

ĐẠI SỐ QUAN HỆ VÀ QUAN ĐIỂM SỬ DỤNG NULL VALUE TRÊN MỘT MÔ HÌNH CƠ SỞ DỮ LIỆU MỜ

HỒ THUẦN, HỒ CẨM HÀ

Abstract. In this paper we propose a new approach to extending the relational database model. This approach is based on the concept of similarity based fuzzy relational database and somewhat of new viewpoint on redundancy. It is shown that, in such an extended database model, we can capture imprecise, uncertain information. The paper also shows an extension of relational operators to the new model and some rules for updates on database, that containing null value.

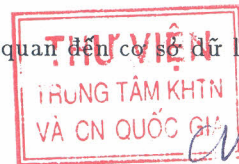
Tóm tắt. Bài báo đề xuất một cách tiếp cận mới để mở rộng mô hình cơ sở dữ liệu quan hệ. Cách tiếp cận này dựa trên khái niệm cơ sở dữ liệu mờ tương tự và một quan điểm mới về dư thừa dữ liệu. Với mô hình cơ sở dữ liệu như vậy có thể nắm bắt được những thông tin không chính xác, không chắc chắn. Bài báo cũng đưa ra một cách xây dựng đại số quan hệ mở rộng cho mô hình mới này và các qui tắc cập nhật dữ liệu trong tình huống chấp nhận null value.

1. GIỚI THIỆU

Như chúng ta đã biết, mô hình quan hệ mặc dù không phải là mô hình được sử dụng trong các DBMS đầu tiên, cũng không phải là một mô hình mới được đề xuất gần đây, nhưng lại đóng vai trò quan trọng và được sử dụng rất phổ biến kể từ khi Codd đưa ra vào năm 1970. Tuy nhiên những hệ thống như vậy chỉ chấp nhận dữ liệu chính xác. Trên thực tế, thông tin về thế giới thực cần xử lý phần nhiều lại là thông tin không đầy đủ, không chắc chắn. Trên một thuộc tính, thông tin về một đối tượng có thể hoàn toàn không biết, hoặc cũng có thể không biết một cách đầy đủ (tức là chỉ biết một tập các giá trị có thể) và có thể nó là thông tin không chắc chắn. Có những trường hợp một vài thuộc tính không thể áp dụng được cho một số đối tượng nằm trong phạm vi xem xét và trong một số tình huống thậm chí chúng ta còn không biết giá trị về thuộc tính đó tồn tại hay không. Có nhiều đề xuất về cách tiếp cận nhằm giải quyết vấn đề này. Việc dùng lý thuyết mờ để mở rộng mô hình cơ sở dữ liệu đã được nhiều nhà nghiên cứu quan tâm. Một trong số đó là P. Buckles và E. Petry. Trong [1] mô hình cơ sở dữ liệu quan hệ được mở rộng để biểu diễn thông tin không chính xác. Mô hình này khác với cơ sở dữ liệu quan hệ truyền thống ở hai điểm quan trọng: giá trị tại mỗi thuộc tính của một đối tượng có thể là một tập và trên mỗi một miền của thuộc tính có một quan hệ mờ thể hiện cấp độ tương tự giữa các phần tử của miền. Trong mô hình này, tuy giá trị của một bộ tại mỗi thuộc tính có thể chứa một hay nhiều phần tử của miền tương ứng, nhưng có một ràng buộc là các phần tử trong cùng một giá trị thuộc tính (của cùng một đối tượng) phải đủ "tương tự" với nhau (nghĩa là cấp độ tương tự của một cặp bất kỳ các phần tử không nhỏ hơn ngưỡng tương tự đã cho).

Cách tiếp cận trong việc mở rộng một cơ sở dữ liệu quan hệ của chúng tôi cũng dựa trên một quan hệ mờ thể hiện khái niệm tương tự trên mỗi miền nhưng thực sự khác với P. Buckles và E. Petry ở chỗ các phần tử của mỗi giá trị thuộc tính không bị đòi hỏi đủ tương tự theo ngưỡng. Ý tưởng này cho phép mỗi giá trị thuộc tính chứa các phần tử biểu diễn những khả năng rất khác xa nhau có thể xảy ra (bởi các giá trị không hề tương tự). Như vậy khi mô hình hóa một cơ sở dữ liệu quan hệ theo cách tiếp cận này không chỉ cho phép nắm bắt thông tin không chính xác mà cả những thông tin không chắc chắn.

Bài báo được trình bày trong 6 phần. Những khái niệm cơ sở liên quan đến cơ sở dữ liệu mờ



và quan hệ mờ thể hiện khái niệm tương tự được nhắc lại ở Phần 2 để thống nhất thuật ngữ được dùng. Một quan niệm mới về dư thừa dữ liệu và cách loại bỏ các bộ thừa được trình bày ở Phần 3. Phần 4 đề xuất một đại số quan hệ cho mô hình dữ liệu này. Phần 5 trình bày các qui tắc cập nhật dữ liệu trong trường hợp có null value. Một số vấn đề có thể tiếp tục nghiên cứu thêm được nêu ở phần cuối của bài báo.

2. CÁC KHÁI NIỆM CƠ BẢN

Quan hệ tương tự (similarity relation) được mô tả bởi định nghĩa của Zadeh [9]. Quan hệ tương tự được dùng để thể hiện cấp độ tương tự của hai phần tử thuộc cùng một miền.

Định nghĩa 2.1. [6] Một quan hệ tương tự $s_D(x, y)$, trên một miền D , là một ánh xạ mọi cặp phần tử của miền vào khoảng đóng $[0, 1]$ thỏa 3 tính chất sau với mọi $x, y, z \in D$:

1. Phản xạ $s_D(x, x) = 1$.

2. Đối xứng $s_D(x, y) = s_D(y, x)$.

3. Bất cầu $s_D(x, z) \geq \max_y(\min[s_D(x, y), s_D(y, z)])$, (T1)

(hoặc 3'. Bất cầu $s_D(x, z) \geq \max_y(|s_D(x, y) * s_D(y, z)|)$, (T2)

trong đó $*$ là phép nhân số học).

Mỗi một thuộc tính thứ j trong cơ sở dữ liệu quan hệ có một miền trị D_j tương ứng. Các miền cho cơ sở dữ liệu mờ ở đây hoặc là một tập vô hướng rời rạc hoặc tập số rời rạc lấy từ một tập vô hạn hay hữu hạn. Một giá trị thuộc tính, d_{ij} trong đó i là chỉ số của bộ thứ i , được định nghĩa là một tập con không rỗng của miền tương ứng D_j . Dùng kí hiệu 2^{D_j} để chỉ tập tất cả các tập con khác rỗng của D_j .

Định nghĩa 2.2. [1] Một quan hệ mờ, r , là một tập con của tập tích Đề-cac $2^{D^1} \times \dots \times 2^{D^m}$.

Định nghĩa 2.3. [1] Một bộ t của một quan hệ mờ là một phần tử của tập $2^{D^1} \times \dots \times 2^{D^m}$. Một cách tổng quát, một bộ $t_i \in r$ có dạng: $t_i = (d_{i1}, d_{i2}, \dots, d_{im})$, $d_{ij} \subseteq D_j$.

Ví dụ. Một bộ t trong một quan hệ mờ

Tên	Nghề nghiệp	Bệnh theo chuẩn đoán
{Lan}	{kỹ sư hóa, dược sĩ, ca sĩ}	{viêm họng, viêm amidan, sốt siêu vi trùng}

3. QUAN NIỆM VỀ DƯ THỪA VÀ LOẠI BỎ CÁC BỘ THỪA

Trong một cơ sở dữ liệu rõ, một bộ được coi là thừa nếu và chỉ nếu nó trùng hoàn toàn với một bộ khác. Theo quan điểm của P. Buckles và E. Petry [1], một bộ là thừa nếu có thể trộn nó với một số bộ khác mà vẫn không vi phạm ngưỡng tương tự đã cho. Mặc dù cho phép giá trị của một bộ tại mỗi thuộc tính là một tập chứ không phải là một giá trị đơn nhất như trong cơ sở dữ liệu truyền thống, nhưng ràng buộc ở đây là các phần tử thuộc cùng giá trị thuộc tính phải đủ tương tự với nhau (nói một cách khác độ tương tự của chúng với một giá trị đơn phải vượt một ngưỡng đã cho). Theo một nghĩa nào đó, nếu chúng ta coi các phần tử đủ tương tự với nhau (theo ngưỡng cho biết) thuộc về cùng một nhánh khả năng có thể xảy ra thì mô hình của P. Buckles và E. Petry [1] chỉ cho phép nắm giữ thông tin của những đối tượng mà với những đối tượng này thông tin biết được về mỗi thuộc tính chỉ thuộc về một nhánh khả năng (tương tự của một đơn giá trị). Tuy nhiên trên thực tế có thể gặp những thông tin không chắc chắn về một đối tượng mà trên một thuộc tính có thể xảy ra nhiều nhánh khả năng. Trong ví dụ trên Lan có thể là dược sĩ, kỹ sư hóa (đại loại thuộc ngành làm việc với hóa chất) nhưng cũng có thể Lan là một ca sĩ. Có thể Lan đang bị viêm họng hoặc viêm amidan và cũng không loại trừ khả năng Lan bị cả 3 loại bệnh đã nêu. Nếu coi một tập

các nhánh khả năng có thể xảy ra và cần được lưu giữ vì nó xác định một thông tin đầy đủ hơn, thì mô hình ở [2] cần được mở rộng. Chúng tôi đã thử làm điều đó.

Giả sử với mỗi D_j có một ngưỡng α_j cho một quan hệ tương tự (với T1 bắc cầu) vốn được xác định trên miền này. Hai bộ được gọi là thừa với nhau nếu chúng có cùng một tập nhánh khả năng trên mỗi thuộc tính. Có thể hình thức hóa điều đó như sau.

Với mỗi ngưỡng α_j của miền D_j , kí hiệu $\text{THRES}(D_j) = \alpha_j$ hay $\text{LEVEL}(D_j) = \alpha_j$. Với $x, y \in D_j$, nếu $s(x, y) \geq \alpha_j$ thì chúng ta viết $x \sim_{\alpha_j} y$. Để nhận thấy \sim_{α_j} là một quan hệ hai ngôi trên D_j và rõ ràng là:

Bổ đề 3.1. \sim_{α_j} là một quan hệ tương đương.

Như vậy quan hệ \sim_{α_j} phân hoạch D_j . Trong lược đồ quan hệ mờ của P. Buckles and E. Petry mỗi giá trị thuộc tính chỉ gồm các phần tử cùng lớp tương đương theo phân hoạch bởi quan hệ \sim_{α_j} và hai bộ là thừa đối với nhau nếu trên mọi thuộc tính, giá trị của hai bộ chỉ chứa các đại diện của cùng một lớp tương đương.

Định nghĩa 3.1. Trong quan hệ mờ r , hai bộ $t_i = (d_{i1}, d_{i2}, \dots, d_{im})$ và $t_k = (d_{k1}, d_{k2}, \dots, d_{km})$ $i \neq k$ được coi là thừa đối với nhau nếu $\forall x \in d_{ij} \exists x' \in d_{kj} : x \sim x', \forall j = 1, 2, \dots, m$ và ngược lại, nghĩa là $\forall x \in d_{kj} \exists x' \in d_{ij} : x \sim x', \forall j = 1, 2, \dots, m$.

Dùng kí hiệu $t_i \approx_{\alpha} t_k$ để nói rằng t_i thừa đối với t_k theo ngưỡng α , trong đó $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_m)$. Để thấy là:

Bổ đề 3.2. \approx_{α} là quan hệ tương đương trên một quan hệ mờ r .

Như vậy quan hệ \approx_{α} cho một phân hoạch duy nhất trên r .

Ví dụ về một quan hệ mờ với các quan hệ tương tự trên các miền

r_1	Tên	Màu xe	Nghề nghiệp
t_1	An	xanh đậm, xanh nhạt, hồng	nhà văn, giáo sư
t_2	Bình	xanh đen, tím đỏ	đạo diễn, giáo viên
t_3	Phúc	trắng, hồng	nhà thơ
t_4	Lộc	hồng, kem	nhà thơ
t_5	Thọ	xanh đen, đỏ	phi công

Hình 1. Một quan hệ mờ

	xanh đậm	xanh nhạt	xanh đen	hồng	đỏ	tím đỏ	trắng	kem
xanh đậm	1,0	0,6	0,8	0,0	0,0	0,0	0,1	0,1
xanh nhạt	0,6	1,0	0,6	0,0	0,0	0,0	0,1	0,1
xanh đen	0,8	0,6	1,0	0,0	0,0	0,0	0,1	0,1
hồng	0,0	0,0	0,0	1,0	0,6	0,6	0,0	0,0
đỏ	0,0	0,0	0,0	0,6	1,0	0,9	0,0	0,0
tím đỏ	0,0	0,0	0,0	0,6	0,9	1,0	0,0	0,0
trắng	0,1	0,1	0,0	0,0	0,0	0,0	1,0	0,7
kem	0,1	0,1	0,1	0,0	0,0	0,0	0,7	1,0

Hình 2. Quan hệ tương tự trên Dom (Màu xe)

	nhà văn	nhà thơ	đạo diễn	giáo viên	giáo sư	phi công
nhà văn	1,0	1,0	0,9	0,5	0,5	0,2
nhà thơ	1,0	1,0	0,9	0,5	0,5	0,2
đạo diễn	0,9	0,9	1,0	0,5	0,5	0,2
giáo viên	0,5	0,5	0,5	1,0	0,8	0,2
giáo sư	0,5	0,5	0,5	0,8	1,0	0,2
phi công	0,2	0,2	0,2	0,2	0,2	1,0

Hình 3. Quan hệ tương tự trên Dom (Nghề nghiệp)

Giả sử rằng ngưỡng $\alpha = (0,0, 0,6, 0,8)$. Khi đó ngưỡng của Dom (Tên) là 0,0, ngưỡng của Dom (Màu xe) là 0,6, ngưỡng của Dom (Nghề nghiệp) là 0,8. Dom (Màu xe) được phân hoạch thành 3 lớp tương đương:

$\{\{\text{xanh đậm, xanh nhạt, xanh đen}\}, \{\text{hồng, đỏ, tím đỏ}\}, \{\text{trắng, kem}\}\}$

và Dom (Nghề nghiệp) được phân hoạch thành:

$\{\{\text{nhà văn, nhà thơ, đạo diễn}\}, \{\text{giáo viên, giáo sư}\}, \{\text{phi công}\}\}$.

Như vậy trong r_1 trên đây, t_1 thừa đối với t_2 và t_3 thừa đối với t_4 .

Để loại bỏ những bộ thừa theo ngưỡng α trong quan hệ r , chúng ta sẽ trộn những bộ thừa lại với nhau cho đến khi không còn tồn tại hai bộ thừa đối với nhau nữa.

Định nghĩa 3.2. Cho một quan hệ mờ r , hai bộ $t_i, t_k \in r$, $t_i = (d_{i1}, d_{i2}, \dots, d_{im})$, $t_k = (d_{k1}, d_{k2}, \dots, d_{km})$. Kết quả của việc trộn hai bộ t_i, t_k là một bộ t sao cho $t = (d_1, d_2, \dots, d_m)$ và $d_h = d_{ih} \cup d_{kh}$, $h = 1, 2, \dots, m$.

Bổ đề 3.3. Việc loại bỏ các bộ thừa (theo một ngưỡng xác định) bằng phép trộn các bộ thừa cho một kết quả duy nhất không phụ thuộc vào thứ tự trộn các bộ.

Như vậy cho một quan hệ r , một ngưỡng tương tự α , có thể đưa ra một r' duy nhất bằng cách loại bỏ các bộ thừa của r . Kí hiệu $r' = M_\alpha(r)$.

Ví dụ. Với quan hệ r_1 cho ở hình 1, $\alpha = (0,0, 0,6, 0,8)$, ta có $M_\alpha(r_1)$ cho ở hình 4.

Tên	Màu xe	Nghề nghiệp
{An Bình}	{xanh đậm, xanh nhạt, xanh đen, hồng, tím đỏ}	{nhà văn, giáo sư, đạo diễn, giáo viên}
{Phúc, Lộc}	{trắng, hồng, kem}	{nhà thơ}
{Thọ}	{xanh đen, đỏ}	{phi công}

Hình 4

4. ĐẠI SỐ QUAN HỆ MỜ

Phép hợp: Cho r_1 và r_2 là hai quan hệ trên cùng một lược đồ R . Hợp theo ngưỡng α của r_1 và r_2 là một quan hệ kí hiệu là $r_1 \cup_\alpha r_2$ được xác định như sau: $r_1 \cup_\alpha r_2 = M_\alpha(r_1 \cup r_2)$.

Phép giao: Cho r_1 và r_2 là hai quan hệ trên cùng một lược đồ R . Giao theo ngưỡng α của r_1 và r_2 là một quan hệ kí hiệu là $r_1 \cap_\alpha r_2$ được xác định như sau:

$$r_1 \cap_\alpha r_2 = M_\alpha(\{t \mid (t \in r_1 \text{ và } \exists t' \in r_2 : t \approx_\alpha t') \text{ hoặc } (t \in r_2 \text{ và } \exists t' \in r_1 : t \approx_\alpha t'))\}.$$

Phép hiệu: Cho r_1 và r_2 là hai quan hệ trên cùng một lược đồ R . Hiệu theo α của r_1 đối với r_2 là một quan hệ kí hiệu là $r_1 -_\alpha r_2$ được xác định như sau:

$$r_1 -_{\alpha} r_2 = M_{\alpha} \{t \in r_1 \mid \forall t' \in r_2 : t \not\approx_{\alpha} t'\}.$$

Phép chiếu: Cho r là quan hệ trên lược đồ $R(A_1, A_2, \dots, A_m)$ với miền trị của A_i là $D_i \forall i = 1, 2, \dots, m$. X là tập con của tập $\{A_1, A_2, \dots, A_m\}$. Chiếu theo ngưỡng α của r trên X là một quan hệ trên lược đồ $R|X$ kí hiệu là $r_{\alpha}[X]$ được xác định như sau: $r_{\alpha}[X] = M_{\alpha}(r[X])$.

Ví dụ. Cho 2 cho quan hệ r_1 (hình 8) và r_2 (hình 9) trên lược đồ $R(A, B, C)$ và các quan hệ tương tự trên các miền ở các hình 5, 6, 7.

	a_1	a_2	a_3	a_5
a_1	1,0	0,3	0,8	0,7
a_2	0,3	1,0	0,3	0,3
a_3	0,8	0,3	1,0	0,8
a_5	0,7	0,3	0,8	1,0

Hình 5. Quan hệ tương tự trên Dom (A)

	b_1	b_2	b_3	b_4
b_1	1,0	0,1	0,6	0,1
b_2	0,1	1,0	0,1	0,9
b_3	0,6	0,1	1,0	0,1
b_4	0,1	0,9	0,1	1,0

Hình 6. Quan hệ tương tự trên Dom (B)

	c_1	c_2	c_3
c_1	1,0	0,0	0,8
c_2	0,0	1,0	0,0
c_3	0,8	0,0	1,0

Hình 7. Quan hệ tương tự trên Dom (C)

r_1

A	B	C
a_1	b_1, b_3	c_1, c_2
a_2, a_3	b_2	c_3

Hình 8

r_2

A	B	C
a_2, a_5	b_4	c_3
a_1, a_3	b_2	c_2

Hình 9

Với $\alpha = (0,7, 0,6, 0,8)$ ta có

$r_1 \cup_{\alpha} r_2$

A	B	C
a_1	b_1, b_3	c_1, c_2
a_2, a_3, a_5	b_2, b_4	c_3
a_1, a_3	b_2	c_2

Hình 10

$r_1 \cap_{\alpha} r_2$

A	B	C
a_2, a_3, a_5	b_2, b_4	c_3

Hình 11

$r_1 -_{\alpha} r_2$

A	B	C
a_1	b_1, b_3	c_1, c_2

Hình 12

$r_{2,\alpha}[B]$

B
b_2, b_4

Hình 13

Phép tích Đề-các: Cho r, s là 2 quan hệ tương ứng trên các lược đồ $R(A_1, A_2, \dots, A_m), S(A'_1, A'_2, \dots, A'_n)$. Tích Đề-các theo α của r với s là một quan hệ trên lược đồ $(A_1, A_2, \dots, A_m, A'_1, A'_2, \dots, A'_n)$ kí hiệu là $r \times_{\alpha} s$, được xác định như sau: $r \times_{\alpha} s = M_{\alpha}(r \times s)$.

Phép chọn: Ở đây chúng tôi đề xuất hai phép chọn: chọn khẳng định và chọn có thể.

Định nghĩa biểu thức của phép chọn

- (1) Một phát biểu f_i có dạng $(\alpha_i.A_i : d)$ là một biểu thức với $\alpha_i \in [0, 1]$, A_i là tên của một thuộc tính, D_i là miền tương ứng của thuộc tính $A_i, d \subseteq D_i$.
- (2) Một phát biểu f_i có dạng NOT $(\alpha_i.A_i : d)$ là một biểu thức với $\alpha_i \in [0, 1]$, A_i là tên thuộc tính, D_i là miền tương ứng của thuộc tính $A_i, d \subseteq D_i$.
- (3) Nếu P, Q là hai biểu thức thì P AND Q là biểu thức, P OR Q là biểu thức.

Cho r là một quan hệ trên lược đồ R . Sau đây chúng tôi trình bày các phép chọn trên r với biểu thức đã cho.

a. *Chọn khẳng định*: Chọn khẳng định trong r , thỏa biểu thức F là một quan hệ trên R kí hiệu là $\sigma_F(r)$ được xác định như sau:

- (1) Nếu F có dạng $(\alpha_i.A_i : d)$ quan hệ $\sigma_F(r)$ sẽ gồm các bộ $t = (d_1, d_2, \dots, d_m)$, $d_j \subseteq D_j$, sao cho $d_i \approx_\alpha d$.
- (2) Nếu F có dạng NOT $(\alpha_i.A_i : d)$ quan hệ $\sigma_F(r)$ sẽ gồm các bộ $t = (d_1, d_2, \dots, d_m)$, $d_j \subseteq D_j$, sao cho $d_i \not\approx_\alpha d$.
- (3) Nếu F có dạng $(P \text{ AND } Q)$ thì $\sigma_F(r) = \sigma_P(r) \cap \sigma_Q(r)$.
- (4) Nếu F có dạng $(P \text{ OR } Q)$ thì $\sigma_F(r) = \sigma_P(r) \cup \sigma_Q(r)$.

b. *Chọn có thể*: Chọn có thể trong r , thỏa biểu thức F là một quan hệ trên R kí hiệu là $\sigma^F(r)$ được xác định như sau:

- (1) Nếu F có dạng $(\alpha_i.A_i : d)$, quan hệ $\sigma_F(r)$ sẽ gồm các bộ $t = (d_1, d_2, \dots, d_m)$, $d_j \subseteq D_j$ sao cho $d_i \cap_\alpha d \neq \emptyset$.
- (2) Nếu F có dạng NOT $(\alpha_i.A_i : d)$ quan hệ $\sigma_F(r)$ sẽ gồm các bộ $t = (d_1, d_2, \dots, d_m)$, $d_j \subseteq D_j$, sao cho $d_i \cap_\alpha d = \emptyset$.
- (3) Nếu F có dạng $(P \text{ AND } Q)$ thì $\sigma^F(r) = \sigma^P(r) \cap \sigma^Q(r)$.
- (4) Nếu F có dạng $(P \text{ OR } Q)$ thì $\sigma^F(r) = \sigma^P(r) \cup \sigma^Q(r)$.

Để thấy nếu quan hệ r không có bộ thừa theo ngưỡng $\beta = (\beta_1, \beta_2, \dots, \beta_m)$ thì các quan hệ kết quả $\sigma_F(r)$ và $\sigma^F(r)$ được xác định như trên cũng không có bộ thừa theo ngưỡng β .

Ví dụ. Cho quan hệ mờ r_3 ở hình 14 cùng các quan hệ tương tự trên các miền tương ứng cho ở hình 2 và hình 3 (ở Mục 3).

r_3	Tên	Màu xe	Nghề nghiệp
t_1	An	xanh đậm, xanh nhạt, hồng	nhà văn, giáo sư
t_2	Bình	xanh đen, tím đỏ	đạo diễn, giáo viên
t_3	Phúc	trắng, hồng	nhà thơ
t_4	Lộc	trắng, kem	nhà văn
t_5	Thọ	xanh đen, đỏ	phi công, đạo diễn
t_6	Tài	xanh đậm, tím đỏ	phi công

Hình 14

$F1 = (0,8. \text{Màu xe: \{xanh đậm, đỏ\}}) \text{ AND } (0,8. \text{Nghề nghiệp: \{nhà văn, giáo viên\}})$

$F2 = (0,8. \text{Màu xe: \{xanh đậm, đỏ\}}) \text{ OR } (0,8. \text{Nghề nghiệp: \{nhà văn, giáo viên\}})$

$F3 = (0,8. \text{Màu xe: \{xanh đậm, đỏ\}}) \text{ AND } (\text{NOT } (0,8. \text{Nghề nghiệp: \{nhà văn, giáo viên\}}))$

Khi đó chúng ta có:

$$\begin{aligned} \sigma_{F1}(r_3) &= \{t_2\} & \sigma_{F2}(r_3) &= \{t_1, t_2, t_5\} & \sigma_{F3}(r_3) &= \{t_5, t_6\} \\ \sigma^{F1}(r_3) &= \{t_1, t_2, t_5\} & \sigma^{F2}(r_3) &= \{t_1, t_2, t_3, t_4, t_5\} & \sigma^{F3}(r_3) &= \{t_6\} \end{aligned}$$

Chọn khẳng định $\sigma_{F1}(r_3)$ cho thông tin về những người mà màu xe chỉ có thể là hai màu tương tự với màu xanh đậm và màu đỏ và nghề nghiệp chỉ có thể là tương tự với nghề viết văn hay giáo viên. Trong khi đó chọn có thể $\sigma^{F1}(r_3)$ sẽ chọn những người có khả năng là màu xe tương tự với màu xanh đậm hay màu đỏ và nghề nghiệp có thể tương tự với nghề viết văn hay giáo viên.

Chọn khẳng định $\sigma_{F2}(r_3)$ cho thông tin về những người mà màu xe chỉ có thể là hai màu tương tự với màu xanh đậm và màu đỏ và những người nghề nghiệp chỉ có thể là tương tự với nghề viết văn hay giáo viên. Trong khi đó chọn có thể $\sigma^{F2}(r_3)$ sẽ chọn những người có khả năng màu xe tương tự với màu xanh đậm hay màu đỏ và những người có khả năng nghề nghiệp tương tự với nghề viết văn hay giáo viên.

5. CẬP NHẬT DỮ LIỆU VÀ QUAN ĐIỂM SỬ DỤNG NULL VALUE

5.1. Các giá trị Null

Trong nhiều nghiên cứu về cơ sở dữ liệu theo mô hình quan hệ, thông tin không đầy đủ được biểu diễn bằng giá trị null. Nhiều người sử dụng thuật ngữ này với những ý nghĩa khác nhau. Nói chung có các trường hợp sau:

- Những giá trị không tồn tại, thường kí hiệu là \perp . Nếu \perp xuất hiện ở bộ t ứng với một thuộc tính A thì điều đó được hiểu là bất cứ một phần tử nào ở $\text{Dom}(A)$ cũng không thể là giá trị của bộ t trên thuộc tính A . Nói cách khác, bộ t là thông tin về một đối tượng mà đối với đối tượng này không thể xét thuộc tính A . Ví dụ không thể có tên cơ quan của một người đang thất nghiệp.
- Những giá trị tồn tại nhưng chưa biết, thường kí hiệu là \mathcal{D} . Nếu \mathcal{D} xuất hiện ở bộ t ứng với một thuộc tính A thì điều đó được hiểu là bất cứ một phần tử nào ở $\text{Dom}(A)$ cũng có thể có giá trị của bộ t trên thuộc tính A . Nói cách khác, biết rằng bộ t có một giá trị trên thuộc tính A nhưng giá trị đó là gì thì chưa xác định được. Ví dụ biết Dân đi làm bằng xe của anh ta nhưng không hề biết xe anh ta màu gì.
- Không có thông tin. Về một thuộc tính A của bộ t , chúng ta không biết một giá trị xác định, lại cũng không biết rơi vào tình huống nào trong hai loại null kể trên. Chẳng hạn không biết nhà Dân có điện thoại hay không khi xét thuộc tính số điện thoại của Dân.

Để tăng cường khả năng biểu diễn thông tin không đầy đủ cho mô hình đã đề xuất, chúng tôi sử dụng hai loại null \perp và \mathcal{D} . Ví dụ sau đây giải thích cụ thể hơn ý nghĩa kí hiệu NULL mà chúng tôi đã sử dụng.

Tên	Màu xe	Nghề nghiệp
{AN}	{xanh đậm, xanh nhạt, hồng}	{bác sĩ, nha sĩ, kế toán}
{Bắc}	{xanh đen, xanh nhạt, \perp }	{ \mathcal{D} , \perp }
{Yến}	{ \mathcal{D} }	{ \perp }

Những thông tin trong bảng trên cho biết Bắc có thể không có xe (mô tô) và cũng có thể có, nếu có xe thì xe của anh ta phải màu xanh đậm hay màu xanh nhạt. Không biết Bắc có nghề nghiệp hay không (thất nghiệp). Yến có xe nhưng không biết một chút gì về màu xe của cô ấy, Yến không có nghề nghiệp.

Qui ước rằng trong một giá trị thuộc tính không xuất hiện \mathcal{D} khi đã xuất hiện một giá trị null. Chẳng hạn không thể có một bộ như sau:

{Vũ} {xanh đậm, xanh lơ, hồng} {nha sĩ, kỹ sư, \mathcal{D} }

Thêm nữa có một hạn chế trên mô hình này. Đó là các NULL không được xuất hiện trong các giá trị của thuộc tính là khóa.

Khi cho phép sử dụng kí hiệu NULL trong các giá trị thuộc tính, cần thiết phải mở rộng quan hệ tương tự trên mỗi miền D_j thành quan hệ tương tự trên $(D_j \cup \{\mathcal{D}, \perp\})$. Giả sử mọi miền D_j đều có nhiều hơn một phần tử. Hình 15 cho thấy sự mở rộng đó (Với mọi $a \in D_j$).

Dễ nhận thấy rằng sau khi mở rộng quan hệ tương tự trên mỗi miền theo bảng trên, quan hệ có được trên $(D_j \cup \{\mathcal{D}, \perp\})$ vẫn thỏa các điều kiện là quan hệ tương tự (T1 bắc cầu).

Các phép tính toán quan hệ xây dựng ở trên không có gì thay đổi khi có sự xuất hiện kí hiệu NULL trong quan hệ.

5.2. Cập nhật dữ liệu

Các phép cập nhật thường được đề cập là: thêm một bộ, thay đổi giá trị của một bộ đã có trong quan hệ, xóa đi một bộ vốn có trong quan hệ. Ở đây chúng tôi chỉ đưa ra những qui tắc cho trường hợp thêm bộ và xóa bộ, bởi phép thay đổi giá trị của bộ có thể biểu diễn qua xóa bộ và thêm bộ.

Giả sử trên một lược đồ R đã xác định một khóa K (gồm k thuộc tính khóa, không mất tính tổng quát giả sử rằng đó là các thuộc tính A_1, A_2, \dots, A_k tương ứng với các miền D_1, D_2, \dots, D_k). Đó là tập các thuộc tính mà bất cứ hai bộ nào “đủ tương tự” trên khóa cũng sẽ “đủ tương tự” trên các thuộc tính còn lại.

Qui tắc 5.1 (Cho thao tác thêm một bộ nào quan hệ)

Cho $INS \langle (d_1, d_2, \dots, d_m), (\alpha_1, \alpha_2, \dots, \alpha_m), r \rangle$ là một câu lệnh thêm một bộ $t = (d_1, d_2, \dots, d_m)$ vào quan hệ r của lược đồ R theo ngưỡng tương tự $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_m)$. Hệ thống sẽ kiểm tra xem.

- (1) Nếu trong quan hệ r không có bộ t' nào để $t[K] \approx_{\alpha|K} t'[K]$ thì thêm t vào r .
- (2) Nếu trong quan hệ r có bộ t' sao cho $t[K] \approx_{\alpha|K} t'[K]$ thì tiếp tục kiểm tra và sẽ hành động tương ứng với các tình huống sau:

(2.1) Nếu $t \approx_{\alpha} t'$ thì thay $t' \in r$ bởi $M_{\alpha}(\{t, t'\})$.

(2.2) Nếu không có $t \approx_{\alpha} t'$ thì

- Trước hết loại t' khỏi r .
- Thành lập bộ $t_* = (d_{*1}, d_{*2}, \dots, d_{*m})$, trong đó $d_{*j} = M_{\alpha}(\{d_j, d'_j\})$ với mọi A_j là thuộc tính khóa và $d_{*j} = t[A_j] \cap_{\alpha} t'[A_j]$ với mọi A_j không là thuộc tính khóa.
- Nếu $d_{*j} \neq \emptyset$ với mọi A_j không là thuộc tính khóa thì thêm t_* vào r .
- Nếu $\exists A_j$ không là thuộc tính khóa để $d_{*j} = \emptyset$ thì không làm gì nữa.

Việc thêm một bộ vào quan hệ r là để lưu trữ thông tin về một đối tượng mới (đối tượng này chưa có mặt trong quan hệ r). Nếu trường hợp (2) xảy ra có nghĩa là t và t' nói về cùng một đối tượng (theo cách quan niệm của mô hình, với một ngưỡng tương tự α). Cụ thể hơn, khi có (2.1), t và t' là thừa đối với nhau, chúng ta trộn hai bộ này để tránh dư thừa thông tin. Khi có (2.2) thì đối với mỗi thuộc tính, giá trị của bộ t và giá trị của bộ t' đều phối bày ra các nhánh khả năng bằng các phần tử đại diện của các nhánh này. Nhánh khả năng nào chỉ xuất hiện ở một bộ thôi (không xuất hiện ở bộ kia), rõ ràng không thể xảy ra trong thực tế bởi cả hai bộ cùng là thông tin đúng về cùng một đối tượng. Nếu tồn tại một thuộc tính mà trên đó cả hai bộ không có chung được một nhánh khả năng nào thì điều này chứng tỏ thông tin do t và t' cung cấp là mâu thuẫn (không đúng), cần loại bỏ cả t và t' .

Qui tắc 5.2 (Cho thao tác xóa một bộ trong quan hệ)

Cho $DEL \langle (d_1, d_2, \dots, d_k), (\alpha_1, \alpha_2, \dots, \alpha_k), r \rangle$ là một câu lệnh xóa khỏi quan hệ r bộ t với $t[K] \approx_{\alpha} (d_1, d_2, \dots, d_k)$ và ngưỡng tương tự $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_k)$. Hệ thống sẽ kiểm tra xem, nếu trong quan hệ r có bộ t để $t[K] \approx_{\alpha} (d_1, d_2, \dots, d_k)$ thì loại bỏ t khỏi quan hệ r .

Ví dụ 5.1. Quan tâm đến quan hệ r_1 ở hình 1 cùng với các quan hệ tương tự đã cho ở hình 2 và hình 3.

Cho khóa chỉ gồm 1 thuộc tính là Tên, ngưỡng tương tự là $(1,0, 0,6, 0,8)$.

- Giả sử muốn thêm bộ $t = (\{\text{Phúc}\}, \{\text{hồng, kem}\}, \{\text{nhà thơ}\})$ vào r_1 . Nghĩa là muốn thực hiện thao tác $INS \langle (\{\text{Phúc}\}, \{\text{hồng, kem}\}, \{\text{nhà thơ}\}), (1,0, 0,6, 0,8), r_1 \rangle$

Do t và t_3 đủ tương tự trên khóa (theo ngưỡng 1,0) và hơn thế $t \approx_{\alpha} t_3$ nên sau khi thực hiện thao tác thêm này r_1 sẽ như bảng cho ở hình 16.

r_1	Tên	Màu xe	Nghề nghiệp
t_1	An	xanh đậm, xanh nhạt, hồng	nhà văn, giáo sư
t_2	Bình	xanh đen, tím đỏ	đạo diễn, giáo viên
t_3	Phúc	trắng, hồng, kem	nhà thơ
t_4	Lộc	hồng, kem	nhà thơ
t_5	Thọ	xanh đen, đỏ	phi công

Hình 16

- Giả sử muốn thêm bộ $p = (\{\text{An}\}, \{\text{xanh đậm, xanh đen}\}, \{\text{đạo diễn, phi công}\})$ vào r_1 . Nghĩa là muốn thực hiện thao tác

$INS \langle (\{\text{An}\}, \{\text{xanh đậm, xanh đen}\}, \{\text{đạo diễn, phi công}\}), (1,0, 0,6, 0,8), r_1 \rangle$

Bộ t_1 cho một khẳng định về An, rằng xe của An chỉ có thể là màu xanh đậm hay màu xanh nhạt hay màu hồng, và An chỉ có thể là nhà văn hay giáo sư. Bộ p cũng cho một khẳng định về An, rằng xe của An chỉ

có thể là màu xanh đậm hay màu xanh đen, và An chỉ có thể là đạo diễn hay là phi công. Nếu quan niệm rằng các bộ phận ánh thông tin đúng đắn. Có thể thấy rằng An không thể làm một nghề tương tự với nghề phi công hay tương tự với nghề giáo viên, xe của An không thể có màu tương tự với màu hồng được. Sau khi thực hiện thao tác thêm bộ p, r_1 sẽ trở nên như hình 17.

r_1	Tên	Màu xe	Nghề nghiệp
t_1	An	xanh đậm, xanh nhạt, xanh đen	nhà văn, đạo diễn
t_2	Bình	xanh đen, tím đỏ	đạo diễn, giáo viên
t_3	Phúc	trắng, hồng	nhà thơ
t_4	Lộc	hồng, kem	nhà thơ
t_5	Thọ	xanh đen, đỏ	phi công

Hình 17

- Giả sử muốn thực hiện thao tác

$INS \langle (\{Thọ\}, \{trắng\}, \{nhà văn, phi công\}), 1, 0, 0, 6, 0, 8, r_1 \rangle$

hệ thống sẽ chỉ loại bỏ bộ t_5 ra khỏi r_1 và kết quả như ở hình 18.

r_1	Tên	Màu xe	Nghề nghiệp
t_1	An	xanh đậm, xanh nhạt, hồng	nhà văn, giáo sư
t_2	Bình	xanh đen, tím đỏ	đạo diễn, giáo viên
t_3	Phúc	trắng, hồng	nhà thơ
t_4	Lộc	hồng, kem	nhà thơ

Hình 18

Ví dụ 5.2. Cho quan hệ mờ r_2 sau đây

Tên	Tên vợ (chồng)	Nghề nghiệp
An	Thanh,	D
Bình	Loan, Diễm	D, \perp
Lạc	Hương, Phương	giáo viên

Hình 19

- Giả sử thực hiện 3 thao tác

$INS \langle (\{An\}, \{Khanh\}, \{phi công, nhà văn\}), \alpha, r_2 \rangle$

$INS \langle (\{Bình\}, \{Lan\}, \{nhà văn, đạo diễn\}), \alpha, r_2 \rangle$

$INS \langle (\{Lạc\}, \{Hương\}, \{D\}), \alpha, r_2 \rangle$

Với giả thiết các quan hệ tương tự đã cho trên các miền và α sao cho Dom (Tên người) được phân hoạch thành: $\{An, Loan, Lan\}, \{Bình\}, \{Lạc\}, \{Thanh, Khanh\}, \{Hương, Phương\}, \{Diễm\}$.

Dom (Nghề nghiệp) được phân hoạch thành: $\{nhà văn, nhà thơ, đạo diễn\}, \{giáo viên, giáo sư\}, \{phi công, nhà du hành\}$.

Kết quả r_2 sẽ như ở hình 20.

Tên	Tên vợ (chồng)	Nghề nghiệp
An	Thanh, Khanh	phi công, nhà văn
Bình	Loan, Lan	nhà văn, đạo diễn
Lạc	Hương, Phương	giáo viên

Hình 20

Cách hiểu về ngữ nghĩa của mỗi bộ trong mô hình dữ liệu này, cùng với các qui tắc cập nhật nêu trên, cho chúng ta thấy:

- Sau khi cập nhật, những thông tin lưu trữ trong quan hệ là không mâu thuẫn.
- Sau khi cập nhật, những thông tin lưu trữ trong quan hệ là đủ (không thiếu và không thừa). Thông thừa thông tin theo nghĩa quan hệ không chứa hai bộ nào thừa đối với nhau (theo ngưỡng tương tự α đang xét đến). Không thiếu thông tin theo nghĩa các thông tin vốn có trước khi cập nhật và các thông tin muốn cập nhật đều có mặt trong quan hệ (kết quả), ngoài trừ những thông tin bị phát hiện là mâu thuẫn.
- Do cập nhật có thể những thông tin chưa đầy đủ, chưa chắc chắn trở nên đầy đủ hơn, chắc chắn hơn.

6. KẾT LUẬN

Qua bài báo này chúng tôi đề xuất một cách mở rộng mô hình quan hệ truyền thống nhờ một quan hệ mờ đánh giá độ tương tự giữa các phần tử thuộc cùng mỗi miền. Một quan điểm mới về dư thừa dữ liệu được trình bày và trên cơ sở đó, chúng tôi trình bày kết quả xây dựng các phép toán quan hệ. Để làm tăng khả năng nắm bắt thông tin không chính xác và không chắc chắn (vốn có), mô hình cho phép có sự xuất hiện của hai loại kí hiệu null. Các qui tắc cho các thao tác cập nhật dữ liệu cũng được đề nghị. Một số vấn đề đặt ra cần được tiếp tục nghiên cứu trên mô hình này, đó là:

- Khóa và các dạng chuẩn.
- Các phụ thuộc dữ liệu trong trường hợp có kí hiệu null.
- Ngôn ngữ hỏi.

TÀI LIỆU THAM KHẢO

- [1] Buckeles B.P. and Petry E., A fuzzy representation of data for relational databases, *Fuzzy Sets and System* **7** (1980) 312–226.
- [2] Codd E.F., A relation model of data for large shared data banks, *Commun. ACM* **13** (6) (1970) 377–387.
- [3] Ho Thuan, Ho Cam Ha, Huynh Van Nam, Some comments about “Axiomatisation of fuzzy multivalued dependencies in a fuzzy relational data model”, *Journal of Computer Science and Cybernetics* **16** (4) (2000) 30–33.
- [4] Ho Thuan, Ho Cam Ha, An approach to extending the relational database model for handling incomplete information and data dependencies, *Journal of Computer Science and Cybernetics* **17** (3) (2001) 41–47.
- [5] K.S. Candan and Hohn Grant and V.S. Subrahmanian, A unified treatment of Null Values using constraints, *Information Science* **98**, 101–156.
- [6] Petry E. and Bosc P., *Fuzzy Databases Principles and Applications*, Kluwer Academic Publishers, 1986.
- [7] Ullman J.D., *Principles of Database System*, 2nd Ed., Computer Science Press, Rockville, MD, 1982.
- [8] Zadeh L.A., Fuzzy sets, *Inform. Control* **12** (1965) 338–353.
- [9] Zadeh L.A., Fuzzy sets as a basis for a theory of possibility, *Fuzzy Sets and System* **1** (1978) 3–28.

Nhận bài ngày 5 - 6 - 2001

Nhận lại sau khi sửa ngày 15 - 10 - 2001

Hồ Thuần - Viện Công nghệ thông tin.

Hồ Cẩm Hà - Trường Đại học Sư phạm Hà Nội.