

CÁC PHỤ THUỘC ĐỐI TƯỢNG MỜ DỰA TRÊN XẤP XỈ NGỮ NGHĨA THEO CÁCH TIẾP CẬN ĐẠI SỐ GIA TỬ*

ĐOÀN VĂN BAN¹, TRƯƠNG CÔNG TUẤN², ĐOÀN VĂN THẮNG³

¹*Viện Công nghệ Thông tin, Viện Hàn lâm Khoa học và Công nghệ Việt Nam*

²*Đại học Khoa học - Đại học Huế*

³*Cao đẳng Công thương*

Tóm tắt. Bài báo tập trung nghiên cứu phụ thuộc mờ giữa các thuộc tính và giữa thuộc tính với phương thức trong một lớp đối tượng mờ dựa trên xấp xỉ ngữ nghĩa theo cách tiếp cận đại số gia tử (ĐSGT). Tương tự như trong cơ sở dữ liệu (CSDL) quan hệ rõ/mờ, bài báo đưa ra các định nghĩa phụ thuộc thuộc tính mờ và phụ thuộc phương thức mờ. Ngoài ra, sử dụng các phụ thuộc mờ này để tìm ra câu trả lời gần đúng cho truy vấn Null trên mô hình CSDL hướng đối tượng (HĐT) với thông tin mờ và không chắc chắn.

Từ khóa. Cơ sở dữ liệu hướng đối tượng mờ, các phụ thuộc mờ, đại số gia tử.

Abstract. In this paper, we study fuzzy dependencies among attributes, between the attributes and methods in a fuzzy object class basing on approximate semantic. Similarly in the clear/fuzzy relational database, we present the definition of fuzzy attribute dependency and fuzzy method dependency. Furthermore, we applied fuzzy dependencies to find quite approximate answers for Null queries on object-oriented database model with fuzzy information and uncertainty.

Keywords. Fuzzy object-oriented database, fuzzy dependencies, hedge algebra.

1. MỞ ĐẦU

Mô hình CSDL quan hệ và hướng đối tượng mờ, các vấn đề liên quan đã được nhiều tác giả trong và ngoài nước quan tâm nghiên cứu rộng rãi trong những năm gần đây [1-11]. Để biểu diễn những thông tin mờ trong mô hình dữ liệu, có nhiều hướng tiếp cận cơ bản: mô hình dựa trên quan hệ tương tự [5] và mô hình phân bố khả năng [1],... Tất cả các cách tiếp cận trên nhằm mục đích nắm bắt và xử lý một cách thỏa đáng các thông tin không đầy đủ, không chính xác hoặc không chắc chắn.

Dựa vào những ưu điểm của cấu trúc đại số gia tử (ĐSGT) [7], các tác giả đã nghiên cứu mô hình CSDL quan hệ [8-10] và hướng đối tượng [2, 3] mờ dựa trên cách tiếp cận của ĐSGT, trong đó ngữ nghĩa ngôn ngữ được lượng hóa bằng các ánh xạ định lượng của ĐSGT. Theo cách tiếp cận của ĐSGT, ngữ nghĩa ngôn ngữ có thể được biểu thị bằng một lân cận các khoảng được xác định bởi độ đo tính mờ của các giá trị ngôn ngữ của một thuộc tính với vai trò là biến ngôn ngữ.

* Bài báo được thực hiện với sự hỗ trợ từ quỹ phát triển KHCVN (Nafosted), mã số 102.01-2011.06.

Trong bài báo này, dựa vào cách thức tính độ xấp xỉ ngữ nghĩa của hai dữ liệu mờ để định nghĩa các phụ thuộc của thuộc tính mờ và phụ thuộc phương thức mờ. Cuối cùng, áp dụng các phụ thuộc mờ này để tìm ra câu trả lời gần đúng cho truy vấn Null.

Bài báo được trình bày như sau: Mục 2 trình bày một số khái niệm cơ bản liên quan đến ĐSGT làm cơ sở cho các mục tiếp theo. Mục 3 trình bày các định nghĩa về phụ thuộc thuộc tính mờ, phụ thuộc phương thức mờ. Mục 4 trình bày các qui trình sử dụng phụ thuộc thuộc tính mờ, phụ thuộc phương thức mờ để tìm ra câu trả lời gần đúng và một số ví dụ minh họa cho ý tưởng của hướng tiếp cận này, và cuối cùng là kết luận.

2. MỘT SỐ KHÁI NIỆM CƠ BẢN

Sau đây là một số khái niệm về ánh xạ định lượng của Nguyễn Cát Hồ [7] và cách thức xác định các hệ lân cận ngữ nghĩa định lượng [8, 10].

2.1. Đại số gia tử

Cho một ĐSGT tuyến tính đầy đủ $\mathcal{AX} = (X, G, H, \Sigma, \Phi, \leq)$, trong đó $Dom(\mathcal{X}) = \mathbf{X}$ là miền các giá trị ngôn ngữ của thuộc tính ngôn ngữ \mathcal{X} được sinh tự do từ tập các phần tử sinh $G = \{\mathbf{1}, c^-, \mathbf{W}, c^+, \mathbf{0}\}$ bằng việc tác động tự do các phép toán một ngôi trong tập H , Σ và Φ là hai phép tính với ngữ nghĩa là cận trên đúng và cận dưới đúng của tập $H(x)$, tức là $\Sigma x = \text{supermum}H(x)$ and $\Phi x = \text{infimum}H(x)$, trong đó $H(x)$ là tập các phần tử sinh ra từ x , còn quan hệ \leq là quan hệ sắp thứ tự tuyến tính trên \mathbf{X} cảm sinh từ ngữ nghĩa của ngôn ngữ.

Cho tập các gia tử $H = H^- \cup H^+$, trong đó $H^+ = \{h_1, \dots, h_p\}$ và $H^- = \{h_{-1}, \dots, h_{-q}\}$, đều là tuyến tính với thứ tự như sau: $h_1 < \dots < h_p$ và $h_{-1} < \dots < h_{-q}$, trong đó $p, q > 1$. Ký hiệu $fm : X \rightarrow [0, 1]$ là độ đo tính mờ của ĐSGT \mathcal{AX} . Khi đó ta có mệnh đề sau:

Mệnh đề 2.1. *Độ đo tính mờ fm và độ đo tính mờ của gia tử $\mu(h)$, $\forall h \in H$, có các tính chất sau:*

- (1) $fm(hx) = \mu(h)fm(x), \forall x \in \mathbf{X}$.
- (2) $fm(c^-) + fm(c^+) = 1$.
- (3) $\sum_{-q \leq i \leq p, i \neq 0} fm(h_i c) = fm(c)$, trong đó $c \in \{c^-, c^+\}$.
- (4) $\sum_{-q \leq i \leq p, i \neq 0} fm(h_i x) = fm(x), x \in \mathbf{X}$.
- (5) $\sum \{\mu(h_i) : -q \leq i \leq -1\} = \alpha$ và $\sum \{\mu(h_i) : 1 \leq i \leq p\} = \beta$, trong đó $\alpha, \beta > 0$ và $\alpha + \beta = 1$.

Trong đại số gia tử, mỗi phần tử $x \in \mathbf{X}$ đều mang dấu âm hay dương, được gọi là PN-dấu và được định nghĩa đệ quy như sau:

Định nghĩa 2.1. (hàm PN-dấu Sgn) $Sgn: X \rightarrow \{-1, 0, 1\}$ là hàm dấu được xác định như sau, ở đây $h, h \in H$, và $c \in \{c^-, c^+\}$:

- (1) $Sgn(c^-) = -1, Sgn(c^+) = +1$.
- (2) $Sgn(h'hx) = 0$, nếu $h'hx = hx$, còn ngược lại ta có
 $Sgn(h'hx) = -Sgn(hx)$, nếu $h'hx \neq hx$ và h' là âm tính đối với h ,
 $Sgn(h'hx) = +Sgn(hx)$, nếu $h'hx \neq hx$ và h' dương tính đối với h .

Mệnh đề 2.2. *Với $\forall x \in \mathbf{X}$, ta có: $\forall h \in \mathbf{H}$, nếu $Sgn(hx) = +1$ thì $hx > x$, nếu $Sgn(hx) = -1$ thì $hx < x$ và nếu $Sgn(hx) = 0$ thì $hx = x$.*

Với các tính chất của tính mờ và hàm PN -dấu, ánh xạ ngữ nghĩa định lượng của ĐSGT được định nghĩa như sau.

Định nghĩa 2.2. Giả sử $\mathcal{A} = (X, G, H, \Sigma, \Phi, \leq)$ là một ĐSGT đầy đủ, tuyến tính và tự do, $fm(x)$ và $\mu(h)$ tương ứng là các độ đo tính mờ của ngôn ngữ và của gia tử h thỏa mãn các tính chất trong Mệnh đề 2.1. Khi đó, ta nói v là ánh xạ cảm sinh bởi độ đo tính mờ fm của ngôn ngữ nếu nó được xác định như sau:

- (1) $v(W) = k = fm(c^-)$, $v(c^-) = k - \alpha fm(c^-) = \beta fm(c^-)$, $v(c^+) = k + \alpha fm(c^+)$.
- (2) $v(h_j x) = v(x) + Sgn(h_j x) \{ \sum_{i=Sgn(j)}^j \mu(h_i) fm(x) - \omega(h_j x) \mu(h_j) fm(x) \}$, trong đó $\omega(h_j x) = \frac{1}{2} [1 + Sgn(h_j x) Sgn(h_p h_j x) (\beta - \alpha)] \in \{ \alpha, \beta \}$, với mọi j , $-q \leq j \leq p$ và $j \neq 0$.
- (3) $v(\Phi c^-) = 0$, $v(\Sigma c^-) = k = v(\Phi c^+)$, $v(\Sigma c^+) = 1$, và với mọi j , $-q \leq j \leq p$ và $j \neq 0$.

Ta có: $v(\Phi h_j x) = v(x) + Sgn(h_j x) \{ \sum_{i=sign(j)}^{j-1} \mu(h_i) fm(x) \}$ và

$$v(\Sigma h_j x) = v(x) + Sgn(h_j x) \{ \sum_{i=sign(j)}^j \mu(h_i) fm(x) \}.$$

2.2. Khoảng mờ của hai khái niệm mờ

Giả sử thuộc tính A có miền tham chiếu thực là khoảng $[a, b]$. Để chuẩn hóa, nhờ một phép biến đổi tuyến tính, ta giả thiết mọi miền như vậy đều là khoảng $[0, 1]$. Khi đó, tính chất (2) trong Mệnh đề 2.1 cho phép ta xây dựng hai khoảng mờ của hai khái niệm nguyên thủy c^- và c^+ , ký hiệu là $I(c^-)$ và $I(c^+)$ với độ dài tương ứng là $fm(c^-)$ và $fm(c^+)$ sao cho chúng tạo thành một phân hoạch của miền tham chiếu $[0, 1]$ và $fm(c^-)$ và $fm(c^+)$ là đồng biến với c^- và c^+ , tức là $c^- \leq c^+$ kéo theo $I(c^-) \leq I(c^+)$.

Một cách quy nạp, giả sử rằng với $\forall x \in X_{k-1} = \{x \in \mathbf{X} : x \text{ có độ dài } |x| = k-1\}$, ta đã xây dựng được hệ các khoảng mờ $\{I(x) : x \in X_{k-1} \text{ và } |I(x)| = fm(x)\}$ sao cho chúng là đồng biến và tạo thành một phân hoạch của đoạn $[0, 1]$. Khi đó, trên mỗi khoảng mờ $I(x)$, độ dài $fm(x)$ của $x \in X_{k-1}$, nhờ tính chất (4) trong Mệnh đề 2.1, ta có thể xây dựng được họ $\{I(h_i x) : q \leq i \leq p, i \neq 0, |I(h_i x)| = fm(h_i x)\}$ sao cho chúng là một phân hoạch của khoảng mờ $I(x)$. Có thể thấy họ $\{I(h_i x) : q \leq i \leq p, i \neq 0, |I(h_i x)| = fm(h_i x) \text{ và } x \in X_{k-1}\} = \{I(y) : y \in X_k \text{ và } |I(y)| = fm(y)\}$ là một phân hoạch của $[0, 1]$. Các khoảng này gọi là các khoảng mờ mức k .

Định nghĩa 2.3. Cho $P^k = \{I(x) : x \in X_k\}$ với $X_k = \{x \in \mathbf{X} : |x| = k\}$ là một phân hoạch của $[0, 1]$. Ta nói rằng u bằng v theo mức k trong P^k , được ký hiệu $u \approx_k v$, khi và chỉ khi $I(u)$ và $I(v)$ cùng thuộc một khoảng trong P^k . Có nghĩa $\forall x, y \in \mathbf{X}$, $u \approx_k v \Leftrightarrow \exists \Delta^k \in P^k : I(u) \subseteq \Delta^k \text{ và } I(v) \subseteq \Delta^k$.

2.3. Lân cận mức k

Xét một CSDL $\{U; Const\}$, trong đó $U = \{A_1, A_2, \dots, A_n\}$ là tập vũ trụ các thuộc tính, $Const$ là một tập các ràng buộc dữ liệu của CSDL. Mỗi thuộc tính A được gắn với một miền giá trị thuộc tính, ký hiệu là $Dom(A)$, trong đó một số thuộc tính cho phép nhận các giá trị ngôn ngữ trong lưu trữ trong CSDL hay trong các câu hỏi truy vấn và được gọi là thuộc tính ngôn ngữ. Những thuộc tính còn lại được gọi là thuộc tính thực hay kinh điển. Thuộc

tính thực A được gắn với một miền giá trị kinh điển, ký hiệu là D_A . Thuộc tính ngôn ngữ A sẽ được gắn một miền giá trị kinh điển D_A và một miền giá trị ngôn ngữ LD_A hay là tập các phần tử của một DSGT. Để bảo đảm tính nhất quán trong xử lý ngữ nghĩa dữ liệu trên cơ sở thống nhất kiểu dữ liệu của thuộc tính ngôn ngữ, mỗi thuộc tính ngôn ngữ sẽ được gắn với một ánh xạ định lượng $v_A : LD_A \rightarrow D_A$ được xác định bởi một bộ tham số định lượng của A . Như vậy, mỗi giá trị ngôn ngữ x của A sẽ được gán một nhân giá trị thực $v_A \in D_A$ được xem như giá trị đại diện của x . Việc đánh giá độ tương tự giữa các dữ liệu của một thuộc tính A được dựa trên khái niệm lân cận mức k của một giá trị ngôn ngữ, với k là số nguyên dương.

Các tác giả trong [8, 10] đã lấy các khoảng mờ của các phần tử độ dài k làm độ tương tự giữa các phần tử, nghĩa là các phần tử mà các giá trị đại diện của chúng thuộc cùng một khoảng mờ mức k là tương tự mức k . Tuy nhiên, theo cách xây dựng các khoảng mờ mức k , giá trị đại diện của các phần tử x có độ dài nhỏ hơn k luôn là đầu mút của các khoảng mờ mức k . Do vậy, khi xác định lân cận mức k mong muốn, các giá trị đại diện như vậy phải là điểm trong của lân cận mức k .

Ta luôn luôn giả thiết rằng mỗi tập H^- và H^+ chứa ít nhất 2 gia tử. Xét \mathbf{X}_k là tập tất cả các phần tử độ dài k . Dựa vào khoảng mờ mức k và mức $k+1$ các tác giả [8, 10] đã xây dựng một phân hoạch của miền $[0, 1]$ như sau:

(1) *Độ tương tự mức 1*: Với $k = 1$, các khoảng mờ mức 1 gồm $I(c^-)$ và $I(c^+)$. Các khoảng mờ mức 2 trên khoảng $I(c^+)$ là $I(h_{-q}c^+) \leq I(h_{-q+1}c^+) \dots \leq I(h_{-2}c^+) \leq I(h_{-1}c^+) \leq v_A(c^+) \leq I(h_1c^+) \leq I(h_2c^+) \leq \dots \leq I(h_{p-1}c^+) \leq I(h_pc^+)$. Khi đó, ta xây dựng phân hoạch về độ tương tự mức 1 gồm các lớp tương đương sau: $S(\mathbf{0}) = I(h_pc^-)$; $S(c^-) = I(c^-) \setminus [I(h_{-q}c^-) \cup I(h_pc^-)]$; $S(\mathbf{W}) = I(h_{-q}c^-) \cup I(h_{-q}c^+)$; $S(c^+) = I(c^+) \setminus [I(h_{-q}c^+) \cup I(h_pc^+)]$ và $S(\mathbf{1}) = I(h_pc^+)$.

Ta thấy, trừ hai điểm đầu mút $v_A(\mathbf{0}) = 0$ và $v_A(\mathbf{1}) = 1$, các giá trị đại diện $v_A(c^-)$, $v_A(\mathbf{W})$, $v_A(c^+)$ đều là điểm trong tương ứng của các lớp tương tự mức 1 $S(c^-)$, $S(\mathbf{W})$ và $S(c^+)$.

(2) *Độ tương tự mức 2*: Với $k = 2$, chẳng hạn, $I(h_ic^+) = (v_A(\Phi h_ic^+), v_A(\Sigma h_ic^+))$ với hai khoảng mờ kề là $I(h_{i-1}c^+)$ và $I(h_{i+1}c^+)$, ta sẽ có các lớp tương đương dạng sau: $S(h_ic^+) = I(h_ic^+) \setminus [I(h_ph_ic^+) \cup I(h_{-q}h_ic^+)]$; $S(\Phi h_ic^+) = I(h_{-q}h_{(i-1)}c^+) \cup I(h_{-q}h_ic^+)$ và $S(\Sigma h_ic^+) = I(h_ph_ic^+) \cup I(h_ph_ic^+)$, với i sao cho $-q \leq i \leq p$ và $i \neq 0$.

Bằng cách tương tự như vậy, có thể xây dựng các phân hoạch các lớp tương tự mức k bất kỳ. Tuy nhiên, trong thực tế ứng dụng theo [6] thì $k \leq 4$, tức có tối đa 4 gia tử tác động liên tiếp lên phần tử nguyên thủy c^- và c^+ . Các giá trị rõ và các giá trị mờ gọi là có độ tương tự mức k nếu các giá trị đại diện của chúng cùng nằm trong một lớp tương tự mức k .

Lân cận mức k của khái niệm mờ: Giả sử phân hoạch các lớp tương tự mức k là các khoảng $S(x_1), S(x_2), \dots, S(x_m)$. Khi đó, mỗi giá trị ngôn ngữ fu chỉ và chỉ thuộc về một lớp tương tự, chẳng hạn đó là $S(x_i)$ và nó gọi là lân cận mức k của fu và ký hiệu là $FRN_k(fu)$.

3. PHƯƠNG PHÁP XÁC ĐỊNH CÁC PHỤ THUỘC MỜ DƯỚI TRÊN XẤP XỈ NGỮ NGHĨA

Ví dụ 1. Xét lớp Sinh Viên được cho ở Hình 3.1. Lớp *SinhVien* được miêu tả bởi tập thuộc tính *tenSV*, *tuoi*, *chieuCao*, *canNang*, và phương thức *sucKhoe()*. Bảng 1 là một số thể hiện của lớp *SinhVien*, để đơn giản giới hạn bảng dữ liệu chỉ gồm những thuộc tính.

SinhVien
tenSV
tuoi
chieuCao
canNang
sucKhoe()

Hình 3.1. Lớp đối tượng Sinh Viên

Bảng 1. Thể hiện của lớp SinhVien

SinhVien					
<i>iDSV</i>	<i>tenSV</i>	<i>tuoi</i>	<i>chieuCao</i>	<i>canNang</i>	<i>sucKhoe()</i>
Id1	An	khoảng 18	hơn cao	khoảng 62	
Id2	Binh	hơn trẻ	[1.62, 1.68]	khoảng 60	
Id3	Hà	[15, 24]	[1.63, 1.66]	trung bình	
Id4	Hương	[26, 30]	cao	khoảng 64	
Id5	Nhân	[21, 25]	[1.6, 1.65]		
Id6	Thủy	trẻ	hơn cao	[70, 71]	

Các đối tượng thuộc lớp *SinhVien* có thể mờ bởi các lý do sau: Thuộc tính *tuoi*, *chieuCao*, *canNang* có thể biểu diễn bằng giá trị số hoặc nhãn không xác định giá trị chính xác, chẳng hạn: *trẻ*, *già*,... đối với thuộc tính *tuoi*; *thấp*, *cao*,... với thuộc tính *chieuCao* hoặc *nặng*, *nhẹ*, ... với thuộc tính *canNang*.

Vậy, giá trị thuộc tính của đối tượng có thể là một trong bốn trường hợp sau: giá trị rõ, giá trị mờ, tham chiếu đến đối tượng (đối tượng này có thể mờ) và sưu tập (collection). Trong bài báo này, sẽ tập trung xem xét trường hợp thứ 1 và 2 đó là: giá trị chính xác (giá trị rõ) và giá trị không chính xác (giá trị mờ) và xem giá trị rõ là trường hợp riêng của giá trị mờ. Giá trị mờ thường rất phức tạp và nhãn ngôn ngữ [1] thường được sử dụng để biểu diễn cho những loại giá trị này. Miền giá trị thuộc tính mờ là hợp của hai thành phần

$$Dom(a_i) = CDom(a_i) \cup FDom(a_i) (1 \leq i \leq n),$$

trong đó:

- $CDom(a_i)$: Miền giá trị rõ của thuộc tính a_i .
- $FDom(a_i)$: Miền giá trị mờ của thuộc tính a_i .

3.1. Chuyển các giá trị thuộc tính về giá trị khoảng $[a, b]$

Giá trị thuộc tính mờ là hợp của hai thành phần giá trị rõ và giá trị mờ. Đối với giá trị rõ, các kiểu dữ liệu bao gồm các loại đơn giản như số nguyên, thực, chuỗi. Đối với giá trị mờ, giá trị ngôn ngữ thường được sử dụng để biểu diễn thông tin không chính xác. Như vậy, ta cần phải có phương pháp biểu diễn một cách thống nhất các dạng dữ liệu, nhằm thuận tiện cho việc đánh giá độ *xấp xỉ ngữ nghĩa* giữa chúng.

Trong phần này, sẽ trình bày phương pháp biến đổi các giá trị này về các khoảng $[a, b]$. Với phương pháp biểu diễn này, sẽ xem xét các kiểu dữ liệu khác nhau trên một quan điểm thống nhất. Phương pháp chuyển về giá trị khoảng được thực hiện như sau:

- (1) Giá trị thuộc tính là giá trị số:
- Nếu giá trị thuộc tính là a thì chuyển thành $[a, a]$. Ví dụ, giá trị thuộc tính $a = 10$ thì khi chuyển thành khoảng sẽ là $[10, 10]$.
 - Nếu giá trị thuộc tính là vào khoảng a thì chuyển thành $[a - \varepsilon, a + \varepsilon]$, với ε là bán kính với tâm a . Ví dụ, giá trị thuộc tính là vào khoảng 10 và chọn $\varepsilon = 1$ thì khi chuyển thành giá trị khoảng sẽ là $[9, 11]$.
 - Nếu giá trị thuộc tính là a đến b thì chuyển thành $[a, b]$. Ví dụ, nếu giá trị thuộc tính là 24 đến 26 thì khi chuyển thành giá trị khoảng sẽ là $[24, 26]$.
- (2) Giá trị thuộc tính là giá trị ngôn ngữ:
- (a) Thuộc tính của đối tượng được xem như là thuộc tính ngôn ngữ và được biểu diễn theo cấu trúc DSGT. Từ đó, xây dựng phân hoạch các lớp tương tự mức k cho thuộc tính ngôn ngữ này.
- (b) Xác định giá trị ngôn ngữ thuộc lớp tương tự mức k và từ đó xác định được khoảng giá trị tương ứng cho giá trị thuộc tính này.

3.2. Hàm đo xấp xỉ ngữ nghĩa

Từ kết quả chuyển các giá trị thuộc tính về các khoảng $[a, b]$ tương ứng, ta xem xét tiếp mối quan hệ xấp xỉ ngữ nghĩa (*Semantic Proximity*) của các giá trị mờ, ký hiệu $SP(f_1, f_2)$ ($0 \leq SP(f_1, f_2) \leq 1$). Dựa vào độ lân cận [11] của các giá trị mờ, xấp xỉ ngữ nghĩa (SP) phải thỏa mãn một số tính chất sau:

Cho $f_1 = [a_1, b_1]$, $f_2 = [a_2, b_2]$, $g_1 = [c_1, d_1]$, $g_2 = [c_2, d_2]$, và $|b_1 - a_1|$ là độ dài của f_1 .

1. Nếu f_1, f_2 là hai khoảng bằng nhau thì xấp xỉ ngữ nghĩa của f_1 và f_2 bằng 1 (nghĩa là, nếu $a_1 = b_1 = a_2 = b_2$ thì $SP(f_1, f_2) = 1$).
2. Nếu f_1, f_2 là hai khoảng không giao nhau, thì xấp xỉ ngữ nghĩa của f_1 và f_2 bằng 0 (nghĩa là $f_1 \cap f_2 = \emptyset$ thì $SP(f_1, f_2) = 0$).
3. Nếu (f_1, f_2) và (g_1, g_2) là hai cặp giống nhau, và chiều dài f_1 nhỏ hơn chiều dài của g_1 , thì $SP(f_1, f_2)$ lớn hơn $SP(g_1, g_2)$ (nghĩa là, $a_1 = a_2, b_1 = b_2, c_1 = c_2, d_1 = d_2$ và $|d_1 - c_1| \geq |b_1 - a_1|$, thì $SP(f_1, f_2) \geq SP(g_1, g_2)$).
4. Nếu độ dài của f_1 bằng độ dài của f_2 và độ dài giao của f_1, g_1 lớn hơn độ dài giao của f_2, g_1 thì $SP(f_1, g_1)$ lớn hơn $SP(f_2, g_1)$ (nghĩa là, nếu $|a_2 - b_2| \geq |a_1 - b_1|$ và $|f_1 \cap g_1| \geq |f_2 \cap g_1|$ thì $SP(f_1, g_1) \geq SP(f_2, g_1)$).

Từ đó, hàm xấp xỉ ngữ nghĩa được định nghĩa như sau.

Định nghĩa 3.1. Cho lớp mờ C xác định trên tập thuộc tính A và tập phương thức M , $a_i \subseteq A$ ($1 \leq i \leq n$). Hàm đánh giá độ xấp xỉ ngữ nghĩa giữa hai giá trị thuộc tính a_i của hai đối tượng $o_1.a_i, o_2.a_i \subseteq C$, có hai khoảng tương ứng là $[l_1, u_1]$ và $[l_2, u_2]$, ký hiệu $SP(o_1.a_i, o_2.a_i)$ (*Semantic Proximity*), được xác định như sau

$$SP(o_1.a_i, o_2.a_i) = [o_1.a_i \cap o_2.a_i] / [o_1.a_i \cup o_2.a_i] - [o_1.a_i \cap o_2.a_i] / \alpha,$$

trong đó: $[h]$ là độ dài của khoảng h được định nghĩa như sau:

$$[h] = \begin{cases} 0 & h = \emptyset \\ \delta & h = [a, a] \\ |b - a| & h = [a, b] \text{ v } \neq b \neq a \\ \alpha & h = \infty \end{cases}$$

và hệ số α được xác định là giá trị lớn nhất của miền giá trị thuộc tính, δ là số tương đối nhỏ, chẳng hạn, $\delta = \alpha/1000$.

Để dàng nhận thấy hàm đo xấp xỉ ngữ nghĩa SP thỏa mãn các tính chất ở trên.

Ví dụ, giả sử phạm vi của miền giá trị là 10, lúc đó $\alpha = 10$ và $\delta = \alpha/1000$.

- 1) Giả sử $f_1 = [10, 10]$ và $f_2 = [10, 10]$, ta có $SP(f_1, f_2) = \delta/\delta - \delta/10 = 1$.
- 2) Giả sử $f_1 = [2, 5]$ và $f_2 = [6, 7]$. ta có $SP(f_1, f_2) = 0/4 - 0/10 = 0$.
- 3) Giả sử $f_1 = [2, 5]$ và $f_2 = [2, 5]$, $g_1 = [2, 8]$ và $g_2 = [2, 8]$, thì $SP(f_1, f_2) = 3/3 - 3/10 = 0.7$ và $SP(g_1, g_2) = 6/6 - 6/10 = 0.4$. Ta có $SP(f_1, f_2) \geq SP(g_1, g_2)$.
- 4) Giả sử $f_1 = [3, 5]$, $f_2 = [2, 4]$, $g_1 = [3, 6]$, thì $SP(f_1, g_1) = 2/3 - 2/10 = 0.5$ và $SP(f_2, g_1) = 1/4 - 1/10 = 0.15$. Ta có $SP(f_1, g_1) \geq SP(f_2, g_1)$.

3.3. Phụ thuộc thuộc tính mờ

Khái niệm phụ thuộc thuộc tính mờ của đối tượng trên cơ sở mở rộng của khái niệm phụ thuộc hàm mờ của mô hình CSDL quan hệ mờ dựa trên DSGT [9]. Phụ thuộc thuộc tính mờ (Fuzzy Attribute Dependency, viết tắt là FAD) được định nghĩa như sau.

Định nghĩa 3.2. Cho lớp mờ C với tập thuộc tính $U, X, Y \subseteq U$ và X là tập các thuộc tính nhận giá trị đơn (giá trị rõ hoặc mờ). Ta định nghĩa Y phụ thuộc thuộc tính mờ vào X , ký hiệu $X \overset{fad}{\sim} Y$ khi và chỉ khi $\forall o_1, o_2 \in C$ nếu $SP(o_1.X, o_2.X) \leq SP(o_1.Y, o_2.Y)$ ($o_1.X$ - giá trị của o_1 trên X).

Định lý 3.1. *Phụ thuộc hàm trong cơ sở dữ liệu quan hệ thỏa mãn định nghĩa phụ thuộc thuộc tính 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 $\forall o_1, o_2 \in C$ ta có: $SP(o_1.X, o_2.X) = SP(o_1.Y, o_2.Y) = 1$.

Do đó, $SP(o_1.X, o_2.X) \leq SP(o_1.Y, o_2.Y)$. Theo định nghĩa phụ thuộc thuộc tính mờ trong CSDL HDT mờ $X \overset{fad}{\sim} Y$. ■

Các luật suy dẫn FAD như sau:

U là tập các thuộc tính của lớp C ; $X, Y, W \subseteq U$:

Luật 1.1: Phản xạ. Nếu $Y \subseteq X$ thì $X \overset{fad}{\sim} Y$.

Luật 1.2: Tăng trưởng. Nếu $X \overset{fad}{\sim} Y$ và $Y \subseteq U$ thì $XZ \overset{fad}{\sim} YZ$.

Luật 1.3: Bắc cầu. Nếu $X \overset{fad}{\sim} Y$ và $Y \overset{fad}{\sim} Z$ thì $X \overset{fad}{\sim} Z$.

Định lý 3.2. *Các luật (suy dẫn) 1.1 đến 1.3 là xác đáng và đầy đủ.*

Việc chứng minh định lý tương tự như trường hợp phụ thuộc hàm mờ trong CSDL quan hệ mờ [4, 8, 9].

Ví dụ 2. Phụ thuộc *tuoi* \xrightarrow{fad} *chieuCao* thỏa mãn định nghĩa FAD trong lớp SinhVien ở Ví dụ 1.

Thật vậy, đặt o_1, o_2 lần lượt là hai đối tượng của lớp SinhVien. Trước tiên, chuyển các giá trị thuộc tính *tuoi* và *chieuCao* của hai đối tượng này thành các giá trị khoảng.

1. Thuộc tính *tuoi*:

a) Áp dụng phương pháp 1 ở Mục 3.1 để chuyển giá trị thuộc tính là giá trị số thành các giá trị khoảng.

b) Xây dựng ngữ nghĩa định lượng cho thuộc tính trong trường hợp giá trị thuộc tính là giá trị ngôn ngữ ngôn ngữ. Xét ĐSGT của biến ngôn ngữ *tuoi*, trong đó $D_{tuoi} = [0, 100]$, các phần tử sinh là $\{0, \text{trẻ}, W, \text{già}, 1\}$, tập các gia tử là $\{\text{ít}, \text{khả năng}, \text{hơn}, \text{rất}\}$, $FD_{tuoi} = H_{tuoi}(\text{già}) \cup H_{tuoi}(\text{trẻ})$. Chọn $fm(\text{già}) = 0.6$, $fm(\text{trẻ}) = 0.4$, $\mu(\text{khả năng}) = 0.25$, $\mu(\text{ít}) = 0.2$, $\mu(\text{hơn}) = 0.15$ và $\mu(\text{rất}) = 0.4$. Phân hoạch đoạn $[0, 100]$ thành các khoảng tương tự mức 1 và 2 như sau: Khoảng tương tự mức 1: $fm(\text{rất trẻ}) \times 100 = 0.4 \times 0.4 \times 100 = 16$. vậy $S(0) = [0, 16]$. $(fm(\text{khả năng trẻ}) + fm(\text{hơn trẻ})) \times 100 = (0.25 \times 0.4 + 0.15 \times 0.4) \times 100 = 16$, vậy $S(\text{trẻ}) = [16, 32]$. Và bằng cách tính tương tự ta có $S(W) = [32, 52]$, $S(\text{già}) = [52, 76]$, $S(1) = [76, 100]$. Và mức 2: $fm(\text{rất rất trẻ}) \times 100 = 0.4 \times 0.4 \times 0.4 \times 100 = 6.4$. vậy $S(0) = [0, 6.4]$. $(fm(\text{khả năng rất trẻ}) + fm(\text{hơn rất trẻ})) \times 100 = (0.45 \times 0.25 \times 0.4 + 0.15 \times 0.4 \times 0.4) \times 100 = 6.4$, vậy $S(\text{rất trẻ}) = (6.4, 12.8]$. Và bằng cách tính tương tự ta có các khoảng là: $S(\text{hơn trẻ}) = (18.4, 20.8]$, $S(\text{khả năng trẻ}) = (26, 30]$, $S(\text{ít trẻ}) = (35.2, 38.4]$, $S(W) = (38.4, 42.4]$, $S(\text{ít già}) = (42.4, 47.2]$, $S(\text{khả năng già}) = (55, 61]$, $S(\text{hơn già}) = (68.8, 72.4]$, $S(\text{rất già}) = (80.8, 90.4]$, $S(1) = (90.4, 100]$.

2. Thuộc tính *chieuCao*:

a) Áp dụng phương pháp 1 ở Mục 3.1 để chuyển giá trị thuộc tính là giá trị số thành các giá trị khoảng.

b) Xây dựng ngữ nghĩa định lượng cho thuộc tính trong trường hợp giá trị thuộc tính là giá trị ngôn ngữ ngôn ngữ. Xét ĐSGT của biến ngôn ngữ *chieuCao*, trong đó $D_{chieuCao} = [0, 2]$, các phần tử sinh là $\{0, \text{thấp}, W, \text{cao}, 1\}$, tập các gia tử là $\{\text{ít}, \text{khả năng}, \text{hơn}, \text{rất}\}$, $FD_{chieuCao} = H_{chieuCao}(\text{cao}) \cup H_{chieuCao}(\text{thấp})$. Chọn $fm(\text{thấp}) = 0.6$, $fm(\text{cao}) = 0.4$, $\mu(\text{khả năng}) = 0.2$, $\mu(\text{ít}) = 0.25$, $\mu(\text{hơn}) = 0.25$ và $\mu(\text{rất}) = 0.2$. Phân hoạch đoạn $[0, 2]$ thành các khoảng tương tự mức 1 và 2. Bằng cách tính tương tự như trên ta có các khoảng tương mức 1 là: $S(1) = (1.84, 2]$, $S(\text{cao}) = (1.48, 1.84]$, $S(W) = (0.98, 1.48]$, $S(\text{thấp}) = (0.44, 0.98]$, $S(0) = [0, 0.44]$; và mức 2 là: $S(1) = (1.968, 2]$, $S(\text{rất cao}) = (1.896, 1.968]$, $S(\text{hơn cao}) = (1.726, 1.816]$, $S(\text{khả năng cao}) = (1.572, 1.644]$, $S(\text{ít cao}) = (1.402, 1.492]$, $S(W) = (1.277, 1.402]$, $S(\text{ít thấp}) = (1.142, 1.277]$, $S(\text{khả năng thấp}) = (0.914, 1.022]$, $S(\text{hơn thấp}) = (0.656, 0.791]$, $S(\text{rất thấp}) = (0.428, 0.536]$, $S(0) = [0, 0.428]$.

Kết quả nhận được như sau:

Áp dụng Định nghĩa 3.1 và 3.2, dễ dàng nhận thấy rằng phụ thuộc *tuoi* \xrightarrow{fad} *chieuCao* thỏa mãn định nghĩa FAD.

Bảng 2. Thể hiện của lớp SinhVien

SinhVien					
<i>iDSV</i>	<i>tenSV</i>	<i>tuoi</i>	<i>chieuCao</i>	<i>canNang</i>	<i>sucKhoe()</i>
Id1	An	[17, 19]	[1.726, 1.816]	khoảng 62	
Id2	Binh	[18.4, 20.8]	[1.62, 1.68]	khoảng 60	
Id3	Hà	[15, 24]	[1.63, 1.66]	trung bình	
Id4	Hương	[26, 30]	[1.44, 1.84]	khoảng 64	
Id5	Nhân	[21, 25]	[1.6, 1.65]		
Id6	Thủy	[16, 32]	[1.726, 1.816]	[70, 71]	

3.4. Phụ thuộc phương thức mờ

Việc xác định sự gắn kết giữa phương thức với các thuộc tính của đối tượng mờ được xác định thông qua việc sử dụng phương thức để xử lý các thuộc tính đó như đọc hay cập nhật, ... Mối quan hệ như vậy được gọi là phụ thuộc phương thức mờ (Fuzzy Method Dependency, viết tắt là FMD).

Định nghĩa 3.3. Cho M_j là một phương thức của một lớp C , X là tập các thuộc tính nhận giá trị đơn (giá trị rõ hoặc mờ) (có thể đọc hoặc sửa đổi). Quan hệ giữa X và M_j , được ký hiệu $X \xrightarrow{fmd} M_j$, nếu X được sử dụng bởi M_j thỏa mãn FMD.

Dựa vào các kết quả [4], ta có các luật suy dẫn FMD như sau:

Luật 2.1: Nếu $X \xrightarrow{fmd} M_j$ và $Y \xrightarrow{fmd} M_j$ thì $XY \xrightarrow{fmd} M_j$.

Luật 2.2: Nếu $X \xrightarrow{fmd} M_i$ và $X \xrightarrow{fmd} M_j$ thì $X \xrightarrow{fmd} M_i M_j$.

Luật 2.3: Nếu $X \xrightarrow{fmd} M_j$ và $Y \subseteq X$ thì $Y \xrightarrow{fmd} M_j$.

Ví dụ 3. Lớp SinhVien ở Ví dụ 1 có phương thức *sucKhoe()*, phương thức này sử dụng hai thuộc tính *tuoi* và *chieuCao* để xác định sức khỏe cho các đối tượng sinh viên. Trong Ví dụ 2 thuộc tính *tuoi* và *chieuCao* thỏa mãn FAD nên dễ dàng nhận thấy rằng phương thức *sucKhoe()* thỏa mãn FMD, nghĩa là $\{tuoi, chieuCao\} \xrightarrow{fmd} sucKhoe()$.

4. THUẬT TOÁN VÀ ỨNG DỤNG

Phụ thuộc hàm đối với lớp, phụ thuộc giữa các thuộc tính và phương thức trong một lớp các đối tượng (mờ) được nghiên cứu để giải quyết các vấn đề khác nhau trên mô hình cơ sở dữ liệu hướng đối tượng (mờ) như chuẩn hóa lớp đối tượng (mờ), xử lý truy vấn CSDL đối tượng (mờ),...

Một câu truy vấn được xem là truy vấn Null khi nó nhận được câu trả lời Null từ CSDL. Câu trả lời Null có thể được sinh ra do nhiều nguyên nhân khác nhau. Nguyên nhân thứ nhất, là khi dữ liệu trong CSDL không thỏa mãn điều kiện của câu truy vấn. Một nguyên nhân tiếp theo, là do dữ liệu không tồn tại trong CSDL, chẳng hạn ta truy vấn các bộ (tuple) dữ liệu từ quan hệ R , nhưng chúng không tồn tại trong CSDL. Một nguyên nhân nữa, đó là do thông tin không đầy đủ trong CSDL, nguyên nhân này thường là do giá trị thuộc tính

của một số đối tượng bị thiếu. Bài báo chỉ tập trung định giá câu trả lời Null được sinh ra do thiếu các giá trị thuộc tính của một số đối tượng.

Xét các đối tượng được cho trong Ví dụ 1. Giả sử rằng giá trị thuộc tính *canNang* của đối tượng *Nhan* là thiếu và câu truy vấn như sau: “*Cho biết cân nặng của Nhan*”; hoặc “*Cho biết sức khỏe của Nhan*”. Khi đó câu trả lời là Null.

Trong lý thuyết suy diễn tương tự [11], nếu gọi S là tập đối tượng nguồn, T là tập đối tượng đích, tập đối tượng nguồn và đích có tính chất tương tự nhau là P . Khi đó, nếu S có tính chất P' thì suy ra T có thể cũng có P' dựa vào tính chất P có trong cả S và T .

Suy diễn tương tự có thể áp dụng để tìm ra câu trả lời gần đúng cho truy vấn Null khá hiệu quả. Có hai vấn đề sẽ được giải quyết trong quá trình suy diễn:

- Thứ nhất, đó là xác định tính chất tương tự nhau P giữa tập đối tượng nguồn và đích, chẳng hạn, xét hai đối tượng nhân viên, đối tượng thứ nhất có tên là *An*, vị trí công việc là *kế toán* và mức lương là *90*, đối tượng thứ hai có tên là *Binh*, vị trí công việc là *kế toán* và mức lương *bị thiếu*. Câu truy vấn “*cho biết lương của Binh*”, rõ ràng câu trả lời là Null. Chúng ta có thể giả thiết rằng có một câu trả lời gần đúng cho câu trả lời Null đó là lương của *Binh* vào khoảng *90*, bởi vị trí công việc của *Binh* là *kế toán* và đó cũng chính là vị trí công việc của *An* và *An* có mức lương *90*. Tất nhiên, câu trả lời gần đúng có thể không đúng vì do nhiều yếu tố ảnh hưởng đến lương của nhân viên. Tuy nhiên, việc có được câu trả lời gần đúng có thể tốt hơn là không có câu trả lời khi định giá các câu truy vấn.
- Thứ hai, là tính chất (thuộc tính) P trong cả S và T có thể xác định tính chất (thuộc tính) Q không rõ. Vấn đề này liên quan đến phụ thuộc hàm (mờ) giữa các thuộc tính.

Bài báo áp dụng phụ thuộc mờ giữa các thuộc tính, phụ thuộc mờ giữa thuộc tính và phương thức để tìm ra câu trả lời gần đúng.

a. Đối với truy vấn thuộc tính, ta có lời giải thuật toán như sau

Thuật toán ASAO (Approximate Semantic for Attributes Objects)

Vào: Một lớp C với m thuộc tính và p phương thức; $O = \{o_1, o_2, \dots, o_n\} \in C$.

Trong đó: $(o_1, o_2, \dots, o_{n-1}).[a_1, a_2, \dots, a_m]$ và $o_n.[a_1, a_2, \dots, a_{m-1}]$ là được định nghĩa (có nghĩa là giá trị dữ liệu tồn tại) và $o_n.a_m$ là chưa được định nghĩa (có nghĩa là giá trị dữ liệu bị thiếu).

Ra: Tìm giá trị tương tự gần đúng cho $o_n.a_m$.

Phương pháp:

Bước 1: Áp dụng phương pháp ở Mục 3.1 để chuyển các giá trị thuộc tính tham gia truy vấn về các giá trị khoảng tương ứng.

Bước 2: Sử dụng phụ thuộc thuộc tính mờ ở Mục 3.3 để kiểm tra các FAD.

Bước 3: Xác định giá trị cho thuộc tính chưa định nghĩa $o_n.a_m$.

Ví dụ 4. Xét câu truy vấn trên “*Cho biết cân nặng của Nhan*”, với giá trị thuộc tính *canNang* của *Nhan* là chưa định nghĩa. Áp dụng qui trình trên để tìm ra câu trả lời gần đúng trong CSDL ở Ví dụ 1 như sau:

Bước 1: Chuyển giá trị thuộc tính *chieuCao* và *canNang* về các giá trị khoảng tương ứng.

- Thuộc tính *chieuCao*: giá trị thuộc tính chuyển về giá trị khoảng đã được tính trong Ví dụ 2.

- Thuộc tính *canNang*:

o Áp dụng phương pháp 1 ở Mục 3.1 để chuyển giá trị thuộc tính là giá trị số thành các giá trị khoảng.

o Xây dựng ngữ nghĩa định lượng cho thuộc tính trong trường hợp giá trị thuộc tính là giá trị ngôn ngữ ngôn ngữ. Xét ĐSGT của biến ngôn ngữ *canNang*, trong đó $D_{canNang} = [0, 100]$, các phần tử sinh là $\{0, nhẹ, W, nặng, 1\}$, tập các gia tử là $\{ít, khả năng, hơn, rất\}$, $FD_{canNang} = H_{canNang}(nặng) \cup H_{canNang}(nhẹ)$. Chọn $fm(nhẹ) = 0.6$, $fm(nặng) = 0.4$, $\mu(khả năng) = 0.15$, $\mu(ít) = 0.25$, $\mu(hơn) = 0.25$ và $\mu(rất) = 0.35$. Phân hoạch đoạn $[0, 100]$ thành các khoảng tương tự mức 2. Bằng cách tính tương tự như ở Ví dụ 2 ta có các khoảng tương tự mức 2 là: $S(1) = (95.1, 100]$, $S(rất nặng) = (89.5, 95.1]$, $S(hơn nặng) = (78.5, 82.5]$, $S(khả năng nặng) = (71.5, 73.9]$, $S(ít nặng) = (62.5, 66.5]$, $S(W) = (56.25, 62.5]$, $S(ít nhẹ) = (50.25, 56.25]$, $S(khả năng nhẹ) = (39.15, 42.75]$, $S(hơn nhẹ) = (26.25, 32.25]$, $S(rất nhẹ) = (7.35, 15.75]$, $S(0) = [0, 7.35]$.

Kết quả nhận được như sau:

Bảng 3. Thể hiện của lớp SinhVien

SinhVien					
<i>iDSV</i>	<i>tenSV</i>	<i>tuoi</i>	<i>chieuCao</i>	<i>canNang</i>	<i>sucKhoe()</i>
Id1	An	khoảng 18	[1.726, 1.816]	[60, 64]	
Id2	Bình	hơn trẻ	[1.62, 1.68]	[58, 62]	
Id3	Hà	[15, 24]	[1.63, 1.66]	[56.25, 62.5]	
Id4	Hương	[26, 30]	[1.44, 1.84]	[62, 66]	
Id5	Nhân	[20, 25]	[1.6, 1.65]		
Id6	Thủy	trẻ	[1.726, 1.816]	[70, 71]	

Bước 2: Áp dụng Định nghĩa 3.1 và Định nghĩa 3.2, dễ dàng nhận thấy rằng phụ thuộc $chieuCao \xrightarrow{fad} canNang$ thỏa mãn định nghĩa FAD.

Bước 3: Giá trị gần đúng cho thuộc tính *canNang* của đối tượng Nhân = $[56.25, 62.5] \in S(W)$.

b. Đối với truy vấn phương thức, ta có lời giải thuật toán như sau

Thuật toán ASMO (Approximate Semantic for Method Objects)

Vào: Một lớp C với m thuộc tính và p phương thức; $O = o_1, o_2, \dots, o_n \in C$.

Trong đó: $(o_1, o_2, \dots, o_{n-1}).[a_1, a_2, \dots, a_m]$ và $o_n.[a_1, a_2, \dots, a_{m-1}]$ là được định nghĩa (có nghĩa là giá trị dữ liệu tồn tại) và $o_n.a_m$ là chưa được định nghĩa (có nghĩa là giá trị dữ liệu bị thiếu).

Ra: Tìm ra giá trị tương tự gần đúng cho phương thức.

Phương pháp:

Bước 1: Áp dụng phương pháp ở Mục 3.1 để chuyển các giá trị thuộc tính tham gia truy vấn về các giá trị khoảng tương ứng.

Bước 2: Sử dụng phụ thuộc thuộc tính mờ ở Mục 3.3 để kiểm tra các FAD.

Bước 3: Xác định giá trị cho thuộc tính chưa định nghĩa $o_n.a_m$.

Bước 4: Sử dụng phụ thuộc phương thức mờ ở Mục 3.4 để kiểm tra các FMD.

Bước 5: Dựa vào FMD chọn hàm kết nhập đại số gia tử cho các thuộc tính.

Bước 6: Xác định giá trị gần đúng cho phương thức.

Ví dụ 5: Xét câu truy vấn ở trên “*Cho biết sức khỏe của Nhân*”, với giá trị thuộc tính *canNang* của *Nhân* là chưa định nghĩa. Áp dụng qui trình trên để tìm ra câu trả lời gần đúng trong CSDL ở Ví dụ 1 như sau:

Từ bước 1 đến bước 3 tính toán để tìm ra giá trị thuộc tính *canNang* cho đối tượng *Nhân* như trong Ví dụ 4. Kết quả ở Bảng 4(a).

Bước 4: Hàm kết nhập ĐSGT trong trường hợp này là trung bình cộng các khoảng của hai ĐSGT là *chieuCao* và *canNang*. Kết quả cho ở Bảng 4(b).

Bảng 4. Các kết quả thực hiện Ví dụ 5

(a) Kết quả thực hiện bước 1 – 3				(b) Kết quả thực hiện bước 4			
SinhVien				SinhVien			
iDSV	chieuCao	canNang	sucKhoe()	iDSV	chieuCao	canNang	sucKhoe()
id1	[1.726, 1.816]	[60, 64]		id1	[1.726, 1.816]	[60, 64]	[30.863, 32.908]
id2	[1.62, 1.68]	[58, 62]		id2	[1.62, 1.68]	[58, 62]	[29.81, 31.84]
id3	[1.63, 1.66]	[56.25, 62.5]		id3	[1.63, 1.66]	[56.25, 62.5]	[28.94, 32.08]
id4	[1.44, 1.84]	[62, 66]		id4	[1.44, 1.84]	[62, 66]	[31.72, 33.92]
id5	[1.6, 1.65]	[56.25, 62.5]		id5	[1.6, 1.65]	[56.25, 62.5]	[28.925, 32.075]
id6	[1.726, 1.816]	[70, 71]		id6	[1.726, 1.816]	[70, 71]	[35.863, 36.408]

Định lý 4.1. *Thuật toán ASAO và ASMO luôn dừng và đúng đắn.*

Chứng minh:

1. *Chứng minh tính dừng:* Tập các thuộc tính, phương thức của đối tượng là hữu hạn (n, p, m là hữu hạn) nên thuật toán sẽ dừng sau khi duyệt xong tất cả các đối tượng.
2. *Chứng minh tính đúng đắn:* Thuật toán **ASAO** và **ASMO** sử dụng Định nghĩa 3.2 phụ thuộc thuộc tính mờ và Định nghĩa 3.3 phụ thuộc phương thức mờ. Khái niệm phụ thuộc thuộc tính mờ và phụ thuộc phương thức mờ là sự mở rộng của khái niệm phụ thuộc hàm mờ trong CSDL quan hệ mờ. Sự suy dẫn logic này dựa vào sự suy dẫn của phụ thuộc mờ trong CSDL quan hệ mờ. Tính đúng đắn của phép suy dẫn bởi phụ thuộc hàm mờ trong CSDL quan hệ mờ đã được chứng minh. ■

Độ phức tạp thuật toán: Thuật toán **ASAO** và **ASMO** có độ phức tạp tính toán là $O(n^*m)$, với n là số đối tượng và m tập thuộc tính.

5. KẾT LUẬN

Các thuộc tính trong bất kỳ CSDL nào cũng đều có mối liên kết, ràng buộc với nhau. Riêng trong các CSDL HDT (mờ) còn có mối quan hệ giữa các thuộc tính và phương thức, trên cơ sở đó bài báo đã định nghĩa hai kiểu phụ thuộc mờ đó là FAD và FMD và các luật suy diễn trên các phụ thuộc đó. Khái niệm phụ thuộc của thuộc tính mờ và phụ thuộc phương thức mờ là sự mở rộng của khái niệm phụ thuộc hàm mờ trong CSDL quan hệ mờ. Trên cơ

sở đó, bài báo đã áp dụng các phụ thuộc mờ này để tìm ra câu trả lời gần đúng cho truy vấn Null. Dựa trên các phụ thuộc mờ, các dạng chuẩn lớp mờ và phương pháp tách lớp về các dạng chuẩn trong lớp đối tượng mờ sẽ được trình bày trong các nghiên cứu tiếp theo.

TÀI LIỆU THAM KHẢO

- [1]. V. Biazzo, R. Giugno, T. Lukasiewicz, V.S.Subrahmanian, Temporal probabilistic object bases, *IEEE Transaction on Knowledge and Engineering*, 15(4) (2003), (921-939).
- [2]. Doan Van Ban, Truong Cong Tuan, Doan Van Thang Querying data with fuzzy information in object-oriented databases based on hedge algebraic semantics, *Proc. of the 4th Inter. Conf. on Knowledge and Systems Engineering (KSE2012)*, Da Nang, Viet Nam, IEEE Computer Society Press, (105-112).
- [3]. D.V.Thang, D.V.Ban, Query data with fuzzy information in object-oriented databases an approach interval values, *International Journal of Computer Science and Information Security (IJCSIS)* **9** (2) (2011) 1-6.
- [4]. Đ.V.Ban, H.C.Hà, V.Đ.Quảng, Phương pháp phát hiện các lớp bộ phận trong lớp đối tượng mờ, *Tạp chí Tin học và Điều khiển học* **26** (4) (2010) 321–331.
- [5]. N.C.Ho, Quantifying Hedge Algebras and interpolation methods in approximate reasoning, *Proc. of the 5th Inter. Conf. on Fuzzy Information Processing*, Beijing, March 1-4, 2003 (105-112).
- [6]. N.C.Ho, *Lý thuyết tập mờ và công nghệ tính toán mềm, Hệ mờ, mạng nơron và ứng dụng*, NXB Khoa học và Kỹ thuật, 2001 (37–74).
- [7]. Nguyễn Cát Hồ, Nguyễn Công Hào, Phụ thuộc đơn điệu trong cơ sở dữ liệu mờ tiếp cận ngữ nghĩa lân cận của đại số gia tử, *Tạp chí Tin học và Điều khiển học* **24** (1) (2008) 20–31.
- [8]. Nguyễn Công Hào, Một số dạng chuẩn trong cơ sở dữ liệu mờ theo cách tiếp cận đại số gia tử, *Tạp chí Tin học và Điều khiển học* **17** (2007) 101–107.
- [9]. Phương Minh Nam, Trần Thái Sơn, Về một cơ sở dữ liệu mờ và ứng dụng trong quản lý tội phạm hình sự, *Tạp chí Tin học và Điều khiển học* **22** (1) (2006) 67–73.
- [10]. S.Y.liao, H.Q.Wang and W.Y.Liu, Functional dependencies with Null values, fuzzy values, and crisp values, *IEEE Transactions on Fuzzy Systems* **17** (1) 97–103.
- [11]. S. L. Wang, and T. J. Huang, Analogical reasoning to answer null queries in fuzzy object-oriented data model, *Proc. of the 6th IEEE Inter. Conf. on Fuzzy Systems*, Barcelona, Spain, Vol.1, 1997, pp 31-36.

Ngày nhận bài 25 - 4 - 2012

Ngày lại sửa ngày 28 - 02 - 2013