

# CHUẨN HÓA CÁC LỚP ĐỐI TƯỢNG TRONG LỢC ĐỒ CƠ SỞ DỮ LIỆU HƯỚNG ĐỐI TƯỢNG MỜ

DOÀN VĂN BAN<sup>1</sup>, HỒ CẨM HÀ<sup>2</sup>, VŨ ĐỨC QUẢNG<sup>3</sup>

<sup>1</sup>*Viện Công nghệ Thông tin, Viện Khoa học và Công nghệ Việt Nam*

<sup>2</sup>*Trường Đại học Sư phạm Hà nội*

<sup>3</sup>*Trường Đại học Quảng Nam*

**Tóm tắt.** Bài báo về nghiên cứu phụ thuộc mờ giữa các thuộc tính trong một lớp đối tượng. Tương tự như trong cơ sở dữ liệu quan hệ mờ/rõ, đã đưa ra các dạng chuẩn đối tượng mờ (1FONF, 2FONF, 3FONF) và các thuật toán chuẩn hóa lớp về các dạng chuẩn trong cơ sở dữ liệu hướng đối tượng mờ.

**Abstract.** In this paper, we study fuzzy dependencies between attributes in an object class. Similarly in the fuzzy (clear) relational database, we present the fuzzy object normal forms (1FONF, 2FONF, 3FONF) and class normalizing algorithms for normal forms in fuzzy object-oriented databases.

## 1. GIỚI THIỆU

Trong những năm gần đây, việc nghiên cứu, ứng dụng cơ sở dữ liệu hướng đối tượng (CSDL HDT) mờ để giải quyết những hạn chế của cơ sở dữ liệu quan hệ/hướng đối tượng rõ trong việc xử lý và lưu trữ các thông tin không chắc chắn, không đầy đủ trở thành một chủ đề nghiên cứu quan trọng trong một số lĩnh vực của khoa học máy tính [5]. Tiếp cận hướng đối tượng mờ trong việc thiết kế các hệ thống phần mềm đã nhận được một sự chú ý đáng kể, đặt biệt là trong lĩnh vực cơ sở dữ liệu. Cũng giống như trong cơ sở dữ liệu kinh điển, trong cơ sở dữ liệu mờ có thể có dư thừa dữ liệu và các dị thường khi cập nhật nếu như các lược đồ cơ sở dữ liệu không được thiết kế một cách thích hợp [2, 6, 10]. Thông thường, các phụ thuộc dữ liệu là tri thức ngữ nghĩa về thế giới thực và được xem như các ràng buộc toàn vẹn đối với việc thiết kế cơ sở dữ liệu.

Dựa trên cách thức tính độ tương đương ngữ nghĩa của hai dữ liệu mờ, đưa ra các định nghĩa về phụ thuộc hàm mờ trên lớp đối tượng mờ. Ngoài ra, còn giới thiệu ba dạng chuẩn đối tượng mờ 1FONF, 2FONF, 3FONF và các thuật toán chuẩn hóa lớp về dạng chuẩn đối tượng mờ.

Nội dung bài báo gồm: Mục 2 trình bày cơ sở lý thuyết về tập mờ, các quan hệ ngữ nghĩa và ước lượng ngữ nghĩa giữa hai tập mờ. Mục 3 nêu các định nghĩa về phụ thuộc hàm mờ, phụ thuộc phương thức mờ trên lớp đối tượng. Mục 4 đưa ra các định nghĩa về dạng chuẩn đối tượng mờ 1, 2, 3 và các thuật toán chuẩn hóa lớp về dạng chuẩn đối tượng mờ. Và cuối cùng là phần kết luận.

## 2. CƠ SỞ LÝ THUYẾT

Theo Zadeh [11], dữ liệu mờ được mô tả như một tập mờ. Cho  $U$  là một vũ trụ, một tập mờ  $F$  trên  $U$  xác định bởi hàm thuộc  $\mu_F : U \rightarrow [0, 1]$ , gán mỗi phần tử  $u$  thuộc  $U$  một giá trị để chỉ độ thuộc của  $u$  vào tập mờ  $F$ . Tập mờ  $F$  được biểu diễn dưới dạng

$$F = \{\mu_F(u_1)/u_1, \mu_F(u_2)/u_2, \mu_F(u_3)/u_3, \dots, \mu_F(u_n)/u_n\} \quad \text{hay} \quad F = \int_{u \in U} \frac{\mu_F(u)}{u}.$$

Khi  $\mu_F(u)$  được xem như độ đo khả năng mà một biến  $X$  nhận giá trị  $u$ , một giá trị mờ được biểu diễn bằng phân bố khả năng  $\pi_X$  như sau

$$\pi_X = \{\pi_X(u_1)/u_1, \pi_X(u_2)/u_2, \pi_X(u_3)/u_3, \dots, \pi_X(u_n)/u_n\},$$

trong đó,  $\pi_X(u_i)$ ,  $u_i \in U$ , biểu thị khả năng  $X$  nhận giá trị  $u_i$ ,  $\pi_X, F$  lần lượt biểu diễn phân bố khả năng và tập mờ đối với một giá trị mờ. Rõ ràng,  $\pi_X = F$ .

Ngữ nghĩa của dữ liệu mờ  $\pi_A$  còn được gọi là không gian ngữ nghĩa, ký hiệu là  $SS(\pi_A)$ . Cho hai dữ liệu mờ  $\pi_A$  và  $\pi_B$ ,  $SS(\pi_A)$  và  $SS(\pi_B)$  lần lượt là không gian ngữ nghĩa của chúng. Nếu  $SS(\pi_A) \supseteq SS(\pi_B)$ , ta nói  $\pi_A$  bao hàm ngữ nghĩa  $\pi_B$  hay  $\pi_B$  được bao hàm ngữ nghĩa bởi  $\pi_A$ . Nếu  $SS(\pi_A) \supseteq SS(\pi_B)$  và  $SS(\pi_A) \subseteq SS(\pi_B)$ , ta nói  $SS(\pi_A)$  và  $SS(\pi_B)$  là tương đương ngữ nghĩa với nhau [12].

Cho  $U = \{u_1, u_2, \dots, u_n\}$  là tập vũ trụ.  $\pi_A$  và  $\pi_B$  là hai dữ liệu mờ trên  $U$  dựa trên phân bố khả năng. Mức độ mà  $\pi_A$  bao hàm ngữ nghĩa  $\pi_B$ , ký hiệu  $SID(\pi_A, \pi_B)$  được xác định như sau [12]:

$$SID(\pi_A, \pi_B) = \sum_{i=1}^n \min_{u_i \in U} (\pi_B(u_i), \pi_A(u_i)) / \sum_{i=1}^n \pi_B(u_i).$$

Với Res là một quan hệ tương đương trên miền  $U$ ,  $\alpha (0 \leq \alpha \leq 1)$  là một ngưỡng tương ứng với Res. Mức độ mà  $\pi_A$  bao hàm ngữ nghĩa  $\pi_B$  theo quan hệ Res được xác định như sau [12]

$$SID(\pi_A, \pi_B) = \sum_{i=1}^n \min_{u_i, u_j \in U, Res(u_i, u_j) \geq \alpha} (\pi_B(u_i), \pi_A(u_i)) / \sum_{i=1}^n \pi_B(u_i).$$

Mức độ tương đương ngữ nghĩa giữa hai dữ liệu mờ  $\pi_A$  và  $\pi_B$ , ký hiệu là  $SE(\pi_A, \pi_B)$  và được biểu thị như sau [12]

$$SE(\pi_A, \pi_B) = \min(SID(\pi_A, \pi_B), SID(\pi_B, \pi_A)).$$

## 3. PHỤ THUỘC HÀM MỜ, PHỤ THUỘC PHƯƠNG THỨC MỜ

Ký hiệu OID là tập các định danh đối tượng, mỗi phần tử của OID xác định duy nhất một đối tượng,  $A$  là tập hữu hạn các tên thuộc tính, mỗi phần tử của  $A$  là tên của một trường.

Tập hợp các miền giá trị  $dom_1, dom_2, \dots, dom_n$  ( $dom_i$  là miền giá trị mờ hay rõ) được ký hiệu là  $dom$ .

Đối tượng mờ. Một đối tượng mờ là một bộ  $o = (id, v)$ ,  $id$  là một phần tử của OID,  $v$  là một giá trị mờ (rõ).  $val(o)$  biểu thị giá trị  $v$  đối với  $o$ , được định nghĩa như sau [4]

(a) một phần tử của  $dom$ , một phần tử của OID và  $nil$  là các giá trị.

(b) nếu  $v_1, v_2, \dots, v_n$  là các giá trị và  $A_1, A_2, \dots, A_n$  là tên các thuộc tính thì  $(A_1 : v_1, A_2 : v_2, \dots, A_n : v_n)$  là giá trị bộ và  $\{v_1, v_2, \dots, v_n\}$  là giá trị tập.

Kiểu (types): Ta có thể định nghĩa một tập kiểu đối tượng một cách đệ qui như sau:

(1) Kiểu dữ liệu nguyên tố (là các số nguyên, số thực hay chuỗi ký tự có độ dài cố định hay thay đổi), kiểu dữ liệu mờ là một kiểu đối tượng.

(2) Nếu  $T$  là kiểu đối tượng thì  $set(T)$  là cũng là một kiểu đối tượng. Một đối tượng có kiểu  $set(T)$  là một tập hợp các đối tượng có kiểu  $T$ ,  $set(T)$  còn được gọi là kiểu tập.

(3) Nếu  $T_1, T_2, \dots, T_n$  là các kiểu đối tượng thì  $tuple(T_1, T_2, \dots, T_n)$  là một kiểu đối tượng. Một đối tượng kiểu  $T_1, T_2, \dots, T_n$  là một  $n$ -bộ, trong đó thành phần thứ  $i$  của  $n$ -bộ này có kiểu  $T_i$ ,  $tuple(T_1, T_2, \dots, T_n)$  còn được gọi là kiểu bộ.

### 3.1. Phụ thuộc hàm mờ trong lớp đối tượng

Trong cơ sở dữ liệu rõ, phụ thuộc hàm giữa hai tập thuộc tính được xác định dựa trên phép so sánh bằng giữa hai giá trị trên mỗi tập thuộc tính. Với cơ sở dữ liệu mờ, phép so sánh bằng giữa hai giá trị trên tập thuộc tính được thay bởi mức độ tương đương giữa hai giá trị của tập thuộc tính đó. Đối với CSDL HDT mờ, kiểu dữ liệu của các thuộc tính không đơn thuần là các kiểu cơ sở mà còn có các kiểu phức tạp khác như kiểu tập, kiểu bộ, ... Vì vậy, độ tương đương ngữ nghĩa trên tập thuộc tính  $X$  của hai đối tượng  $O_1, O_2$  của lớp đối tượng mờ  $c(U)$  được xác định đệ qui như sau:

(i) Nếu  $X = \{A_1, A_2, \dots, A_k\}$  thì

$$SE(O_1.X, O_2.X) = \min(SE(O_1.A_1, O_2.A_1), SE(O_1.A_2, O_2.A_2), \dots, SE(O_1.A_k, O_2.A_k));$$

(ii) Nếu  $A_i$  có kiểu lớp  $c_i(U_i)$  thì  $SE(O_1.A_i, O_2.A_i) = 1$  nếu  $O_1.A_i \equiv O_2.A_i$ . Ngược lại,  $SE(O_1.A_i, O_2.A_i) = SE(O_1.A_i.U_i, O_2.A_i.U_i)$

(iii) Nếu  $A_i$  là thuộc tính kiểu tập, giả sử  $O_1.A_i = \{s_1, s_2, \dots, s_k, \dots, s_n\}$ ,  $O_2.A_i = \{s_1, s_2, \dots, s_j, \dots, s_m\}$  thì độ tương đương ngữ nghĩa của 2 giá trị mờ  $O_1.A_i$  và  $O_2.A_i$  được tính như sau:

$$SE(O_1.A_i, O_2.A_i) = \max \left( \min \left( \max_k \left( \frac{SID(s_j, s_k)}{k=1..n, j=1..m} \right) \right), \min \left( \max_j \left( \frac{SID(s_k, s_j)}{j=1..m, k=1..n} \right) \right) \right).$$

(iv) Nếu  $A_i$  là thuộc tính nhận giá trị bộ, giả sử  $O_1.A_i = \{v_1, v_2, \dots, v_k, \dots, v_n\}$ ,  $O_2.A_i = \{w_1, w_2, \dots, w_j, \dots, w_m\}$  thì:  $SE(O_1.A_i, O_2.A_i) = \min_k \left( \frac{SE(v_k, w_j)}{k, j=1..n} \right)$ .

**Định nghĩa 3.1.** Cho lớp mờ  $c$  với tập các thuộc tính  $U$  (ký hiệu  $c(U)$ );  $X, Y \subseteq U$ , lớp  $c$  thỏa mãn phụ thuộc hàm mờ  $X \xrightarrow{fc} Y$  nếu  $\forall O_1, O_2$  là các đối tượng của lớp  $c$  thì  $SE(O_1.X, O_2.X) \leq SE(O_1.Y, O_2.Y)$ .

**Ví dụ 1:** Cho lớp  $c$  với tập thuộc tính  $U, X, Y \in U$ ,  $Res(X)$  và  $Res(Y)$  là hai quan hệ tương

tự trên  $X$  và  $Y$ ,  $\alpha_1 = 0.9, \alpha_2 = 0.95$  lần lượt là hai ngưỡng trên  $Res(X)$  và  $Res(Y)$ .

$Res(X)$	a	b	c	d	e		$Res(Y)$	f	g	h	i	j
a	1.0	0.2	0.3	0.2	0.4		f	1.0	0.3	0.2	0.96	0.2
b		1.0	0.92	0.4	0.1		g		1.0	0.4	0.2	0.3
c			1.0	0.1	0.3		h			1.0	0.3	0.1
d				1.0	0.2		i				1.0	0.4
e					1.0		j					1.0

Các thể hiện của lớp đối tượng  $c$  được cho như sau.

ID	X	Y	...
ID1	{0.7/a, 0.4/b, 0.5/d}	{0.9/f, 0.6/g, 1.0/h}	...
ID2	{0.5/a, 0.4/c, 0.8/d}	{0.6/g, 0.9/h, 0.9/i}	...
ID3	{0.3/d, 0.8/e}	{0.6/h, 0.4/i, 0.1/j}	...

$\forall O_1, O_2$  là các đối tượng thuộc lớp  $c$ , ta có:

$SE(O_1.X, O_2.X) = \min(SID(O_1.X, O_2.X), SID(O_2.X, O_1.X)) = \min(0.824, 0.875) = 0.824$  và  $SE(O_1.Y, O_2.Y) = \min(SID(O_1.Y, O_2.Y), SID(O_2.Y, O_1.Y)) = \min(1.0, 0.96) = 0.96$ . Vậy  $SE(O_1.Y, O_2.Y) > SE(O_1.X, O_2.X)$ . Theo định nghĩa phụ thuộc hàm mờ, lớp đối tượng mờ  $c$  thỏa mãn phụ thuộc hàm mờ  $X \xrightarrow{fc} Y$ .

**Bổ đề 3.1.** Cho lớp  $c(U)$ ,  $X \subseteq U, A \subseteq U, X \xrightarrow{fc} A$ , giả sử  $A : tuple(A_1 : T_1, A_2 : T_2, \dots, A_k : T_k)$ . Nếu  $A$  được thay thế bởi  $A_1, A_2, \dots, A_k$  trong  $c$  thì  $X \xrightarrow{fc} A_i, i = 1, 2, \dots, k$ .

*Chứng minh:* Vì  $X \xrightarrow{fc} A$  nên với  $O_1, O_2$  bất kỳ thuộc lớp đối tượng mờ  $c$ , ta có  $SE(O_1.A, O_2.A) \geq SE(O_1.X, O_2.X)$  và vì  $A$  được thay thế bởi  $A_1, A_2, \dots, A_k$  nên

$$SE(O_1.A, O_2.A) = \min(SE(O_1.A_i, O_2.A_i)), i = 1..k.$$

Do đó  $SE(O_1.A_i, O_2.A_i) \geq SE(O_1.X, O_2.X), i = 1..k$ . Theo định nghĩa phụ thuộc hàm mờ, suy ra  $X \xrightarrow{fc} A_i, i = 1, 2, \dots, k$ . ■

**Định lý 3.1.** Phụ thuộc hàm trong cơ sở dữ liệu quan hệ thỏa mãn định nghĩa phụ thuộc hàm mờ trong CSDL HDT mờ.

*Chứng minh:* Cho quan hệ  $r(U)$  thỏa mãn phụ thuộc hàm  $X \rightarrow Y$ , với  $X, Y \subseteq U$ . Nên  $\forall t_1, t_2 \in r, t_1[X] = t_2[X]$  thì  $t_1[Y] = t_2[Y]$ . Giả sử  $U$  là tập các thuộc tính của lớp mờ  $c$  thì  $\forall O_1, O_2$  là các đối tượng của  $c$  ta có:  $SE(O_1.X, O_2.X) = 1, SE(O_1.Y, O_2.Y) = 1$ .

Do đó,  $SE(O_1.Y, O_2.Y) \geq SE(O_1.X, O_2.X)$ . Theo định nghĩa phụ thuộc hàm mờ trong CSDL HDT ta có  $X \xrightarrow{fc} Y$ . ■

### 3.2. Tách các thuộc tính kiểu bộ theo phụ thuộc hàm mờ

Tương tự như trong CSDL HDT rõ, việc nhóm một số thuộc tính lại thành một thuộc tính kiểu bộ trong một lớp có thể dẫn đến một số ràng buộc ngữ nghĩa hay phụ thuộc hàm mờ

giữa các thuộc tính không được bảo toàn [7]. Trong trường hợp này, cần phải tách các thuộc tính kiểu bộ để bảo toàn các ràng buộc ngữ nghĩa đó.

Giả sử  $A_* : tuple(X_i : T_i), i = 1, 2, 3, \dots, k (k > 1)$ . Nếu  $A_*$  được thay thế bởi  $X_i$  và tồn tại phụ thuộc hàm mờ  $X \xrightarrow{fc} Y$ , với  $X = \bigcup_{i=1}^k X'_i, X'_i \subseteq X_i$  và  $Y'_i \subseteq X_k$  thì  $A_* : tuple(X_i : T_i), i = 1, 2, 3, \dots, k$  được tách thành các thuộc tính như sau:

$$B_i : tuple(X_i \setminus X'_i : T_{X_i \setminus X'_i}), i = 1, 2, 3, \dots, k - 1,$$

$$B_k : tuple(X_k \setminus X'_k \setminus Y : T_{X_k \setminus X'_k \setminus Y}), i = 1, 2, 3, \dots, k - 1,$$

$$C_i : tuple(X'_i : T_{X'_i}), i = 1, 2, 3, \dots, k,$$

$$D : tuple(T : T_Y).$$

### 3.3. Các luật suy dẫn trên các phụ thuộc hàm mờ

Các luật suy dẫn của phụ thuộc hàm mờ trên lớp đối tượng mờ  $c(U), X, Y, Z \subseteq U$ .

**Luật 31** *Phản xạ.* Nếu  $X \supseteq Y$  thì  $X \xrightarrow{fc} Y$ .

**Luật 32** *Tăng trưởng.* Nếu  $X \xrightarrow{fc} Y$  thì  $XZ \xrightarrow{fc} YZ$ .

**Luật 33** *Bắc cầu.* Nếu  $X \xrightarrow{fc} Y$  và  $Y \xrightarrow{fc} Z$  thì  $X \xrightarrow{fc} Z$ .

**Luật 34** *Hợp.* Nếu  $X \xrightarrow{fc} Y$  và  $X \xrightarrow{fc} Z$  thì  $X \xrightarrow{fc} YZ$ .

**Luật 35** *Bắc cầu giả.* Nếu  $X \xrightarrow{fc} Y$  và  $YW \xrightarrow{fc} Z$  thì  $XW \xrightarrow{fc} Z$ .

**Luật 36** *Tách.* Nếu  $X \xrightarrow{fc} Y$  và  $Z \subset Y$  thì  $X \xrightarrow{fc} Z$ .

**Định lý 3.2.** Các luật suy dẫn 3.1 đến 3.6 trên lớp đối tượng mờ là xác đáng và đầy đủ.

Việc chứng minh định lý tương tự như trường hợp phụ thuộc hàm mờ trong cơ sở dữ liệu quan hệ mờ [10, 13].

### 3.4. Phụ thuộc phương thức mờ

Phương thức định nghĩa trong lớp được mô tả

$$M_j(N, I, R) \Rightarrow (u, v, g),$$

trong đó,  $N$ : tên phương thức;  $I$ : tập các tham số đầu vào  $\{<\text{tên}, \text{kiểu}>\}$ ;  $R$ : tập các thuộc tính mà giá trị của nó được đọc bởi phương thức;  $u$ : tập các tham số đầu ra bao gồm kiểu giá trị trả về  $\{<\text{tên}, \text{kiểu}>\}$ ;  $v$ : tập các thuộc tính mà giá trị của nó bị thay đổi bởi phương thức;  $g$ : tập các thông báo được đưa ra bởi phương thức có dạng  $\{[o, msg, p]\}$ ,  $o$  là nơi nhận thông báo,  $msg$  là thông báo và  $p$  là tập các tham số có trong thông báo  $\{<n, t>\}$ .

Mối quan hệ giữa phương thức với các thuộc tính được xác định thông qua việc phương thức sử dụng các thuộc tính đó để đọc hay sửa đổi. Sự phụ thuộc của phương thức vào các

thuộc tính như thế được gọi là phụ thuộc phương thức mờ (fuzzy method dependency – viết tắt  $fm$ -phụ thuộc).

**Định nghĩa 3.2.**  $Fm$ -phụ thuộc: Cho  $M_j$  là một phương thức của một lớp  $c$ ,  $M_j(N, I, R) \Rightarrow (u, v, g)$  và  $X = R \cup v \neq \emptyset$  là tập các thuộc tính (đọc, sửa đổi) được sử dụng bởi  $M_j$ . Quan hệ giữa  $M_j$  và  $X$  được gọi là  $fm$ -phụ thuộc, ký hiệu  $X \xrightarrow{fm} M_j$ .

Các luật suy dẫn của  $fm$ -phụ thuộc:

**Luật 37** Nếu  $X \xrightarrow{fm} M_j$  và  $Y \xrightarrow{fm} M_j$  thì  $XY \xrightarrow{fm} M_j$ .

**Luật 38** Nếu  $X \xrightarrow{fm} M_i$  và  $X \xrightarrow{fm} M_j$  thì  $X \xrightarrow{fm} M_i M_j$ .

**Luật 39** Nếu  $X \xrightarrow{fm} M_j$  và  $Y \subseteq X$  thì  $Y \xrightarrow{fm} M_j$ .

**Luật 310** Nếu  $M_j$  gọi đến  $M_i$  và  $X \xrightarrow{fm} M_i, Y \xrightarrow{fm} M_j$  thì  $XY \xrightarrow{fm} M_j$ .

#### 4. CHUẨN HÓA LỚP ĐỐI TƯỢNG MỜ

Tương tự như mô hình cơ sở dữ liệu quan hệ, trước khi chuẩn hóa lớp, ta xem xét một số khái niệm như phụ thuộc hàm mờ bộ phận, phụ thuộc hàm mờ bắc cầu, khóa của lớp, thuộc tính khóa.

**Định nghĩa 4.1.** (Phụ thuộc hàm mờ bộ phận) Cho lớp  $c(U)$ ,  $X, Y \subseteq U$ ,  $Y$  được gọi là phụ thuộc hàm mờ bộ phận vào  $X$  khi và chỉ khi và  $X \xrightarrow{fc} Y$  và  $\exists X' \subset X, X' \neq \emptyset$  mà  $X' \xrightarrow{fc} Y$ .

**Định nghĩa 4.2.** (Phụ thuộc hàm mờ bắc cầu) Cho lớp  $c(U)$ ,  $X, Y, Z \subseteq U$ ,  $Y$  được gọi là phụ thuộc hàm mờ bắc cầu vào  $X$  nếu  $X \xrightarrow{fc} Z$ ,  $Z \xrightarrow{fc} Y$  và  $Z \xrightarrow{fc} X$ .

**Định nghĩa 4.3.** (Khóa của lớp đối tượng mờ) Cho lớp mờ  $c(U)$ ,  $K \subseteq U$ ,  $K$  được gọi là khóa của lớp đối tượng mờ  $c$  nếu thỏa mãn hai điều kiện sau đây:

1.  $K \xrightarrow{fc} U$ ,
2.  $\forall K' \subset K, K' \not\xrightarrow{fc} U$ ,

$U$  là tập thuộc tính của lớp đối tượng  $c$  không bao gồm định danh của đối tượng.

**Định nghĩa 4.4.** (Thuộc tính khóa) Cho lớp mờ  $c(U)$ ,  $K \subseteq U$ ,  $K$  là một khóa của lớp mờ  $c$ ,  $A$  được gọi là thuộc tính khóa nếu  $A \in K$ . Những thuộc tính không phải là thuộc tính khóa gọi là thuộc tính không khóa.

#### 4.1. Các dạng chuẩn đối tượng mờ

Mục đích của các dạng chuẩn trong CSDL HDT mờ (rõ) là giải quyết các vấn đề dư thừa dữ liệu và dị thường trong khi thao tác dữ liệu. Tương tự cách tiếp cận trong [2, 7] trên mô hình CSDL HDT rõ, dựa trên khái niệm phụ thuộc hàm mờ trên tập thuộc tính không bao gồm định danh đối tượng của lớp, kết hợp với phương pháp chuẩn hóa lược đồ quan hệ trên cơ sở dữ liệu quan hệ để đưa ra các dạng chuẩn đối tượng mờ.

**Định nghĩa 4.5.** (Dạng chuẩn đối tượng mờ 1(1FONF)) Lớp  $c(U)$  được gọi là ở 1FONF nếu  $\forall A \in U$ , hoặc  $A$  là thuộc tính có kiểu bộ không nhận giá trị tập hoặc  $A$  chỉ chứa các thuộc tính không có kiểu bộ nhận giá trị tập hoặc  $A$  không phải là thuộc tính khóa kiểu bộ.

**Định nghĩa 4.6.** (Dạng chuẩn đối tượng mờ 2) (2FONF) Lớp  $c(U)$  được gọi là ở 2FONF nếu  $c$  ở 1FONF và không tồn tại thuộc tính không khóa  $A$  phụ thuộc hàm mờ bộ phận vào khóa.

**Định nghĩa 4.7.** (Dạng chuẩn đối tượng mờ 3 (3FONF)). Lớp  $c(U)$  được gọi là ở 3FONF nếu  $c$  ở 2FONF và không tồn tại thuộc tính không khóa  $A$  phụ thuộc hàm mờ bắc cầu mờ vào khóa.

Một lược đồ CSDL HDT mờ được gọi là ở 1FONF (2FONF, 3FONF) nếu mọi lớp thuộc lược đồ CSDL HDT mờ đều ở 1FONF (2FONF, 3FONF).

#### 4.2. Chuẩn hóa lớp đối tượng mờ

##### 4.2.1. Thuật toán chuẩn hóa lớp về 1FONF

Vào: Lớp đối tượng mờ  $c(U)$ .

Ra: Phép tách lớp  $c$  thành các lớp  $c_i$  ở 1FONF không mất mát thông tin.

**Bước 1.** Nếu lớp  $c$  có dạng (i), (ii) và  $A_i$  không phải là thuộc tính khóa thì ta tách lớp  $c$  thành 2 lớp  $c$  và  $c_1$  ở 1FONF như sau:

(i) *Class*  $c$  *tuple*( $A_1 : T_1, A_2 : T_2, \dots, A_i : set(tuple(X : T_X)), \dots, A_n : T_n$ ).

Tách lớp  $c$  thành 2 lớp  $c$  và  $c_1$  như sau

*Class*  $c$  *tuple*( $A_1 : T_1, A_2 : T_2, \dots, A_i : set(c_1), \dots, A_n : T_n$ ),

*Class*  $c_1$  *tuple*( $X : T_X$ ).

(ii) *Class*  $c$  *tuple*( $A_1 : T_1, A_2 : T_2, \dots, A_i : tuple(B_1 : I_1, B_2 : I_2, \dots, B_i : set(tuple(X : T_X)), \dots, B_m : T_m), \dots, A_n : T_n$ ).

Tách  $c$  thành 2 lớp  $c$  và  $c_1$  như sau:

*Class*  $c$  *tuple*( $A_1 : T_1, A_2 : T_2, \dots, A_i : tuple(B_1 : I_1, B_2 : I_2, \dots, B_i : set(c_1)), \dots, B_m : T_m), \dots, A_n : T_n$ ),

*Class*  $c_1$  *tuple*( $X : T_X$ ).

Các phương thức phụ thuộc vào  $A_1, A_2, A_3, \dots, A_n$  được phân vào lớp  $c$ , các phương thức phụ thuộc vào tập thuộc tính  $X$  được phân vào lớp  $c_1$ . Viết lại các phương thức trong  $c, c_1$  cho phù hợp dựa vào tập thuộc tính và kiểu của nó trong lớp đó.

**Bước 2.** Nếu lớp  $c$  có dạng (ii), (iii) và  $A_i$  là thuộc tính khóa thì ta tách lớp  $c$  thành 3 lớp  $c, c_1$  và  $c_2$  (đối với (ii)), thành 2 lớp  $c, c_1$  (đối với (iii)) như sau:

(ii) *Class c tuple*( $A_1 : T_1, A_2 : T_2, \dots, A_i : tuple(B_1 : I_1, B_2 : I_2, \dots, B_i : set(tuple(X : T_X)), \dots, B_m : T_m), \dots, A_n : T_n$ ).

Tách  $c$  thành 3 lớp  $c, c_1$  và  $c_2$  như sau:

*Class c tuple*( $A_1 : T_1, A_2 : T_2, \dots, A_i : c_1, \dots, A_n : T_n$ ),

*Class c<sub>1</sub> tuple*( $B_1 : I_1, B_2 : I_2, \dots, B_i : set(c_2), \dots, B_m : T_m$ ),

*Class c<sub>2</sub> tuple*( $X : T_X$ ).

Các phương thức phụ thuộc vào  $A_1, A_2, A_3, \dots, A_n$  được phân vào lớp  $c$ , các phương thức phụ thuộc vào tập thuộc tính  $A_i$  được phân vào lớp  $c_1$ , các phương thức phụ thuộc vào tập thuộc tính  $X$  được phân vào lớp  $c_2$ . Viết lại các phương thức trong  $c, c_1, c_2$  cho phù hợp dựa vào tập thuộc tính và kiểu của nó trong lớp đó.

(iii) *Class c tuple*( $A_1 : T_1, A_2 : T_2, \dots, A_i : tuple(X : T_X), \dots, A_n : T_n$ ).

Tách  $c$  thành 2 lớp  $c$  và  $c_1$  như sau:

*Class c tuple*( $A_1 : T_1, A_2 : T_2, \dots, A_i : c_1, \dots, A_n : T_n$ ),

*Class c<sub>1</sub> tuple*( $X : T_X$ ).

Các phương thức phụ thuộc vào  $A_1, A_2, A_3, \dots, A_n$  được phân vào lớp  $c$ , các phương thức phụ thuộc vào tập thuộc tính  $X$  được phân vào lớp  $c_1$ . Viết lại các phương thức trong  $c, c_1$  cho phù hợp dựa vào tập thuộc tính và kiểu của nó trong lớp đó.

#### 4.2.2. Thuật toán chuẩn hóa lớp về 2FONF

Vào: Lớp đối tượng mờ  $c(U)$  ở 1FONF,  $F$ : tập các phụ thuộc hàm mờ tối thiểu, khóa  $K$  của lớp  $c$ .

Ra: Tách lớp  $c$  thành các lớp  $c_i$  ở 2FONF không mất mát thông tin.

**Bước 1.** Với mỗi phụ thuộc hàm mờ  $X_i \xrightarrow{f_c} A_j \in F, X_i \subset K$  và  $A_j$  là thuộc tính không khóa ( $A_j$  phụ thuộc hàm mờ bộ phận vào khóa  $K$ ), đặt  $S_i = X_i \cup A_j$ . Tạo các lớp mới  $c_i$  với tập thuộc tính  $S_i$  và khóa tương ứng là  $X_i$ , các phương thức  $M$  phụ thuộc vào  $S_i$  được đưa vào lớp  $c_i$ .

**Bước 2.** Đặt  $U = U - \{A_i\}$  ( $A_i$  là các thuộc tính phụ thuộc hàm mờ bộ phận vào khóa  $K$ , tạo lớp mới  $c(U)$ . Các phương thức  $M$  phụ thuộc vào  $U$  được đưa vào lớp  $c(U)$ .

**Bước 3.** Đặt  $T$  là tập các lớp được tạo ra ở bước 1. Với mỗi lớp  $c_i \in T$ , nếu  $c(U)$  chứa tập thuộc tính  $H$  mà  $H$  là khóa chính của  $c_i$  thì thay tập thuộc tính  $H$  trong  $c$  bởi thuộc tính  $h$  có kiểu lớp  $c_i$ .

Với mỗi phương thức  $M$  trong các lớp được viết lại cho phù hợp dựa vào các thuộc tính và kiểu của nó trong lớp đó.

**Ví dụ 2.** Cho lớp EMPPROJ mô tả người làm việc trong các dự án như sau.

```
Class EMPPROJ
ATTRIBUTE
    ssn: integer;
```

*eName*: string;  
*projNum*: integer;  
*hours*: fuzzy integer;  
*pName*: string;  
*pLocation*: string;

WEIGH ...

METHODS

InputEMP(*ssn*, *eName*);  
 InputPROJ(*projNum*, *pName*, *pLocation*);  
 InputHour(*hours*);  
*pD*: domain[0,1] of real;  
 END

Tập phụ thuộc hàm mờ trên lớp EMPPROJ là  $F = \left\{ ssn \xrightarrow{fc} eName; projNum \xrightarrow{fc} pName; ssn \xrightarrow{fc} pLocation; ssn, projNum \xrightarrow{fc} hours \right\}$

Khóa của lớp EMPPROJ là *ssn*, *projNum*.

Theo Định nghĩa 1,2, lớp EMPPROJ đã ở 1FONF nhưng không ở 2FONF vì tồn tại các thuộc tính không khóa phụ thuộc hàm mờ bộ phân vào khóa, chẳng hạn *eName*, *pName*, *pLocation*. Tương tự như trong cơ sở dữ liệu quan hệ dư thừa dữ liệu xảy ra khi một người làm việc trong nhiều dự án khác nhau. Áp dụng thuật toán chuẩn hóa lớp về 2FONF như sau:

Bước 1, 2: Ta có các lớp EMPLOYEE(*ssn*, *eName*, *inputEMP*(...)), PROJECT(*projNum*, *pName*, *pLocation*, *inputPROJ*(...)), EMPPROJ(*ssn*, *projNum*, *hours*, *inputHour*(...)).

Bước 3: Lớp EMPPROJ được chuyển đổi thành EMPPROJ(employee: EMPLOYEE, project: PROJECT, hours, *inputHour*(...)).

#### 4.2.3. Thuật toán chuẩn hóa lớp về 3FONF

Vào: Lớp đối tượng mờ  $c(U)$  ở 1FONF,  $F$ : tập các phụ thuộc hàm mờ tối thiểu, khóa  $K$  của lớp  $c$ .

Ra: Tách lớp  $c$  thành các lớp  $c_i$  ở 3FONF không mất mát thông tin.

##### **Bước 1.**

Đặt  $S = \bigcup A_i$  với  $A_i$  không có mặt trong bất kỳ phụ thuộc hàm mờ nào.

Với mỗi vế trái  $X_i$  của phụ thuộc hàm mờ trong  $F$ , đặt  $S_i = X_i \cup \{A_1\} \cup \{A_2\} \cup \dots \cup \{A_k\}$  trong đó  $X_i \xrightarrow{fc} A_1, X_i \xrightarrow{fc} A_2, \dots, X_i \xrightarrow{fc} A_k$  là các phụ thuộc hàm mờ trong  $F$ .

Với mỗi  $S_i$  tạo lớp mới  $c_i(S_i)$  với khóa là  $X_i$ , các phương thức  $M$  phụ thuộc vào  $S_i$  được đưa vào lớp  $c_i$ .

##### **Bước 2.**

Đặt  $T$  là tập các lớp  $c_i$  được tạo ra ở bước 1, nếu không tồn tại một lớp  $c_i$  có khóa  $K$  thì tạo một lớp mới  $c$  bằng cách bổ sung tập thuộc tính  $S$  vào  $c_i \in T$  mà có khóa  $(K - S)$ . Các

phương thức  $M$  phụ thuộc vào  $K$  được phân vào lớp  $c(K)$ .

**Bước 3.**

Với mỗi lớp  $c_i \in T$ , nếu  $c_i$  có khóa  $H$  mà  $H \subset K$  thì thay tập thuộc tính  $H$  trong  $c$  bởi thuộc tính  $h$  có kiểu lớp  $c_i$ .

Với mỗi phương thức  $M$  trong các lớp được viết lại cho phù hợp dựa vào các thuộc tính và kiểu của nó trong lớp đó.

Thuật toán tách lớp  $c$  thành các lớp ở 1FONF, (2FONF, 3FONF) là không làm mất mát thông tin. Ta có thể kiểm tra việc tách lớp  $c(U)$  thành các lớp  $c_i$  là không làm mất thông tin dựa vào Bổ đề 2 trong [7].

**Ví dụ 3.** Cho lớp Teacher như sau:

Class TEACHER

ATTRIBUTE:

*ssn*: integer;

*tName*: string;

*salary*: fuzzy real;

*dept*: string;

*dName*: string;

*dLocation*: string;

WEIGHT: ...

METHODS:

InputTEA(*ssn*, *tName*, *salary*);

InputDEPT(*dept*, *dName*, *dLocation*);

*pD*: domain[0,1] of real;

END;

Khóa của lớp Teacher là *ssn*

Tập phụ thuộc hàm mờ trên lớp Teacher  $F = \left\{ ssn \xrightarrow{fc} tName; ssn \xrightarrow{fc} salary; ssn \xrightarrow{fc} dept; dept \xrightarrow{fc} dLocation, dept \xrightarrow{fc} dName \right\}$ .

Theo Định nghĩa 12, lớp TEACHER đã ở 1FONF nhưng không ở 3FONF vì tồn tại các thuộc tính không khóa phụ thuộc hàm mờ bậc cao vào khóa, chẳng hạn *dName*, *dLocation*. Cũng như trong cơ sở dữ liệu kinh điển, dị thường khi xóa bộ xuất hiện khi xóa một Khoa nào đó thì thông tin của tất cả các Giáo viên trong khóa đó sẽ bị mất đi. Áp dụng thuật toán chuẩn hóa lớp về 3FONF như sau:

Bước 1,2: Ta có các lớp TEACHER(*ssn*, *tName*, *salary*, *dept*, *inputTEA*(...)),

DEPARTMENT(*dept*, *dName*, *dLocation*, *inputDEPT*(...)).

Bước 3: Lớp TEACHER được chuyển đổi thành TEACHER(*ssn*, *tName*, *salary*, *department* : DEPARTMENT, *inputTEA*(...)), phương thức *inputDEPT*(...) phụ thuộc vào thuộc tính *department* có thể bỏ đi.

## 5. KẾT LUẬN

Một lược đồ CSDL HDT mờ/rõ ban đầu có thể được thiết kế chưa tốt có thể dẫn đến dư thừa dữ liệu. Sự dư thừa dữ liệu làm tổn hại đến tính toàn vẹn dữ liệu. Giống như các cơ sở dữ liệu khác, để người thiết kế có được một lược đồ CSDL HDT mờ thích hợp cần phải xem xét các phụ thuộc dữ liệu trong mỗi lớp đối tượng. Các phụ thuộc dữ liệu là tri thức ngữ nghĩa về thế giới thực và được xem như các ràng buộc toàn vẹn đối với việc thiết kế. Tương tự như trong cơ sở dữ liệu quan hệ rõ (mờ), có thể đưa các dạng chuẩn đối tượng mờ và các thuật toán chuẩn hóa lớp đối tượng về 1FONF, 2FONF, 3FONF để hỗ trợ cho việc thiết kế cơ sở dữ liệu đối tượng, đối tượng quan hệ. Trên cơ sở bài báo này, phụ thuộc đường dẫn và phương pháp truy vấn đối tượng dựa trên đường dẫn trong CSDL HDT mờ sẽ được trình bày trong nghiên cứu tiếp theo.

## TÀI LIỆU THAM KHẢO

- [1] B.Bouchon-Meunier, Hồ Thuần, Đặng Thanh Hà, *Logic mờ và Ứng dụng*, Đại học Quốc gia Hà Nội, 2007.
- [2] Byung S.Lee, Normalization in OODB design, *ACM SIGMOD Record* **24** (3) (1995) 23–27.
- [3] Catriel Beeri, Ronald Fagin, and John H. Howard, A complete axiomatization for functional and multivalued dependencies in database relations, *Proceedings of ACM SIGMOD*, August, 1997 (47–63).
- [4] Christophe Lécluse, Philippe Richard, Fernando Velez,  $O_2$  - an object-oriented data model, *Proceedings of the ACM SIGMOD* **17** (3) (june 1988) 424–433.
- [5] Gloria Bordogna, Gabriella Pasi, *Recent Issue on Fuzzy Databases*, Physica-Verlag Heidelberg New York, 2000.
- [6] Gillian Dobbie, Toward normalization on object-oriented databases, “Technical Report CS-RS-96/16”.
- [7] Nguyễn Kim Anh, Chuẩn hoá sơ đồ cơ sở dữ liệu hướng đối tượng, *Tạp chí Tin học và Điều khiển học* (2003) 125–130.
- [8] P. J. Pratt and J. J. Adamski, *Database Systems Management and Design* (3<sup>rd</sup> edition), Boyd & Fraser Publishing Company, Danvers, MA, 1994 (597–598).
- [9] M. I. Sozat, A. Yazici, A complete axiomatization for fuzzy functional and multivalued dependencies in fuzzy database relations, *Fuzzy Sets and Systems* (2001) 161–183.
- [10] Tarek Sobh, *Advances in Computer and Information Sciences and Engineering*, Springer, NewYork, 2008 (300–304).
- [11] L. A. Zadeh, Fuzzy sets, *Information & Control* **8** (3) (1965) 338–353.
- [12] Zongmin Ma, *Fuzzy Database Modeling with XLM*, Springer, NewYork, 2005.

Ngày nhận bài 6 - 4 - 2011