức độc lập miền là công thức $A(x_1, x_2, ..., x_n)$ gồm n quan hệ chứa các phần tử của miền sao cho chúng là thích hợp. Nghĩa là $\forall t$ thì tập các bộ của các phần tử trên miền $(v(x_1), v(x_2), ..., v(x_n))$ có dạng $D, v, t \models A(x_1, x_2, ..., x_n)$ là hữu hạn và chỉ gồm các phần tử thích hợp với $A(x_1, x_2, ..., x_n)$.

Điều kiện lý tưởng của truy vấn là các công thức độc lập trên miền, tuy nhiên, theo Ullman [11] thì không thể xác định chắc chắn rằng trong trường hợp tổng quát một công thức là độc lập trên miền hay không.

Sau đây, chúng ta sẽ xét khái niệm truy vấn tin cậy theo thời gian dựa vào sự giới nội của các biến độc lập trong các phép $\vee, \wedge, \neg$ theo thời gian.

Nhận xét. có một số biểu thức hoặc luật tạo ra các quan hệ vô hạn, chẳng hạn: Lớn-hơn $(X, Y) : -X > Y$ Nếu X và Y có miền giá trị là số nguyên hoặc là một tập vô hạn nào sẽ tạo thành một quan hệ vô hạn.

Với mục đích tránh các công thức tạo ra các quan hệ vô hạn, tính giới nội của các biến được định nghĩa như sau ([9]):

(i) Tất cả các biến trong vị từ là giới nội.

(ii) Mỗi biến X có dạng $X = a$ với a là const thì X là giới nội.

(iii) Biến X là giới nội nếu $X = Y$ với Y là giới nội.

Một luật hoặc công thức là tin cậy nếu tất cả các biến đều là giới nội.

Trong ([7]), đối với phép tuyển $A \vee B$, chỉ đơn giản đòi hỏi A và B chia sẻ cùng chung số biến độc lập, như vậy nếu công thức $A(x) \vee B(x)$ là không tin cậy với $A(x)$ tại giá trị $x_0 \in D$, thì sẽ có vô số cặp $(x_0, d), d \in D$ thỏa mãn truy vấn.

Đối với phép phủ định $\neg$ thì tính giới nội của các biến tương tự như giới nội trên miền, nghĩa là tất cả các biến độc lập có xuất hiện bên trong dấu phủ định thì phải xuất hiện trong một khẳng định ngoài phủ định này. Chẳng hạn, công thức $\neg A(x, y)$ là không tin cậy, nhưng $\neg A(x, y) \wedge B(x) \wedge C(y)$ là tin cậy.

Khái niệm này được mở rộng trên công thức có tính thời gian TF như sau: Công ty TF có dạng $S(A(x), B(y))$ là không tin cậy nếu $A(x)$ là True tại một thời điểm trước đó với các giá trị $x_0 \in D$, thì có vô số cặp $(x_0, d), d \in D$, thỏa mãn công thức sau: $S(A(x) \wedge B(y), B(y))$, tuy nhiên điều này không xảy ra đối với công thức tin cậy.

Tương tự, các công thức mở rộng TF cho $S(A, B)$ và $U(A, B)$ như sau: Các công thức TF dạng $S(A, B)$ và $U(A, B)$ là tin cậy nếu:

(i) Tất cả các biến tự do là giới nội trong A thì cũng giới nội trong $S(A, B)$ và $U(A, B)$,

(ii) Tất cả các biến tự do là giới nội trong B thì cũng giới nội ngoài B.

Chẳng hạn, biến y là không giới nội ngoài $B(y)$ trong công thức không tin cậy $S(A(x), B(y))$, nhưng nó giới nội trong công thức tin cậy như $S(A(x) \wedge B(y), B(y))$ hoặc $S(A(x), B(y)) \wedge C(y)$.

Định nghĩa 9. Một công thức TF là tin cậy khi:

a) Nếu chứa các công thức con dạng tuyển thì các công thức con này phải tin cậy và có cùng biến độc lập. Nghĩa là, có dạng $B_1(x_1, x_2, \cdots, x_n) \vee B_2(x_1, x_2, \cdots, x_n)$.

b) Nếu chứa các công thức con dạng hội $B_1 \wedge \cdots \wedge B_m$, thì tất cả các biến tự do xuất hiện trong B_i phải giới hạn theo các trường hợp sau:

- Một biến là giới nội nếu nó chỉ có trong vài B_i nào đó mà các B_i này không là phủ định và không phụ thuộc vào thời gian.
- Nếu B_i có dạng $x = c$(const) thì x là giới nội.
- Nếu B_i có dạng $x = y$ với y là giới nội thì x cũng là giới nội.
- Nếu B_i có dạng $S(A, C)$ hoặc $U(A, C)$ thì tất cả các biến là giới nội trong A thì cũng giới nội trong B_i.

c) Nếu chứa công thức dạng $\neg B(x_1, x_2, \cdots, x_k)$ thì phải thỏa b) nghĩa là $\neg B(x_1, x_2, \cdots, x_k)$ phải là có dạng hội hoặc là tất cả các x_i giới nội.

d) Nếu chứa công thức dạng $S(A, B)$ hoặc $U(A, B)$ thì phải thỏa b) nghĩa là $S(A, B)$ hoặc $U(A, B)$ phải là có dạng hội hoặc là tất cả các biến tự do là giới nội.

Định nghĩa 10. Một truy vấn TQ (Temporal Query) $Q = \{x_1, x_2, \cdots, x_m; t | A(x_1, x_2, ..., x_m)\}$ là tin cậy nếu $A(x_1, x_2, ..., x_m)$ là công thức tin cậy với các biến tự do $x_1, x_2, ..., x_m$.

Một quan hệ được tạo ra bởi truy vấn Q trên CSDL D là

$$\{(v(x_1), ..., v(x_m)) \text{ với } v \text{thỏa } D, v, t \models A(x_1, x_2, ..., x_m)\}$$

và mỗi bộ phận của quan hệ được tạo ra thỏa mãn truy vấn Q tại thời điểm t.

Một truy vấn TQ là độc lập trên miền nếu với mỗi giá trị $t \in Z$, nó tạo ra chỉ một quan hệ gồm hữu hạn phần tử.

Định lý 1. *Các truy vấn TQ tin cậy là độc lập trên miền.*

Chứng minh: Ullman đã chứng minh rằng mỗi công thức non-temporal với các biến giới nội thì tạo ra quan hệ chỉ gồm hữu hạn phần tử trên miền thích hợp, như vậy sẽ thực hiện các phép tuyển tin cậy với mỗi thời điểm t bất kỳ.

Với nhận xét về tính tin cậy của các công thức con trong công thức TF là $S(A, B)$ và $U(A, B)$, A chỉ tạo ra các quan hệ gồm hữu hạn bộ đối với các phần tử thích hợp trên miền tại bất kỳ thời điểm t và B bao gồm tất cả các biến tự do giới nội. Do đó CSDL là bị chặn theo thời gian và chỉ có hữu hạn các bộ của các phần tử trên miền (ứng với biến tự do) là có thể thỏa mãn $S(A, B)$ và $U(A, B)$ tại mọi thời điểm bất kỳ.

Vì vậy, các công thức con này chỉ có thể tạo ra hữu hạn các quan hệ chỉ gồm hữu hạn phần tử trên miền thích hợp, do đó công thức TF tin cậy là độc lập với miền. ■

Nhận xét. Định lý trên không thể có điều ngược lại, có nghĩa là có thể có các truy vấn không tin cậy những vẫn độc lập miền. Phản ví dụ sau:

$$\text{Công thức } A(x, y, z) \wedge \neg(FB(x, y) \vee C(y, z))$$

là không tin cậy nhưng nó sẽ tạo ra những hữu hạn bộ trên R_D trong CSDL, bởi vì công thức trên còn được viết lại $A(x, y, z) \wedge \neg FB(x, y) \vee \neg C(y, z))$.

Ví dụ 3. Trong CSDL trên, chúng ta muốn biết tên của các nhân viên đã nghỉ việc trong quá khứ, ví dụ như tháng 4/2002, ta có công thức truy vấn tin cậy như sau:

$$\{x, 4/2002 | P\exists\, y\exists\, z \text{ Nhân viên } (x, y, z) \wedge \neg\exists w \exists v \text{ Nhân viên}(x, w, v)\}$$

sẽ tạo ra một quan hệ gồm có một thể hiện là {(John)}.

Nếu muốn biết Tên, Mức lương của các nhân viên đã từng làm việc tại phòng Kinh doanh, ta có truy vấn tin cậy sau:

$$\{x, y; 4/2002 | P \text{ Nhân viên } (x, y, \text{Kinh doanh}) \;\; \text{Nhân viên}(x, y, \text{Kinh doanh})\}$$

sẽ tạo ra quan hệ gồm {(Peter 2000), (John, 2000)}.

3. CÁC PHỤ THUỘC VÀ THAO TÁC CẬP NHẬT DỮ LIỆU

3.1. Giới thiệu

Định nghĩa 11. ([4]) Một luật TR (Temporal Rule) là công thức có dạng:

$$\langle \text{ Biểu thức điều kiện}(x_1, x_2, ..., x_m) \rangle \;\; \rightarrow \;\; \langle \text{Thao tác } (y_1, y_2, ..., y_m) \rangle \text{ , hay } C_i \;\rightarrow\; A_i,$$

trong đó, biểu thức điều kiện là Công thức TQL (Temporal Query Formula) với các biến độc lập $x_1, x_2, ..., x_m$ và thao tác $(y_1, y_2, ..., y_m)$ là cập nhật nào đó xác định.

Định nghĩa 12. ([4]) Cho M là mô hình CSDL có tính thời gian hai chiều. Thời điểm giao tác x (transaction time) và thời điểm hợp lệ t (valid time), thời điểm chứa các luật TR như sau $C_i \rightarrow A_i$.

Ta gọi tập hữu hạn:

$$A(t) = \{[A_i^v]\} \text{ trong đó } i, v \text{ thỏa mãn } M, t, t \models [C_i^v].$$

Ngữ nghĩa của luật trên là nếu $A(t)$ thỏa mãn thao tác thì thao tác tại thời điểm tiếp theo cũng phải thỏa mãn, nghĩa là:

$$M, t, t \models C_i \text{ thì } M, t, t+1 \models A_i.$$

Giả sử rằng trong mô hình này chúng ta chỉ có các luật dạng "từ quá khứ đến hiện tại" và các thao tác cập nhật dựa vào các luật không làm thay đổi các dữ liệu trong quá khứ. Vì vậy, chúng ta phải lưu trữ các điều kiện vế trái (ta gọi là hệ sinh của luật) và các thao tác đã thực hiện. Như vậy tính hợp lệ của các luật theo là quan trọng. Tuy nhiên có thao tác được tạo ra không phải từ hệ sinh. Sau khi thực hiện một luật về cập nhật, có thể tạo ra thay đổi các hệ sinh của nó, xóa đi một thao tác đã ghi trong CSDL, mà các thao tác này sự tồn tại của nó là không cần thiết. Nói khác, các điều kiện C_i là không còn đúng.

Gọi M là mô hình CSDL có tính thời gian đang xét, ta nói rằng thao tác $\{[A^v]\}$ là không được sinh ra từ hệ sinh tại thời điểm thao tác x (transaction time) và thời điểm hợp lệ t (valid time) nếu thao tác này được thực hiện tại thời điểm $t < x$ ký hiệu $M, t, t \models \{[C^v]\}$.

Cho đến thời điểm giao tác $(x-1)$ được tạo ra từ hệ sinh và các thao tác đã ghi từ các luật vẫn còn hợp lệ có nghĩa là với $t < y \Leftarrow x-1, M, t, y \not\models \{[(C^v])$ và $M, t, y \not\models \{[(A^v])$.

Nhưng sau khi thực hiện cập nhật Y_{x-1}, tại thời điểm thao tác x thì $M, t, x \models \{[(C \wedge \neg A)^v]\}$. Như vậy, một cập nhật nào đó trong quá khứ có thể là nguyên nhân tạo ra thao tác không phải từ hệ sinh trong CSDL, như từ khi hệ sinh có 1 luật không hạn chế chỉ trong quá khá thì bất kỳ cập nhật nào cũng có thể là nguyên nhân gây ra chúng.

Ví dụ 4. Giả sử trong CSDL trong Ví dụ 1

Thao tác trả lương (Mary, 3000) được thực hiện từ tháng 9/2000 đến tháng 4/2002. Giả sử tại thời điểm tháng 4/1999 lương của Mary được tăng lên 5000 nhưng việc tăng lương này có hiệu lực từ tháng 1/2002.

Điều này cho phép CSDL thực hiện thao tác Trả lương(Mary,3000), thao tác này được tạo ra từ hệ sinh từ tháng 2/2002 đến 4/2002. Như vậy, để tránh các mâu thu việc cập nhật trong CSDL thì thời gian hợp lệ của CSDL phải trong thời gian thao tác của CSDL.

3.2. Thao tác cập nhật trong mô hình có tính thời gian

Với d_+, d_- là tập các thời điểm t, x là thời điểm hiện hành, $M_x = (T, <, g)$ là mô hình CSDL, có tính thời gian ứng với thời điểm x.

Ta gọi phụ thuộc thời gian là tại thời điểm t suy dẫn biểu thức A là công thức có dạng:

$$M_x \models (t, d_+, d_-) : A,$$

nghĩa là biểu thức A có giá trị là True trong M_x tại thời điểm t, ứng với các tập thời gian điểm d_+ và False trên tập phủ định d_-.

Chú ý: $M_x \not\models (t, d_+, d_-) : A$ thì không suy ra $M_x \models (t, d_+, d_-) : \neg A$, nghĩa là nếu A là False tại thời điểm t ứng với các tập phụ thuộc có thời gian trên d_+ và trên d_-, nhưng là True đối với các tập phụ thuộc có thời gian khác, nói khác $\neg A$ là True trên tập d_+, d_- khác.

Định lý 2. ([3]) *Điều kiện cần và đủ để tồn tại các tập phụ thuộc thời điểm d_+, d_- sao cho*

$$M_x \models (t, d_+, d_-) : A \text{ là } M_x, t : A.$$

Nhận xét: Khi có thao tác cập nhật có tình chất chung thì một số thời điểm có hiệu ứng. Gọi Y_x là cập nhật xác định bởi cặp (υ_+, υ_-). Tập hợp các thời điểm có hiệu ứng xác định bởi việc cập nhật được, ký hiệu:

$$Af_+(Y_x) = \{t | t : q \in \upsilon_+\}.$$

Tương tự, tập hợp các thời điểm có hiệu ứng phủ định được định nghĩa:

$$Af_-(Y_x) = \{t | t : q\ \varepsilon\ \upsilon_-\}.$$

Bổ đề 1. ([3]) *Với mỗi thời điểm giao tác x, mỗi thời điểm hợp lệ t, nếu công thức A thỏa:*
i) $M_x \models (t, d_+, d_-) : A$,
ii) $Af_+(Y_{x-1}) \cap d_- = Af_-(Y_{x-1}) \cap d_+ = \phi$,
thì $M_{x-1} \models (t, d_+, d_-) : A$.

Nhận xét: Bằng cách kết hợp các phụ thuộc có tính thời gian và các hiệu ứng có tính thời gian để có thể xác định thao tác nào là không được tạo ra từ hệ sinh.

Định lý 3. (Các thao tác được tạo ra không phải từ hệ sinh) *Cho M là mô hình CSDL có tính thời gian gồm các luật có dạng*

$$C \rightarrow A \text{ sao cho } M_{x-1} \models (t, d_+, d_-) : C.$$

Khi đó điều kiện cần để một thao tác trở thành thao tác không phải từ hệ sinh tại thời điểm giao tác x và thời điểm hợp lệ t là

$$[Af_+(Y_{x-1}) \cap d_-] \cup [Af_-(Y_{x-1}) \cap d_+] \neq \phi.$$

Chứng minh: Giả sử rằng $M_{x-1} \models (t, d_+, d_-) : C$, khi đó theo Bổ đề 1 ta có $M_{x-1}, t \models C$.

Nếu $Af_+(Y_{x-1}) \cap d = Af_-(Y_{x-1}) \cap d_+ = \phi$, khi đó theo Bổ đề 2, ta có $M_x \models (t, d_+, d_-) : C$ và như vậy $M_x, t \models C$ (!) điều này mâu thuẫn với giả thiết là thao tác được tạo ra không phải từ hệ sinh tại thời điểm giao tác x và thời điểm hợp lệ t.

4. KẾT LUẬN

Trong phần trên chúng ta đã xét truy vấn là tin cậy và thao tác cập nhật nào được tạo ra không phải từ hệ sinh. Vấn đề tiếp tục nghiên cứu là phải dò tìm các thao tác có thể tạo nên các nghịch lý nhưng không phải tạo ra từ hệ sinh, chẳng hạn như do việc cập nhật từ trường hợp quá khứ nào khác gây nên.

Lời cám ơn. Xin chân thành cảm ơn GS. Nguyễn Đình Ngọc, PGS. Nguyễn Xuân Huy đã có những gợi ý định hướng quan trọng và người phản biện đã có những nhận xét, góp ý cho bài báo được hoàn thành.

TÀI LIỆU THAM KHẢO

[1] J. P. Burgess, *Basic Tense Logic*, Handbook of Philosophical logic, Reidel Publishing Company, 1984.

[2] M. Finger, D. M. Gabbay, Adding a Temporal Dimension to a logic System, *Journal of Logic Language anh Information* (1993).

[3] D. M. Gabbay, P. Mc Bien, Temporal Logic and Historical Database, *In 17th Conference on Very Large Database* (1994).

[4] D. M. Gabbay, I. M. Hodkinson, M. A. Reynolds, Temporal Logic - Mathematical Foundations and Computational Aspects (to appear in OUP).

[5] S. Y. Liao, H. Q. W. Y. Wang, Stability Constraints and Stability Normal Forms for Temporal Relational Database, *Information and Software Tecnology* (1999).

[6] J. W. Lloyd, *Foundations of Logic Programming* Springer-Verlag, 1987.

[7] Nguyễn Đình Thuân, Các khái niệm và các dạng chuẩn trong CSDL, Quan hệ có tính thời gian. "Báo cáo tại Hội tho Quốc gia về Công nghệ Thông tin. Hải Phòng" tháng 6/2001.

[8] J. D. Ullman, *Database And Knowledge-Base Systems,* Computer Science Press, 1988.

Nhận bài ngày 20 - 11 -2002
Nhận lại sau sửa ngày 20 - 12 -2003