

MỘT CÁCH TIẾP CẬN XỬ LÝ RÀNG BUỘC TRONG MÔ HÌNH CƠ SỞ DỮ LIỆU HƯỚNG ĐỐI TƯỢNG MỜ

Nguyễn Công Hào

Ban Đào tạo và Công tác sinh viên, Đại học Huế

Email: nchao@hueuni.edu.vn

Ngày nhận bài: 01/12/2022; ngày hoàn thành phần biện: 21/12/2022; ngày duyệt đăng: 20/12/2022

TÓM TẮT

Trong bài báo này, chúng tôi đề xuất một cách tiếp cận xử lý ràng buộc như ràng buộc miền giá trị và ràng buộc giữa các thuộc tính trên các lớp của mô hình cơ sở dữ liệu hướng đối tượng mờ dựa trên đại số gia tử. Theo cách tiếp cận này, việc kiểm chứng đối tượng thỏa mãn ràng buộc theo từng mức k ($k \in \mathbb{Z}^+$), với k là mức độ tương tự được xác định ứng với mỗi thuộc tính mờ phù hợp với các ràng buộc dữ liệu trong mô hình, đồng thời dễ dàng đối sánh dữ liệu trực tiếp trên các giá trị ngôn ngữ so với các cách tiếp cận khác.

Từ khóa: Đại số gia tử, mô hình cơ sở dữ liệu hướng đối tượng mờ, ràng buộc dữ liệu.

1. MỞ ĐẦU

Đối với một cơ sở dữ liệu hướng đối tượng (CSDL HĐT) mờ, ràng buộc là thành phần rất quan trọng dùng để quy định ngữ nghĩa dữ liệu đối với CSDL và nó được xem như các yêu cầu cần được thỏa mãn giữa các thuộc tính bên trong một lớp hoặc các thuộc tính giữa các lớp với nhau [5-6]. Chẳng hạn, khi lưu trữ thông tin về giảng viên trong một trường đại học, một số ràng buộc có thể được sử dụng trên miền giá trị của các thuộc tính của giảng viên như số căn cước công dân, mã giảng viên, tuổi, năng lực giảng dạy và nghiên cứu khoa học.... trong đó giá trị số căn cước công dân, mã giảng viên không được trùng nhau, tuổi của giảng viên là *tương đối trẻ*, năng lực giảng dạy và nghiên cứu khoa học của giảng viên là *rất tốt*. Các ràng buộc tồn tại là một thực tế nên vấn đề là cần đưa chúng vào các mô hình dữ liệu để làm cho dữ liệu có ý nghĩa và phù hợp với ngữ cảnh hơn [1]. Các ràng buộc cũng được sử dụng để đưa ra các tiêu chuẩn cho việc cập nhật, xử lý dữ liệu nhằm đảm bảo tính nhất quán, đúng đắn và đầy đủ.

Một cách tổng quát, một CSDL HĐT mong muốn đối với người dùng khi nó thỏa mãn tất cả các ràng buộc có thể có trên các lớp của CSDL đó, tạo ra một sự linh hoạt và đầy đủ về mặt ngữ nghĩa trên phương diện lưu trữ và thao tác dữ liệu. Ngược lại, sẽ có những hạn chế trên các dữ liệu nếu một CSDL không thỏa mãn hoặc không đưa ra được các ràng buộc. Cấu trúc bài báo được trình bày trong 5 phần, ngoài phần mở đầu, phần 2 trình bày một số kiến thức cơ sở về đại số gia tử (ĐSGT) và CSDL HĐT dựa trên ĐSGT, phần 3 và 4 trình bày một số ràng buộc dữ liệu như ràng buộc miền giá trị và ràng buộc giữa các thuộc tính, phần 5 trình bày kết luận và hướng nghiên cứu tiếp theo của bài báo.

2. MỘT SỐ KIẾN THỨC CƠ SỞ

2.1. Giới thiệu đại số gia tử

ĐSGT là một trong những cách tiếp cận để phát hiện cấu trúc đại số của miền giá trị của các biến ngôn ngữ [4]. Theo quan điểm đại số, mỗi miền giá trị của biến ngôn ngữ X có thể được hiểu như là một đại số $AX = (X, G, H, \leq)$, trong đó $Dom(X)=X$ là miền các giá trị ngôn ngữ của biến ngôn ngữ X được sinh ra từ tập các phần tử sinh $G = \{c, c^+\}$ bởi sự tác động của các gia tử trong $H=H \cup H^+$, W là phần tử trung hòa, \leq là một mối quan hệ thứ tự có ngữ nghĩa trên X , nó được cảm sinh từ ý nghĩa định tính tự nhiên của từ ngữ. Cấu trúc thứ tự cảm sinh trực tiếp như vậy chính là điểm khác biệt so với các cách tiếp cận khác. Khi thêm một số phần tử đặc biệt, thì ĐSGT trở thành một đại số trừu tượng $\underline{X}=(X, G, H, \Sigma, \Phi, \leq)$, trong đó Σ, Φ là các toán tử lấy giới hạn của tập các phần tử được sinh ra khi chịu tác động liên tiếp của các gia tử trong H . Một cách khác, nếu ký hiệu $H(x)=\{h_1...h_px/h_1, \dots h_p \in H\}$ thì $\Phi_x=\text{infimum}H(x)$ và $\Sigma_x=\text{supremum}H(x)$. Các hàm định lượng ngữ nghĩa (ν), hàm độ đo tính mờ (f_m), hàm dấu (Sgn) và các tính chất của ĐSGT có thể tham khảo trong các công trình liên quan [1,4].

2.2. Độ tương tự mức k

Khi định nghĩa lân cận mức k chúng ta mong muốn các giá trị đại diện như vậy phải là điểm trong của lân cận mức k . Vì vậy ta định nghĩa độ tương tự mức k như sau: Chúng ta luôn luôn giả thiết rằng mỗi tập H^- và H^+ chứa ít nhất 2 gia tử. Xét X_k là tập tất cả các phần tử độ dài k . Dựa trên các khoảng mờ mức k và các khoảng mờ mức $k+1$ chúng ta mô tả không hình thức việc xây dựng một phân hoạch của miền $[0,1]$ như sau: Với $k=1$, các khoảng mờ mức 1 gồm $I(c^-)$ và $I(c^+)$. Các khoảng mờ mức 2 trên khoảng $I(c^-)$ là $I(h_{p^-}c^-) = I(h_{p-1}c^-) = \dots = I(h_2c^-) = I(h_1c^-) = \nu_A(c^-)=I(h_{-1}c^-) = I(h_{-2}c^-) = \dots = I(h_{-q+1}c^-)=I(h_{-q}c^-)$. Khi đó, ta xây dựng phân hoạch về độ tương tự mức 1 gồm các lớp tương đương sau: $S(0)=I(h_{p^-}c^-), S(c^-)=I(c^-) \setminus [I(h_{-q}c^-) \cup I(h_{p^-}c^-)]$; $S(W) = I(h_{-q}c^-) \cup I(h_{-q}c^+)$; tương tự ta có $S(c^+) = I(c^+) \setminus [I(h_{-q}c^+) \cup I(h_{p^+}c^+)]$ và $S(1) = I(h_{p^+}c^+)$. Tương tự, với $k=2$, ta có thể xây dựng phân

hoạch các lớp tương tự mức 2. chẳng hạn trên khoảng mờ mức 2, $I(h_i c^+) = (\cup_A(\Phi h_i c^+), \cup_A(\Sigma h_i c^+))$ với hai khoảng mờ kề là $I(h_{i-1} c^+)$ và $I(h_{i+1} c^+)$ chúng ta sẽ có các lớp tương đương dạng sau: $S(h_i c^+) = I(h_i c^+) \setminus [I(h_p h_i c^+) \cup I(h_q h_i c^+)]$, $S(\Phi h_i c^+) = I(h_q h_{i-1} c^+) \cup I(h_q h_i c^+)$ và $S(\Sigma h_i c^+) = I(h_p h_{i-1} c^+) \cup I(h_p h_i c^+)$, với I sao cho $-q \leq i \leq p$ và $i \neq 0$. Bằng cách tương tự như vậy ta có thể xây dựng các phân hoạch các lớp tương tự mức k bất kỳ. Tuy nhiên, trong thực tế ứng dụng chúng ta có thể giới hạn các gia tử tác động liên tiếp lên các phần tử nguyên thủy c và c^+ là một số nguyên k^* nào đó. Các giá trị kinh điển 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 .

Ví dụ 2.1. Ta xét lược đồ quan hệ $U = \{ \text{MAGV}, \text{TENGV}, \text{NLV}, \text{LUONG} \}$ với ý nghĩa: Mã số giảng viên (MAGV), Tên giảng viên (TENGV) là 2 thuộc tính rõ, Số ngày làm việc trong tháng (NLV), Lương (LUONG) là 2 thuộc tính mờ. Trong đó $\text{DNLV} = [0; 30]$ và $\text{DLUONG} = [0; 10.000.000]$. LDNLV và LDLUONG có cùng tập các râu giống nhau với tập các phần tử sinh là $\{0, \text{thấp}, W, \text{cao}, 1\}$ và tập các gia tử là $\{\text{ít}, \text{khả năng}, \text{hơn}, \text{rất}\}$. Mặc dù các thuộc tính ngôn ngữ đang xét có cùng tập các râu, nhưng ngữ nghĩa định lượng của chúng khác nhau.

Đối với thuộc tính SONLV: $fm(\text{cao}) = 0,35$, $fm(\text{thấp}) = 0,65$, $\mu(\text{khả năng}) = 0,25$, $\mu(\text{ít}) = 0,20$, $\mu(\text{hơn}) = 0,15$ và $\mu(\text{rất}) = 0,40$. Ta phân hoạch đoạn $[0, 30]$ thành 5 khoảng tương tự mức 1 là: $fm(\text{rất cao}) \times 30 = 0,35 \times 0,35 \times 30 = 3,675$. Vậy $S(1) \times 30 = (26,325; 30]$;

$(fm(\text{khả năng cao}) + fm(\text{hơn cao})) \times 30 = (0,25 \times 0,35 + 0,15 \times 0,35) \times 30 = 4,2$ và $S(\text{cao}) \times 30 = (22,125; 26,325]$; $(fm(\text{ít thấp}) + fm(\text{ít cao})) \times 30 = (0,25 \times 0,65 + 0,25 \times 0,35) \times 30 = 7,5$ và $S(W) \times 30 = (14,625; 22,125]$; $(fm(\text{khả năng thấp}) + fm(\text{hơn thấp})) \times 30 = (0,25 \times 0,65 + 0,15 \times 0,65) \times 30 = 7,8$ và $S(\text{thấp}) \times 30 = (6,825; 14,625]$; $S(0) \times 30 = [0; 6,825]$.

Đối với thuộc tính LUONG: $fm(\text{cao}) = 0,6$, $fm(\text{thấp}) = 0,4$, $\mu(\text{khả năng}) = 0,15$, $\mu(\text{ít}) = 0,25$, $\mu(\text{hơn}) = 0,25$ và $\mu(\text{rất}) = 0,35$. Ta phân hoạch đoạn $[0; 10.000.000]$ thành 5 khoảng tương tự mức 1 là: $fm(\text{rất cao}) \times 10.000.000 = 0,35 \times 0,6 \times 10.000.000 = 2.100.000$. Vậy $S(1) \times 10.000.000 = (7.900.000; 10.000.000]$; $(fm(\text{khả năng cao}) + fm(\text{hơn cao})) \times 10.000.000 = (0,25 \times 0,6 + 0,15 \times 0,6) \times 10.000.000 = 2.400.000$ và $S(\text{cao}) \times 10.000.000 = (5.500.000; 7.900.000]$; $(fm(\text{ít thấp}) + fm(\text{ít cao})) \times 10.000.000 = (0,25 \times 0,6 + 0,25 \times 0,4) \times 10.000.000 = 2.500.000$ và $S(W) \times 10.000.000 = (3.000.000; 5.500.000]$; $(fm(\text{khả năng thấp}) + fm(\text{hơn thấp})) \times 10.000.000 = (0,25 \times 0,4 + 0,15 \times 0,4) \times 10.000.000 = 1.600.000$ và $S(\text{thấp}) \times 10.000.000 = (1.400.000; 3.000.000]$, $S(0) \times 10.000.000 = [0; 1.400.000]$.

2.3. Cơ sở dữ liệu hướng đối tượng mờ

Trong CSDL HĐT mờ [5], bốn trường hợp sau đây có thể được dùng để phân biệt cho các mối quan hệ lớp đối tượng:

a) Lớp rõ và đối tượng rõ: trường hợp này giống như trong CSDL HDT, nghĩa là đối tượng thuộc hay không thuộc lớp một cách chắc chắn.

b) Lớp rõ và đối tượng mờ: lớp được xác định chính xác và có ranh giới chính xác, còn đối tượng là mờ vì giá trị thuộc tính của nó có thể mờ. Trong trường hợp này, đối tượng có thể là thành viên của lớp với độ thuộc nào đó.

c) Lớp mờ và đối tượng rõ: giống như trường hợp ở 2.2.2, các đối tượng có thể thuộc về lớp với độ thuộc mức k . Ví dụ một đối tượng học viên cao học và một lớp sinh viên trẻ.

d) Lớp mờ và đối tượng mờ: trong trường hợp này, đối tượng cũng thuộc về lớp với độ thuộc mức k .

Các mối quan hệ lớp đối tượng trong b), c) và d) trên đây được gọi là quan hệ lớp đối tượng mờ. Trong thực tế, trường hợp a) có thể được xem như là trường hợp đặc biệt của mối quan hệ lớp đối tượng mờ, với độ thuộc của đối tượng vào lớp là 1. Với mỗi giá trị ngôn ngữ mờ x , chúng ta sẽ định nghĩa một biểu diễn khoảng cho x . Trong thực tế, số gia tử trong các giá trị ngôn ngữ là hữu hạn nên tồn tại một số nguyên dương k^* sao cho $0 < |x| \leq k^*$, $\forall x \in X$. Với bất kỳ $x \in X$, đặt $j = |x|$, với mỗi số nguyên k cho trước với $1 \leq k \leq k^*$, lân cận tối thiểu k của x ký hiệu là $O_{\min,k}(x)$ được định nghĩa như sau: nếu $k=j$ thì $O_{\min,k}(x) = I_{k+1}(h_{-1}x) \cup I_{k+1}(h_1x)$, nếu $1 \leq k < j$ thì $O_{\min,k}(x) = I_j(x)$ và nếu $j+1 \leq k \leq k^*$ thì $O_{\min,k}(x) = I_{k+1}(h_{-1}y) \cup I_{k+1}(h_1y)$. Từ đó, ta biểu diễn dữ liệu ngôn ngữ mờ theo định nghĩa sau:

Định nghĩa 2.1. Cho $x \in X \cup C$, một biểu diễn khoảng của x là một tập $IRp(x)$ các khoảng được xác định: $IRp(x) = \{O_{\min,k}(x) \mid 1 \leq k \leq k^*\}$

Xét \underline{X} là ĐSGT tuyến tính đầy đủ, với $H^+ = \{h_1, \dots, h_p\}$ và $H^- = \{h_{-1}, \dots, h_{-q}\}$, trong đó $p, q > 1$. Đặt H_1 là tập các gia tử yếu, H_2 là tập các gia tử mạnh theo nghĩa khi tác động nó sẽ làm thay đổi nghĩa mạnh hơn số gia tử trong H_1 , tức là các tập H_1 và H_2 gồm: $H_1 = \{h_i, h_j \mid 1 \leq i \leq [p/2], 1 \leq j \leq [q/2]\}$, $H_2 = \{h_i, h_j \mid [p/2] \leq i \leq p, [q/2] \leq j \leq q\}$. Đặt $P_{k+1}(H_n) = \{I(h_i y) \mid y \in X_k, h_i \in H_n\}$, với $n=1,2$. Hai khoảng $I(x)$ và $I(y)$ trong $P_{k+1}(H_n)$ được gọi là liên thông với nhau nếu tồn tại các khoảng thuộc $P_{k+1}(H_n)$ liên tiếp nhau xếp từ $I(x)$ đến $I(y)$. Quan hệ này sẽ phân $P_{k+1}(H_n)$ thành các thành phần liên thông. Ta lại có, với mỗi $y \in X_k$, $P_{k+1}(H_1)$ được phân thành các cụm có dạng $\{I(h_i y) \mid h_i \in H_1\}$. Hơn nữa, do $I(h_{-1}y) \leq \nu(y) \leq I(h_1 y)$ hoặc là $I(h_1 y) \leq \nu(y) \leq I(h_{-1}y)$ nên bao giờ ta cũng có $\nu(y) \in \{I(h_i y) \mid h_i \in H_1\}$. Bây giờ ta phân cụm các khoảng mờ của $P_{k+1}(H_2)$. Giả sử $X_k = \{x_s \mid s = 0, \dots, m-1\}$ gồm m phần tử được sắp thành một dãy sao cho $x_i \leq x_j$ khi và chỉ khi $i \leq j$. Ký hiệu $H_2^- = H_2 \cap H^-$ và $H_2^+ = H_2 \cap H^+$. Để ý rằng $h_{-q} \in H_2^-$ và $h_p \in H_2^+$. Các cụm được sinh ra từ các khoảng mờ thuộc $P_{k+1}(H_2)$ có ba loại là cụm nằm bên trái x_0 : $\{I(h_i x_0) \mid h_i \in H_2^+\}$; cụm nằm bên phải x_{m-1} : $\{I(h_i x_{m-1}) \mid h_i \in H_2^+\}$; cụm nằm giữa x_s và x_{s+1} với $s = 0, \dots, m-2$: phụ thuộc vào $Sgn(h_p x_s)$ và $Sgn(h_p x_{s+1})$ như sau: $C = \{I(h_i x_s), I(h_j x_{s+1}) \mid h_i \in H_2^+, h_j \in H_2^-\}$, nếu

$Sgn(h_p x_s) = +1$ và $Sgn(h_p x_{s+1}) = +1$; $C = \{ I(h_i x_s), I(h_j' x_{s+1}) \mid h_i \in H_2^+, h_j' \in H_2^+ \}$, nếu $Sgn(h_p x_s) = +1$ và $Sgn(h_p x_{s+1}) = -1$; $C = \{ I(h_i x_s), I(h_j' x_{s+1}) \mid h_i \in H_2^-, h_j' \in H_2^- \}$, nếu $Sgn(h_p x_s) = -1$ và $Sgn(h_p x_{s+1}) = +1$; $C = \{ I(h_i x_s), I(h_j' x_{s+1}) \mid h_i \in H_2^-, h_j' \in H_2^+ \}$, nếu $Sgn(h_p x_s) = -1$ và $Sgn(h_p x_{s+1}) = -1$.

Tập tất cả các cụm được ký hiệu là C . Vì $\{ S_k(C) \mid C \in C \}$ là một phân hoạch trên miền trị tham chiếu nên nó xác định một quan hệ tương đương và chúng ta sẽ gọi là quan hệ tương tự mức k . Do tính chất của phân hoạch nên với mỗi giá trị x của thuộc tính, tồn tại duy nhất một cụm C sao cho $u(x) \in S_k(C)$ và ta định nghĩa khoảng tương tự mức k như sau:

Định nghĩa 2.2. Với mỗi C thuộc C , ta gọi khoảng tương tự mức k ứng với C là: $S_k(C) = \cup \{ I(u) \mid I(u) \in C \}$, khi đó $S_k(x) = S_k(C)$.

Định nghĩa 2.3. Cho một đối tượng bất kỳ o trên tập thuộc tính $\{ A_1, A_2, \dots, A_n \}$ của lớp C , \underline{X} là một ĐSGT tuyến tính đầy đủ, với mỗi k , $1 \leq k \leq k^*$, S_k là quan hệ tương tự mức k trên miền trị thuộc tính A_i của lớp C . Khi đó, với mọi $u \in \underline{X}$, giá trị $o(A_i)$ và u được gọi là bằng nhau mức k , ký hiệu $o(A_i) =_k u$, khi và chỉ khi $O_{\min, k}(o(A_i)) \in S_k(u)$.

Định nghĩa 2.4. Cho hai đối tượng bất kỳ o_1, o_2 trên tập thuộc tính $\{ A_1, A_2, \dots, A_n \}$ của lớp C , \underline{X} là một ĐSGT tuyến tính đầy đủ, với mỗi k , $1 \leq k \leq k^*$, S_k là quan hệ tương tự mức k trên miền trị thuộc tính A_i của lớp C . Khi đó:

1) Hai giá trị $o_1(A_i)$ và $o_2(A_i)$ được gọi là bằng nhau mức k , ký hiệu $o_1(A_i) =_k o_2(A_i)$, khi và chỉ khi tồn tại một lớp tương đương $S_k(u)$ của quan hệ tương tự S_k sao cho $O_{\min, k}(o_1(A_i)) \in S_k(u)$ và $O_{\min, k}(o_2(A_i)) \in S_k(u)$.

2) Hai giá trị $o_1(A_i)$ và $o_2(A_i)$ được gọi là khác nhau mức k , ký hiệu $o_1(A_i) \neq_k o_2(A_i)$, nếu không tồn tại một lớp tương đương $S_k(u)$ của quan hệ tương tự S_k sao cho $O_{\min, k}(o_1(A_i)) \in S_k(u)$ và $O_{\min, k}(o_2(A_i)) \in S_k(u)$.

Định nghĩa 2.5: Gọi $C = \{ c_1, c_2, \dots, c_n \}$ và $O = \{ o_1, o_2, \dots, o_m \}$ lần lượt là tập các ràng buộc và tập các đối tượng trên lớp C . Đối với mỗi ràng buộc $c_i \in C$ ràng buộc cho một hay nhiều đối tượng $o_j \in O$ được định nghĩa đầy đủ thông qua cú pháp và ngữ nghĩa của nó như sau:

Cú pháp của một ràng buộc:

$C^{<name>}$
 $C^{<properties>} [\varepsilon] \text{ in } <class\ name> [\text{reference to } <class\ name^*>]$

[satisfied with level of $<k>$]

Trong đó: $<name>$: tên của ràng buộc; $<properties>$: Tập thuộc tính của lớp tham gia vào ràng buộc; $[\varepsilon]$: với ε hoặc là giá trị rỗng; hoặc là tập các thuộc tính; hoặc là một giá trị số, khoảng hay giá trị mờ tùy thuộc vào mỗi ràng buộc đưa ra; $<class\ name>$: tên lớp tham gia vào ràng buộc; $[\text{reference to } <class\ name^*>]$: tên lớp có tham chiếu ràng buộc nếu có; $[\text{satisfied with level of } <k>]$: Mức k thỏa mãn ràng buộc nếu có.

3. RÀNG BUỘC MIỀN GIÁ TRỊ THUỘC TÍNH

Theo hướng tiếp cận ĐSGT, khi xét về mối quan hệ giữa lớp/đối tượng hoặc các đối tượng thỏa mãn các điều kiện truy vấn mờ thì đối tượng đó thuộc về một lớp hay thỏa mãn các biểu thức truy vấn mờ với một mức k xác định nào đó. Với ý tưởng này, trên quan điểm ràng buộc dữ liệu ta cũng xét các đối tượng thỏa mãn các ràng buộc với một mức k do tính mờ thể hiện trong các ràng buộc được đưa ra. Chẳng hạn, như một ràng buộc đối với đối tượng giảng viên có thuộc tính lương là *tương đối cao* và số công nghiên cứu khoa học đã công bố *khá nhiều* nên việc thỏa mãn ràng buộc này chỉ nên được đánh giá bởi một mức nào đó tùy thuộc vào khoảng mờ mức k . Để đánh giá được mức k như thế buộc chúng ta phải biểu diễn các giá trị trên các thuộc tính của đối tượng có tham gia vào ràng buộc dưới dạng các khoảng mờ theo từng mức. Sau đó việc xác định các đối tượng thỏa mãn ràng buộc được thực hiện bởi việc đối sánh các giá trị thuộc tính của đối tượng tham gia vào ràng buộc theo từng mức k với cách đối sánh thông thường, đây chính là một ưu điểm nổi bật của cách tiếp cận ĐSGT.

Đối với các thuộc tính mờ thì các kiểu dữ liệu cũng dựa trên các kiểu dữ liệu cơ sở như trên cho phép biểu diễn các thông tin mờ. Một số ràng buộc được đề xuất cụ thể như sau:

Ràng buộc “not_null”: Đối với ràng buộc này, thuộc tính trên đối tượng quy định luôn phải có một giá trị xác định và hợp lệ. Biểu diễn cho ràng buộc này ta sử dụng cú pháp:

$$C_{\{id\}}^{not_null} \text{ in } \langle \text{class name} \rangle \quad (3.1)$$

Trong đó, id mô tả thuộc tính được áp dụng ràng buộc “not_null” trên lớp có tên là $\langle \text{class name} \rangle$.

Ràng buộc “null”: Đối với ràng buộc này, thuộc tính trên đối tượng có thể chứa giá trị null ngược lại đối với trường hợp “not_null” ở trên. Biểu diễn cho ràng buộc này ta sử dụng cú pháp:

$$C_{\{id\}}^{null} \text{ in } \langle \text{classname} \rangle \quad (3.2)$$

Ràng buộc “value” (ràng buộc giá trị): Để biểu diễn cho ràng buộc này ta sử dụng cú pháp:

$$C_{\{id\}}^{value} [\varepsilon] \text{ in } \langle \text{class name} \rangle \text{ Satisfied with level of } \langle k \rangle \quad (3.3)$$

Trong đó, ε là một ràng buộc giá trị có thể là một giá trị rõ, giá trị khoảng và giá trị ngôn ngữ mờ để giới hạn miền giá trị của thuộc tính được chỉ ra là id và k là mức cho trước chỉ ra mức thỏa mãn ràng buộc của đối tượng trên lớp có tên là $\langle \text{class name} \rangle$.

Ví dụ 3.1. Chẳng hạn, ta ràng buộc thuộc tính Lương của lớp Giảng viên với các giá trị như khoảng $7.000.000$ hoặc $10.000.000 \leq \text{Lương} \leq 15.000.000$ hoặc “ tương đối cao” thì ta có các biểu diễn ràng buộc tương ứng như sau:

$C_{\{\text{Lương}\}}^{\text{value}}$ [khoảng 7.000.000] in Giảng viên Satisfied with level of $\langle k \rangle$

$C_{\{\text{Lương}\}}^{\text{value}}$ [$10.000.000 \leq \text{Lương} \leq 15.000.000$] in Giảng viên Satisfied with level of $\langle k \rangle$

$C_{\{\text{Lương}\}}^{\text{value}}$ [tương đối cao] in Nhân viên Satisfied with level of $\langle k \rangle$.

Nếu quan hệ bằng nhau mức $k \in \mathbb{Z}^+$ xảy ra tức là tồn tại lớp tương đương $S_k(x) \in S_k$, với S_k là một quan hệ tương đương mức k trên thuộc tính id sao cho lân cận tối thiểu biểu diễn giá trị thuộc tính đang xét và ràng buộc giá trị ε cùng thuộc vào lớp tương đương $S_k(x)$ đó.

Thuật toán 3.1. Kiểm tra một đối tượng o có thỏa mãn ràng buộc giá trị c với mức k cho trước hay không.

Đầu vào: Đối tượng o và thuộc tính id cần xét. Ràng buộc c được biểu diễn từ (3.3), $k \in \mathbb{Z}^+$ cho trước để chỉ ra mức thỏa mãn ràng buộc của đối tượng o , S_k là một quan hệ tương đương trên miền trị thuộc tính id.

Đầu ra: Đối tượng o có thỏa mãn ràng buộc giá trị hoặc không thỏa mãn.

Phương pháp:

1. Begin

2. Xây dựng ĐSGT AX_{id} cho thuộc tính id bao gồm xây dựng các tham số mờ, các gia tử và các phần tử sinh.

3. Xác định miền xác định thuộc tính id là $D(id) = (\min_{id}, \max_{id})$ với \min_{id} , \max_{id} là giá trị nhỏ nhất và giá trị lớn nhất của id.

4. Tính $O_{\min,k}(o(id))$ và $O_{\min,k}(\varepsilon)$

5. if $(O_{\min,k}(o(id)) =_k O_{\min,k}(\varepsilon))$ then

6. return “đối tượng o thỏa mãn ràng buộc c với mức k ”

7. else

8. return “ đối tượng o không thỏa mãn ràng buộc c với mức k ”

9. End.

Thuật toán 3.1 đúng đắn, độ phức tạp thuật toán là $O(1)$.

4. RÀNG BUỘC GIỮA CÁC THUỘC TÍNH

Trong thực tế có những ràng buộc có dạng “Nếu hai giảng viên o1, o2 có năng lực giảng dạy khá tốt như nhau thì lương của o1, o2 là khá cao như nhau”, nhận thấy rằng tồn tại mối liên hệ năng lực giảng dạy và lương của o1, o2. Nếu chúng ta quan niệm rằng mối liên hệ giữa năng lực giảng dạy và lương của hai giảng viên o1, o2 là một mối ràng buộc trong một cơ sở dữ liệu hướng đối tượng mờ thì ta gọi ràng buộc này là một ràng buộc giữa các thuộc tính trên lớp C chứa hai thuộc tính này, hay nói cách khác đó là một dạng phụ thuộc dữ liệu mờ giữa hai thuộc tính.

Định nghĩa 4.1. Cho $A = \{A_1, A_2, \dots, A_m\}$ là tập các thuộc tính và C là lớp xác định trên tập thuộc tính đó. Xét $X, Y \subseteq A$. Ta nói rằng, đối tượng $o \in C$ thỏa mãn ràng buộc giữa thuộc tính X, Y kí hiệu $X \sim_k Y$ đọc là X xác định mờ Y với mức k: nếu $\forall o_i \in C, o_i(X) =_k o(X)$ thì $o_i(Y) =_k o(Y)$.

Cú pháp ràng buộc liên thuộc tính là :

$$C_{\{Y_1, Y_2, \dots, Y_m\}}^{inter_properties} [X_1, X_2, \dots, X_n] \text{ in } \langle \text{class name} \rangle \text{ Satisfied with level of } \langle k \rangle \quad (4.1)$$

Trong đó, $X = \{X_1, X_2, \dots, X_n\}$ và $Y = \{Y_1, Y_2, \dots, Y_m\}$ là tập các thuộc tính trên lớp có tên là $\langle \text{class name} \rangle$, yêu cầu mức thỏa mãn ràng buộc giữa X và Y.

Thuật toán 4.1. Thuật toán kiểm tra xem đối tượng $o \in C$ có thỏa mãn ràng buộc liên thuộc tính X, Y được đưa ra với mức $k \in Z^+$ cho trước hay không.

Đầu vào: $A = \{A_1, A_2, \dots, A_p\}$ là tập các thuộc tính trên lớp C, đối tượng $o \in C$ cần xét. Trên C chứa tập các đối tượng $\{o_1, o_2, \dots, o_q\}$ và $X, Y \subseteq A, X = \{X_1, X_2, \dots, X_n\}, Y = \{Y_1, Y_2, \dots, Y_m\}, m, n \leq p$; ràng buộc liên thuộc tính được chỉ ra ở (4.1), $k \in Z^+$.

Đầu ra: Đối tượng o có thỏa mãn ràng buộc liên thuộc tính $X \sim_k Y$ với mức $k \in Z^+$ hay không.

Phương pháp:

1. Begin
2. Xây dựng các ĐSGT cho các thuộc tính mờ trong X và Y, giả sử tập X, Y lần lượt có s, r thuộc tính mờ trong X và Y là $S = \{X_1, X_2, \dots, X_s\}, R = \{Y_1, Y_2, \dots, Y_r\}$.
3. Xác định miền giá trị các thuộc tính $X_i \in S$ và $Y_i \in R$ là $D(X_i) = (\min_{X_i}, \max_{X_i})$ và $D(Y_i) = (\min_{Y_i}, \max_{Y_i})$ với (\min_{X_i}, \max_{X_i}) và (\min_{Y_i}, \max_{Y_i}) lần lượt là giá trị nhỏ nhất và giá trị lớn nhất của X_i, Y_i .
4. $i := 1$; stop := false;
5. while ($i \leq q-1$) and (stop = false) do
6. begin


```
7.         if thoaX( $o_i(X) =_k o(X)$ ) then
8.             begin
9.                 if thoaY( $o_i(Y) =_k o(Y)$ ) then Stop:= false;
10.                else Stop:=true;
11.            end i: = i+1;
12.        end
13. if ( $i \neq q-1$ ) then return “ o không thỏa mãn ràng buộc”
14. else         if ( $i = q-1$ ) then return “ o thỏa mãn ràng buộc”
15. End.
```

// Xây dựng hàm thoaX, hàm thoaY

```
1. Begin
2. r: = 1; Stop:= false;
3. while ( $r \leq n$ ) and (stop = false) do
4.     begin
5.         if ( $X_r$  là thuộc tính rõ) then
6.             begin
7.                 if  $o_i(X_r) = o(X_r)$  then stop:= false
8.                 else  stop:= true;
9.             end
10.        else if ( $X_r$  là thuộc tính mờ) then
11.            begin
12.                Tìm  $O_{\min,k}(o_i(X_r))$  và  $O_{\min,k}(o(X_r))$ ;
13.                if ( $O_{\min,k}(o_i(X_r)) =_k O_{\min,k}(o(X_r))$ ) then stop:=false
14.                else stop:= true;
15.            end
16.            r: = r+1;
17.        end
18. if ( $r = n$ ) then return thoaX: = true
19. else return thoaX: = false;
```

```
20. End.
// Xây dựng hàm thoaY
1. Begin
2. z: = 1; Stop:= false;
3. while (z ≤ m) and (stop = false) do
4.     begin
5.         if (Yz là thuộc tính rõ) then
6.             begin
7.                 if oi(Yz) = o(Yz) then stop:= false
8.                 else stop:= true;
9.             end
10.        else if (Yz là thuộc tính mờ) then
11.            begin
12.                Tìm Omin,k(oi(Yz)) và Omin,k(o(Yz));
13.                if (Omin,k(oi(Yz)) =k Omin,k(o(Yz))) then stop:=false
14.                else stop:= true;
15.            end
16.            z: = z+1;
17.        end
18. if (z = m) then return thoaY: = true
19. else return thoaY: = false;
20. End.
```

Với đối tượng o cần xét thỏa mãn ràng buộc giữa các thuộc tính được đề xuất, thuật toán 4.1 cần duyệt qua $t-1$ đối tượng còn lại trên lớp đó. Thuật toán 4.1 là đúng đắn, đảm bảo tính dừng và dừng khi hàm thoa $X(o_i(X) =_k o(Y))$ thỏa và hàm thoa $Y(o_i(Y) =_k o(Y))$ không thỏa hoặc đã duyệt qua hết các đối tượng của lớp. Độ phức tạp thuật toán hàm thoa X và thoa Y lần lượt là $O(n)$ và $O(m)$. Vậy, Độ phức tạp thuật toán 4.1 là $O(t)$, với $n, m \leq t$.

5. KẾT LUẬN

Trong quá trình thiết kế CSDL HDT mờ, việc xác định được các dạng ràng buộc dữ liệu là vấn đề quan trọng. Chẳng hạn như dạng ràng buộc trên miền trị thuộc tính là nền tảng xác định các khóa mờ, ràng buộc giữa các thuộc tính là cơ sở để xác định các dạng phụ thuộc hàm mờ, một cơ sở quan trọng trong việc xây dựng các dạng chuẩn mờ nhằm hạn chế tối đa sự dư thừa dữ liệu. Ràng buộc trên miền trị thuộc tính nhằm đảm bảo việc xác định dữ liệu theo một quy tắc nào đó được quy định sẵn khi thiết kế CSDL và đảm bảo tính đúng đắn dữ liệu của một đối tượng trong lớp đối tượng mờ. Việc nghiên cứu và đề xuất các thuật toán liên quan đến các dạng ràng buộc dữ liệu này là có ý nghĩa về mặt lý thuyết cũng như thiết kế ứng dụng trong thực tiễn. Một số vấn đề liên quan đến dạng ràng buộc đặc biệt được chúng tôi nghiên cứu và trình bày trong những công trình tiếp theo.

TÀI LIỆU THAM KHẢO

- [1]. Nguyen Cong Hao, Le Thi My Le, "Data dependencies in fuzzy object oriented databases model based on hedge algebras ", *Annales Uni. Sci. Budapest. Sect. Comp*, Vol 44 (2015), pp 165-182.
- [2]. Nguyen Cong Hao, "Fuzzy functional dependencies based on hedge algebras", *International Journal of Computer Technology and Applications*, Vol (6) 6, 2015, pp 1052-1059.
- [3]. Nguyễn Công Hào, Trương Thị Mỹ Lê, "Mô hình cơ sở dữ liệu hướng đối tượng mờ theo cách tiếp cận đại số gia tử", *Tạp chí Tin học và Điều khiển học*, T.20, S.3 (2012), 129-140.
- [4]. Nguyễn Công Hào, "Phụ thuộc hàm mờ với lượng từ ngôn ngữ theo cách tiếp cận đại số gia tử", *Tạp chí Công nghệ thông tin và truyền thông*, T 22, S S2 (2009), 87-93.
- [5]. Cristina-Maria Vladarean, "Extending object-oriented databases for fuzzy information modeling", *S.C. WATERS Romania S.R.L, Romai J.*, 2, 1(2006), 225-237. Vol 9. No. I, pp 149-1&1.
- [6]. Z. M. Ma, "Advances in Fuzzy Object-Oriented Databases: Modeling and Applications", Idea Group Publishing, 2004.

AN APPROACH FOR PROCESSING CONSTRAINTS IN FUZZY OBJECT-ORIENTED DATABASE MODEL

Nguyen Cong Hao

Department of Academic Students' Affair, Hue University

Email: nchao@hueuni.edu.vn

ABSTRACT

In this paper, we propose an approach to process data constraints such as the domain of values of attribute constraints and between attribute constraints in the class of the fuzzy object-oriented database model based on Hedge algebras. In this approach, object validation satisfies the constraint of each level k , where k is the similarity degree of precision for each fuzzy attribute that matches the data constraints in this model. Matching the data directly to the language values is easier than many other approaches.

Keywords: Hedge algebras, fuzzy object-oriented database model, data constraints.



Nguyễn Công Hào sinh ngày 25/07/1976 tại Thừa Thiên Huế. Ông tốt nghiệp đại học ngành Toán - Tin học tại Trường ĐHSP, ĐH Huế năm 1997; tốt nghiệp thạc sĩ ngành Công nghệ thông tin tại Trường ĐHBK Hà Nội năm 2022; bảo vệ luận án tiến sĩ chuyên ngành Bảo đảm toán học cho máy tính và Hệ thống tính toán tại Viện Công nghệ thông tin, Viện Hàn lâm Khoa học và Công nghệ Việt Nam năm 2008. Hiện nay, ông đang công tác tại Đại học Huế

Lĩnh vực nghiên cứu: Cơ sở dữ liệu mờ, logic mờ, đại số gia tử, các phương pháp tính toán mềm.