PHƯƠNG PHÁP SO SÁNH ĐỐI TƯỢNG MỜ

Trần Minh Cảnh*, Trịnh Minh Thiên
Trường Đại học Phú Yên
**Email: tranminhcanh@pyu.edu.vn*
Ngày nhận bài:12/08/2024; Ngày nhận đăng: 04/02/2025

Tóm tắt

Khái niệm so sánh đóng một vai trò quyết định trong nhiều vấn đề liên quan đến quản lý trong mô hình cơ sở dữ liệu hướng đối tượng. Trong mô hình này, việc so sánh đối tượng được xử lý một cách dễ dàng dựa trên khái niệm định danh đối tượng và sự bằng nhau của các giá trị. Tuy nhiên, khi xử lý với các đối tượng không chính xác hoặc mơ hồ thì câu hỏi như "hai đối tượng bằng nhau ở mức độ nào?" hoặc "hai đối tượng tương đương nhau như thế nào?"... sẽ không có một câu trả lời rõ ràng, vì khái niệm "bằng" trở nên mờ. Bài báo này trình bày phương pháp đối sánh các đối tượng trong môi trường mờ dựa trên hướng tiếp cận đại số gia tử. Dựa trên lân cận ngữ nghĩa của đại số gia tử, việc đối sánh dữ liệu với thông tin không chắc chắn và dữ liệu kinh điển hoàn toàn nhất quán trên cơ sở đảm bảo tính thuần nhất kiểu dữ liệu.

Từ khoá: *So sánh, đại số gia tử, định danh đối tượng, thông tin không chắc chắn.*

The methods of comparing fuzzy objects

Tran Minh Canh, Trinh Minh Thien
Phu Yen University
Received: August 12, 2024; Accepted:February 4,2025

Abstract

The concept of comparison plays a decisive role in the management-related issues in the object-oriented database model. In the model, the object comparison is easily handled based on the concept of the object identification and the equality of values. However, when dealing with either inaccurate or vague objects, such questions as "to what extent are the two objects equal? Or "how equivalent are the two objects?" will not be clearly answered, as the concept of "equal" becomes fuzzy. This paper presents the methods of comparing the objects in a fuzzy environment which are presented on the base of the hedge algebra approach. Based on the semantic proximity of the hedge algebra, the comparison of data with uncertain information and canonical data are completely consistent in order to ensure the consistency of the data types.

Keywords: *Comparison, fuzzy algebra, object identity, uncertain information.*

1. Mở đầu

Trong những năm gần đây, mô hình cơ sở dữ liệu (CSDL) quan hệ, hướng đối tượng (HĐT) mờ và các vấn đề liên quan đã được nhiều tác giả quan tâm nghiên cứu và đã có những kết quả đáng kể (Biazzo & cs., 2003; Nguyễn Cát Hồ & Nguyễn Công Hào, 2008; D.V.Thang & D.V. Ban, 2011). Có nhiều cách tiếp cận khác nhau như cách tiếp cận dựa trên quan hệ tương tự (M.Koyuncu & A.Yazici, 2001), phân bố khả năng (Biazzo & cs., 2003). Tất cả các cách tiếp cận trên nhằm mục đích biểu diễn và xử lý thông tin một cách thỏa đáng trên một luận điểm nào đó về các thông tin không chính xác, không chắc chắn hay những thông tin không đầy đủ trên các mô hình CSDL quan hệ hay HĐT mờ.

Định danh đối tượng là một khái niệm cơ bản trong mô hình CSDL HĐT và được sử dụng để phân biệt các đối tượng. Khái niệm bằng nhau giữa hai đối tượng cho rằng hai đối tượng là đồng nhất nếu chúng có cùng định danh (N.Marín & cs., 2003). Tuy nhiên, trong mô hình CSDL HĐT mờ thì khái niệm đó là không đầy đủ bởi các đối tượng trong mô hình này là mờ, không chắc chắn và không đầy đủ thông tin (Ho Cam Ha & Vu Duc Quang, 2011). Do đó, để đối sánh đối tượng trong mô hình này chúng ta sử dụng khái niệm bằng nhau của các giá trị thuộc tính tức là hai đối tượng bằng nhau nếu các giá trị thuộc tính bằng nhau hoặc tương tự nhau ở một mức nào đó.

Hiện nay có nhiều hướng nghiên cứu đã được đề xuất để so sánh đối tượng trong môi trường mờ dựa trên cơ sở của cách tiếp cận quan hệ tương tự, phân bố xác suất… Trong bài báo này, chúng tôi dựa vào những ưu điểm của cấu trúc đại số gia tử (ĐSGT) trình bày một hướng tiếp cận để so sánh đối tượng trong mô hình CSDL HĐT mờ dựa trên lân cận ngữ nghĩa của ĐSGT mà 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 diễn 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 trình bày như sau: Ngoài phần mở đầu, phần 2 trình bày một số khái niệm cơ bản liên quan đến ĐSGT làm cơ sở cho các mục tiếp theo; phần 3 trình bày phương pháp đối sánh đối tượng dựa trên lân cận ngữ nghĩa của ĐSGT; phần 4 trình bày hai thuật toán **CTAC** và **CTO** để thực hiện việc so sánh hai đối tượng thuộc cùng một lớp có *xấp xỉ* với nhau *mức k* hay không và một số ví dụ minh họa cho ý tưởng của hướng tiếp cận này, cuối cùng là kết luận.

2. Một số khái niệm cơ bản

Trong phần này, chúng tôi tóm tắt lại một số khái niệm về ánh xạ định lượng (Nguyễn Cát Hồ, 2001) và cách thức xác định các hệ lân cận ngữ nghĩa định lượng.

2.1 Đại số gia tử

Cho một ĐSGT $\underline{X} = (\mathbf{X}, G, H, \leq)$, trong đó $\mathbf{X} = LDom(\underline{X})$, $G = \{\mathbf{1}, c^-, \mathbf{W}, c^+, \mathbf{0}\}$ là tập các phần tử sinh, H là tập các gia tử được xem như là các phép toán một ngôi và $\leq$ là quan hệ thứ tự ngữ nghĩa trên $\mathbf{X}$. Tập $\mathbf{X}$ được sinh ra từ tập G bởi các phép toán trong H. Như vậy, mỗi phần tử của $\mathbf{X}$ sẽ có dạng biểu diễn $x = h_n h_{n-1} \ldots\ldots h_1 x$, $x \in$ G. Tập tất cả các phần tử được sinh ra từ một phần tử x được ký hiệu là $H(x)$. Cho tập các gia tử $H = H^- \cup H^+$, trong đó $H^+ = \{h_1, \ldots, h_p\}$ và $H^- = \{h_{-1}, \ldots, h_{-q}\}$, đều là tuyến tính với thứ tự như sau: $h_1 < \ldots < h_p$ và $h_{-1} < \ldots < h_{-q}$, trong đó $p, q > 1$. Khi đó, ta có các định nghĩa liên quan như sau:

Định nghĩa 2.1 Hàm *fm*: $\mathbf{X} \to [0,1]$ được gọi là độ đo tính mờ trên X nếu thoả mãn các điều kiện sau:

(1) fm là độ đo mờ đầy đủ trên ***X***, tức là $\sum_{-q\le i\le p, i\ne 0} fm(h_i u) = fm(u)$.

(2) Nếu x là khái niệm rõ, tức là *H(x) = {x}* thì *fm(x) = 0*, do đó *fm(0)=fm(W)=fm(1)=0*.

(3) Với ∀x,y ∈ ***X***, ∀*h* ∈ *H*, ta có $\frac{fm(hx)}{fm(x)} = \frac{fm(hy)}{fm(y)}$, nghĩa là tỉ số này không phụ thuộc vào x và y, được kí hiệu là μ(h) gọi là độ đo tính mờ (fuzziness measure) của gia tử h.

Định nghĩa 2.2 (Hàm định lượng ngữ nghĩa *v*)

Cho *fm* là độ đo tính mờ trên ***X***, hàm định lượng ngữ nghĩa *v* trên ***X*** được định nghĩa như sau:

(1) $v(W) = \theta = fm(c^-)$, $v(c^-) = \theta - \alpha.fm(c^-)$ và $v(c^+) = \theta + \alpha.fm(c^+)$

(2) Nếu 1 ≤ j ≤ p thì: $v(h_j x) = v(x) + Sign(h_j x) \times \left[\sum_{i=1}^{j} fm(h_i x) - \omega(h_j x) fm(h_j x)\right]$

Nếu -q ≤ j ≤ -1 thì: $v(h_j x) = v(x) + Sign(h_j x) \times \left[\sum_{i=j}^{-1} fm(h_i x) - \omega(h_j x) fm(h_j x)\right]$

Trong đó:

$$\omega(h_j x) = \frac{1}{2}\left[1 + Sign(h_j x) Sign(h_q h_j x)(\beta - \alpha)\right] \in \{\alpha, \beta\}$$

Định nghĩa 2.3 Gọi *fm* là độ đo tính mờ trên ĐSGT ***X***, Với mỗi x ∈ ***X***, ta ký hiệu *I(x)* ⊆ [0, 1] và |I(x)| là độ dài của I(x).

Một họ J = {I(x):x∈***X***} được gọi là phân hoạch của [0, 1] nếu:

(1): $\{I(c^+), I(c^-)\}$ là phân hoạch của [0, 1] sao cho |I(c)| = *fm*(c), với $c \in \{c^+, c^-\}$.

(2): Nếu đoạn I(x) đã được định nghĩa và |I(x)| = *fm(x)* thì $\{I(h_i x)$: i = 1...p+q} được định nghĩa là phân hoạch của I(x) sao cho thoả mãn điều kiện: $|I(h_i x)| = fm(h_i x)$ và $|I(h_i x)|$ là tập sắp thứ tự tuyến tính, tức là: hoặc $I(h_1 x) < I(h_2 x) < I(h_3 x) <...< I(h_{p+q} x)$, hoặc $I(h_1 x) > I(h_2 x) > I(h_3 x) >...> I(h_{p+q} x)$.

Tập $\{I(h_i x)\}$ được gọi là phân hoạch gắn với phần tử x. Ta có $\sum_{i=1}^{p+q} |I(h_i x)| = |I(x)| = fm(x)$

Định nghĩa 2.4 Cho $X_k = \{x \in X : |x| = k\}$, xét $P^k = \{I(x) : x \in X_k\}$ 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 bao hàm trong một khoảng mờ mức k. Có nghĩa ∀u, v ∈ ***X***, $u \approx_k v \Leftrightarrow \exists \Delta^k \in P^k : I(u) \subseteq \Delta^k$ và $I(v) \subseteq \Delta^k$.

2.2 Lân cận mức k

Các tác giả Nguyễn Công Hào và Trương Thị Mỹ Lệ (2012), N.Marín & cs. (2003) 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* chúng ta mong muốn các giá trị đại diện như vậy phải là điểm trong của *lân cận mức k*.

Dựa vào khoảng mờ mức *k* và mức *k+1* các tác giả trên 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 $I(h_{-q}c^+) \leq I(h_{-q+1}c^+) \ldots \leq I(h_{-2}c^+) \leq I(h_{-1}c^+) \leq \upsilon_A(c^+) \leq I(h_1c^+) \leq I(h_2c^+) \leq \ldots \leq I(h_{p-1}c^+) \leq I(h_pc^+)$. Khi đó, ta xây dựng phân hoạch về độ *tương tự mức 1* gồm các lớp tương đương sau: $S(\boldsymbol{0}) = I(h_pc^-)$; $S(c^-) = I(c^-) \setminus [I(h_{-q}c^-) \cup I(h_pc^-)]$; $S(\boldsymbol{W}) = I(h_{-q}c^-) \cup I(h_{-q}c^+)$; $S(c^+) = I(c^+) \setminus [I(h_{-q}c^+) \cup I(h_pc^+)]$ và $S(\boldsymbol{1}) = I(h_pc^+)$.

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

(2). *Độ tương tự mức 2*: với $k = 2$, các khoảng mờ mức 2 gồm $I(h_ic^+)$ và $I(h_ic^-)$, với $-q \leq i \leq p$. Chúng ta sẽ có các lớp tương đương sau: $S(\boldsymbol{0}) = I(h_ph_pc^-)$; $S(h_ic^-) = I(h_ic^-) \setminus [I(h_{-q}h_ic^-) \cup I(h_ph_ic^-)]$; $S(\boldsymbol{W}) = I(h_{-q}h_{-q}c^-) \cup I(h_{-q}h_{-q}c^+)$; $S(h_ic^+) = I(h_ic^+) \setminus [I(h_{-q}h_ic^+) \cup I(h_ph_ic^+)]$ và $S(\boldsymbol{1}) = I(h_ph_pc^+)$, với $-q \leq i \leq p$.

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 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 $S(x_1)$, $S(x_2)$, …, $S(x_m)$. Khi đó, mỗi giá trị mờ *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 đối sánh đối tượng

Ví dụ 1: Chúng ta xét mô hình CSDL HĐT được cho như sau:

Lớp PhongHoc có các thuộc tính là dienTich và tập các sinh viên tham dự buổi học của họ trong mỗi phòng học. Lớp SinhVien được miêu tả bởi tập thuộc tính tenSV, tuoi, chieuCao. Sau đây là một số thể hiện của lớp PhongHoc và lớp SinhVien, để đơn giản chúng tôi chỉ giới hạn bảng dữ liệu chỉ gồm những thuộc tính.

PhongHoc		
iDPH	dienTich	sv
iD1	$30m^2$	{sv1,sv2,sv3,sv4}
iD2	lớn	{sv1,sv3,sv5,sv6}

SinhVien			
iDSV	tenSV	tuoi	chieuCao
iD1	An	trẻ	1.85
iD2	Bình	trẻ	1.7
iD3	Hà	20	Thấp
iD4	Hương	24	Cao
iD5	Nhân	25	1.8
iD6	Thủy	trẻ	1.9

Trong ví dụ này, các đối tượng của lớp PhongHoc là các đối tượng mờ (không chính xác) do các lý do sau đây:

- Thuộc tính dienTich được biểu diễn bởi giá trị số hoặc nhãn không xác định giá trị chính xác (ví dụ: nhỏ, lớn).

- Tập các đối tượng sinh viên là mờ, và tùy theo mỗi bài học khác nhau mà số lượng sinh viên trong phòng học cũng khác nhau.

Ngoài ra, các đối tượng thuộc lớp SinhVien có thể mờ bởi các lý do sau:

- Thuộc tính tuoi có thể biểu diễn bằng giá trị số hoặc nhãn không xác định giá trị chính xác (ví dụ: trẻ, già...).

- Thuộc tính chieuCao cũng có thể biểu diễn bằng giá trị số hoặc nhãn không xác định giá trị chính xác (ví dụ: thấp, cao...).

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). Để thực hiện việc đối sánh hai đối tượng, chúng ta cần phải giải quyết các vấn đề: Thứ nhất, đó là đối sánh các giá trị thuộc tính trên miền cơ sở (ví dụ: thuộc tính dienTich) và thứ hai, là thực hiện đối sánh các sưu tập mờ của các đối tượng mờ (không chính xác).

3.1. Đối sánh giá trị thuộc tính trong miền cơ sở

Trong ví dụ 1, thuộc tính dienTich nhận các giá trị đơn (giá trị rõ hoặc mờ). Những thuộc tính như vậy được chúng tôi xem là những thuộc tính kinh điển hoặc thuộc tính mờ.

Cho lớp mờ C với tập thuộc tính A = $\{a_i$: i = 1,.., n$\}$, trong đó a_i là các thuộc tính kinh điển hoặc thuộc tính mờ. Miền giá trị của thuộc tính kinh điển của a_i là $CDom(a_i)$ và ta viết $Dom(a_i) = CDom(a_i)$, miền giá trị của thuộc tính mờ là $Dom(a_j) = CDom(a_j) \cup FDom(a_j)$ $(1 \le i \le n)$.

Trên cơ sở khái niệm lân cận, chúng ta trình bày định nghĩa về quan hệ đối sánh giữa các phần tử trong miền giá trị của thuộc tính mờ 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, o_1, $o_2 \in$ C và *k* là mức phân hoạch. Ta nói rằng $o_1.a_i$ *bằng nhau bậc k* với $o_2.a_i$ được ký hiệu $o_1.a_i =_k o_2.a_i$ nếu:

(1) Nếu $o_1.a_i, o_2.a_i \in CDom(a_i)$ thì $o_1.a_i = o_2.a_i$ hoặc tồn tại $FRN_k(o_2.a_i)$ sao cho $o_1.a_i \in FRN_k(o_2.a_i)$.

(2) Nếu $o_1.a_i$ hoặc $o_2.a_i \in FDom(a_i)$, chẳng hạn $o_1.a_i$ thì ta phải có $o_2.a_i \in FRN_k(o_1.a_i)$.

(3) Nếu $o_1.a_i, o_2.a_i \in FDom(a_i)$, thì $FRN_k(o_2.a_i) = FRN_k(o_1.a_i)$.

Ví dụ 2: Giả sử trong CSDL HĐT mờ, có một lớp tên là BoPhan, với một thuộc tính mờ a, các giá trị thuộc tính a tương ứng trong 4 đối tượng thuộc lớp BoPhan là o_1.a = cao; o_2.a = 80; o_3.a = 70; o_4.a = 90. Hệ lân cận được xây dựng như sau:

Xét ĐSGT của biến ngôn ngữ a, trong đó D_a = [0, 100], các phần tử sinh là {0, thấp, W, cao, 1}, tập các gia tử là {ít, khá, hơn, rất}, $FD_a = H_a(cao) \cup H_a(thấp)$.

Chọn fm(cao) = 0.60, fm(thấp) = 0.40, μ(khá) = 0.15, μ(ít) = 0.25, μ(hơn) = 0.25 và μ(rất) = 0.35. Đoạn [0, 100] được phân hoạch thành 5 khoảng tương tự mức 1 như sau:
fm(rấtcao) * 100 = 0.35*0.60*100=21. vậy S(1)*100= (79, 100].
(fm(khácao)+fm(hơncao))*100=(0.25*0.60+0.15*0.60)*100=24, vậy S(cao)*100=(55, 79].
(fm(ítthấp) + fm(ítcao))*100 = (0.25*0.60 + 0.25*0.40)*100 =25, vậy S(W)*100 = (30, 55].
(fm(kháthấp) + fm(hơnthấp))*100 = (0.25*0.40 + 0.15*0.40)*100 = 16, vậy S(thấp)*100 = (14, 30], và S(0)*100 = [0, 14].

Từ đó, ta có lân cận mức 1 của các lớp tương tự như sau: $FRN_1(0)$ = [0, 14], FRN_1(thấp) = (14, 30], $FRN_1(W)$ = (30, 55], FRN_1(cao) = (55, 79] và $FRN_1(1)$ = (79, 100].

Vậy ta nói $o_1.a =_1 o_3.a$ vì $o_1.a = cao \in FRN_1(cao)$ và $o_3.a = 70 \in FRN_1(cao)$; hoặc $o_2.a =_1 o_4.a$ vì $o_2.a = 80 \in FRN_1(1)$ và $o_4.a = 90 \in FRN_1(1)$. Với *bậc k =1*.

3.2. Đối sánh các sưu tập đối tượng mờ

Trong phần 3.1, chúng ta đã xem xét giải pháp đối sánh giá trị của hai thuộc tính nhận các giá trị đơn (giá trị rõ hoặc mờ). Trong phần này, chúng ta tiếp tục xác định các giải pháp để đối sánh những thuộc tính mà giá trị của nó là sưu tập.

Trong lớp PhongHoc được định nghĩa ở ví dụ 1 thì thuộc tính sV nhận giá trị sưu tập. Có nghĩa là thuộc tính này sẽ lưu giữ một sưu tập các đối tượng SinhVien của mỗi đối tượng PhongHoc thuộc lớp PhongHoc.

Trên cơ sở khái niệm lân cận ngữ nghĩa, định nghĩa về *xấp xỉ mức k* cho 2 thuộc tính nhận giá trị sưu tập được xác định như sau:

Định nghĩa 3.2

Cho lớp mờ C xác định trên tập thuộc tính A, $a_i \subseteq A$ và a_i là thuộc tính nhận giá trị sưu tập, $o_1, o_2 \in C$ và *k* là mức phân hoạch. Ta nói rằng $o_1.a_i$ *xấp xỉ bậc k* với $o_2.a_i$ được ký hiệu $o_1.a_i \approx_k o_2.a_i$, nếu hai tập $\Omega(X)$ và $\Omega(Y)$ có cùng các lớp tương tự. Trong đó $\Omega(X)$ và $\Omega(Y)$ là tập các lớp tương tự *mức k* của $o_1.a_i$ và $o_2.a_i$.

Ví dụ 3: Xét mô hình CSDL HĐT được cho như sau:

Giả sử chúng ta có các đối tượng được thể hiện trong mô hình CSDL HĐT trên như sau:

MonHoc		
iDMH	tenMH	hV
id1	CTDL	{hV_1, hV_4, hV_6}
id2	CSDL	{hV_2, hV_3, hV_5}

HocVien		
iDHV	tenHV	tuoi
iD1	An	18
iD2	Bình	trẻ
iD3	Hà	24
iD4	Hương	32
iD5	Nhân	34
iD6	Thủy	trẻ

Chúng ta kiểm tra $o_1.hV \approx_1 o_2.hV$, với mức phân hoạch *k=1* và $o_1, o_2 \in MonHoc$.

Để thực hiện kiểm tra xem $o_1.hV$ có *xấp xỉ bậc k* với $o_2.hV$ dựa vào định nghĩa 3.2 chúng ta thực hiện như sau:

Bước 1: Xây dựng các lớp tương tự cho thuộc tính *tuoi* cho sáu đối tượng thuộc lớp HocVien, bởi vì thuộc tính này gây ra tính mờ cho các đối tượng này.

Chúng ta xét ĐSGT của biến ngôn ngữ tuoi, trong đó D_{tuoi} = [0, 100], các phần tử sinh là {0, trẻ, W, già, 1}, tập các gia tử là {ít, khả năng, hơn, rất}, $FD_{tuoi} = H_{tuoi}$(trẻ) $\cup$ H_{tuoi}(già).

Chọn fm(trẻ) = 0.4, fm(già) = 0.6, μ(khả năng) = 0.25, μ(ít) = 0.2, μ(hơn) = 0.15 và μ(rất) = 0.4. Đoạn [0, 100] được phân hoạch thành 5 khoảng tương tự mức 1 như sau:
fm(rất trẻ) * 100 = 0.4*0.4*100=16. Vậy S(0)*100= [0, 16).
(fm(khảnăng trẻ)+fm(hơntrẻ))*100=(0.25*0.4+0.15*0.4)*100=16, vậy S(trẻ)*100=[16, 32).
(fm(ít trẻ) + fm(ít già))*100= (0.2*0.6 + 0.25*0.4)*100 =20, vậy S(W)*100 = [32, 52).
(fm(khả năng già) + fm(hơn già))*100= (0.25*0.6 + 0.15*0.6)*100= 24, vậy S(già)*100 = [52, 76), và S(1)*100 = [76, 100].

Từ đó, ta có lân cận mức 1 của các lớp tương tự như sau: FRN_1(0)= [0, 16), FRN_1(trẻ)= [16, 32), FRN_1(W) = [32, 52), FRN_1(già) = [52, 76) và FRN_1(1) = [76, 100].

Bước 2: Vì hV_1.tuoi= 18, hV_6.tuoi= trẻ $\in$ FRN_1(trẻ), và hV_4.tuoi= 32 $\in$ FRN_1(W), nên tập các lớp tương của thuộc tính $o_1.hV$ là $\Omega(o_1.hV)$ = {FRN_1(trẻ), FRN_1(W)}.

Và vì hV_2.tuoi= trẻ, hV_3.tuoi= 24 $\in$ FRN_1(trẻ), và hV_5.tuoi= 34 $\in$ FRN_1(W), nên tập các lớp tương của thuộc tính $o_2.hV$ là $\Omega(o_2.hV)$ = {FRN_1(trẻ), FRN_1(W)}.

Dễ dàng nhận thấy rằng tập các lớp tương tự trong $\Omega(o_1.hV)$ và $\Omega(o_2.hV)$ là hoàn toàn như nhau, nghĩa là $o_1.hV \approx_1 o_2.hV$ với mức phân hoạch *k= 1*.

4. Thuật toán so sánh hai đối tượng mờ

Phần này trình bày hai thuật toán dựa vào những khái niệm đã được nêu trên. Đó là các thuật toán: Thuật toán **CTAC** thực hiện việc kiểm tra 2 thuộc tính nhận giá trị tập có *xấp xỉ* nhau *mức k* hay không; và thuật toán **CTO** thực hiện việc kiểm tra xem hai đối tượng thuộc cùng một lớp có *xấp xỉ* nhau *mức k* hay không.

Thuật toán CTAC ***(Comparing Two Attributes Collection)***

Vào: Một lớp C_1 với m thuộc tính; O = {o_1, o_2, ..., o_n} $\in$ C_1.

X, Y hai thuộc tính chứa tập các đối tượng O = {o_i: i = 1..n} $\in$ C_1.

Ra: *True*: Nếu X xấp xỉ mức k với Y và *False* nếu ngược lại.

Phương pháp:

// Khởi tạo các giá trị.

(1) **For** (i = 1; i<=n; i++)

(2) **For** (j = 1; i<=m; j++)

(3) { //**Begin**

(4) $G_{o_i.a_j}$ = {$\boldsymbol{0}$, $c^-_{o_i.a_j}$, $\boldsymbol{W}$, $c^+_{o_i.a_j}$, $\boldsymbol{1}$}; $H_{o_i.a_j} = H^+_{o_i.a_j} \cup H^-_{o_i.a_j}$. Trong đó $H^+_{o_i.a_j}$ = {h_1, h_2}, $H^-_{o_i.a_j}$ = {h_3, h_4}, với $h_1 < h_2$ và $h_3 > h_4$. Chọn độ đo tính mờ cho các phần tử sinh và gia tử.

(5) $D_{o_i.a_j} = [\min_{o_i.a_j}, \max_{o_i.a_j}]$.

(6) $FD_{o_i.a_j} = H_{o_i.a_j}(c^+_{o_i.a_j}) \cup H_{o_i.a_j}(c^-_{o_i.a_j})$.

(7) } //**End**

//Xây dựng các *khoảng tương tự mức k =1.*

(8) ***For*** (i = 1; i<=n; i++)

(9) ***For*** (j = 1; j<=m; j++)

(10) ***For*** (h = 1; h<=8; h++)

(11) *Xây dựng các khoảng tương tự mức k:* $S^k_{o_i.a_j}(x_h)$*;*

// Xác định tập các lớp tương tự cho o_i.

(12) **For** (i = 1; i<=n; i++)

(13) **For** (ij= 1; j<=m; j++)

(14) { //**Begin**

(15) t=0;

(16) **Do**

(17) t=t+1;

(18) **While** *(*$o_i.a_j \in S^k_{o_i.a_j}(x_t)$ or t > *8)*;

(19) ***If*** $o_i \in$ X $\Omega_1(XX) = \Omega_1(XX) \cup S^k_{o_i.a_j}(x_t)$;

(20) ***Else*** $\Omega_2(YY) = \Omega_2(YY) \cup S^k_{o_i.a_j}(x_t)$;

(21) } **//End**

//So sanh hai tập các lớp tương tự.

(22) ***For*** *each* $S^k_{o_i.a_j}(x) \in \Omega_1(XX)$

(23) ***If*** $S^k_{o_i.a_j}(x) \notin \Omega_2(YY)$ Return False;

(24) Return True;

Thuật toán CTO *(Compare Two Objects)*

Vào: Một lớp C với m thuộc tính; O = {o_1, o_2} $\in$ C.

Trong đó: (o_1, o_2,).[a_1, a_2, ..., a_{p-1}] là những thuộc tính nhận giá trị đơn và (o_1, o_2,).a_p là thuộc tính nhận giá trị sưu tập.

Ra: *True*: Nếu o_1 xấp xỉ mức k với o_2 và *False* nếu ngược lại.

Phương pháp:

// Khởi tạo các giá trị.

(1) **For** (i = 1; i<=2; i++)

(2) **For** (j = 1; j<=p-1; j++)

(3) **{** //**Begin**

(4) $G_{o_i.a_j} = \{\boldsymbol{0}, c^-_{o_i.a_j}, \boldsymbol{W}, c^+_{o_i.a_j}, \boldsymbol{1}\}$; $H_{o_i.a_j} = H^+_{o_i.a_j} \cup H^-_{o_i.a_j}$. Trong đó $H^+_{o_i.a_j} = \{h_1, h_2\}, H^-_{o_i.a_j} = \{h_3, h_4\}$, với $h_1<h_2$ và $h_3>h_4$. Chọn độ đo tính mờ cho các phần tử sinh và gia tử.

(5) $D_{o_i.a_j} = [\min_{o_i.a_j}, \max_{o_i.a_j}]$.

(6) $FD_{o_i.a_j} = H_{o_i.a_j}(c^+_{o_i.a_j}) \cup H_{o_i.a_j}(c^-_{o_i.a_j})$.

(7) **} //End**

//Xây dựng các *khoảng tương tự mức k =1.*
For (i = 1; i<=2; i++)
    For (j = 1; j<=p-1; j++)
        For (h = 1; h<=8; h++)
            Xây dựng các khoảng tương tự mức k: $S^k_{o_i.a_j}(x_h)$;
// Xác định *lân cận mức k = 1* của $o_i.a_j$.
For (i = 1; i<=2; i++)
    For (j = 1; j<=p-1; j++)
    { //**Begin**
        t=0;
        Do
            t=t+1;
        While $(o_i.a_j \in S^k_{o_i.a_j}(x_t)$ or $t > 8)$;
        If *i = 1* $\Omega_1(X) = \Omega_1(X) \cup S^k_{o_i.a_j}(x_t)$;
        Else $\Omega_2(Y) = \Omega_2(Y) \cup S^k_{o_i.a_j}(x_t)$;
    } //**End**
//Kiểm tra hai đối tượng có xấp xỉ mức k hay không.
    If $(o_1.a_m, o_2.a_m)$ = *True*
        { //**Begin**
        For *each* $S^k_{o_i.a_j}(x) \in \Omega_1(X)$
            If $S^k_{o_i.a_j}(x) \notin \Omega_2(Y)$ Return False;
        } //**End**
    Else Return False;
    Return True;

Định lý: Thuật toán **CTAC** và **CTO** 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, m, p là hữu hạn) nên thuật toán sẽ dừng sau khi duyệt xong tất cả các đối tượng.

2. *Chứng minh tính đúng đắn*:

(i). Đối với mỗi một thuộc tính a_i ($1 \le i \le n$) trong đối tượng o $\in$C, mà giá trị thuộc tính có thể nhận là giá trị kinh điển (giá trị rõ) hay giá trị ngôn ngữ (giá trị mờ). Trong quan hệ đối sánh đối với dữ liệu chúng tôi chia thành 2 trường hợp sau:

Trường hợp 1: Đối với giá trị thuộc tính kinh điển (giá trị rõ) vẫn sử dụng phép toán = để thực hiện việc đối sánh dữ liệu.

Trường hợp 2: Đối với giá trị ngôn ngữ, chúng tôi sử dụng các phép toán đối sánh theo mức $=_k$, với *k* là khoảng *lân cận mức k* theo ĐSGT. Dựa vào định lượng ngữ nghĩa của ĐSGT chúng ta có thể tính được *lân cận mức k* của một từ *x* là $FRN_k(x)$ = [a, b], theo các trường hợp sau:

a) Nếu y là giá trị kinh điển (rõ) mà y $\in$[a, b] thì y $=_k$ x.

b) Nếu y là giá trị ngôn ngữ trong khoảng $[x_1, x_2]$ (tính được thông qua định lượng ngữ nghĩa) mà a <= x_1 và x_2 <= b thì $y =_k x$.

(ii). Đối với việc đối sánh hai thuộc tính mà giá trị thuộc tính là giá trị sưu tập. Trong trường hợp này cũng dựa vào định lượng ngữ nghĩa của ĐSGT chúng ta xây dựng được hai tập các lớp tương đương $FRN_k(x)$ cho hai thuộc tính này là X và Y. Nếu các phần tử trong hai tập X và Y là hoàn toàn như nhau thì thuật toán **CTAC** trả về *True*, có nghĩa là $X \approx_k Y$, còn ngược lại là *False*.

Vậy, thuật toán **CTO** luôn cho đúng kết quả đầu ra như mong muốn.

Độ phức tạp tính toán của thuật toán **CTAC** là max(O(n*m), O(n*n)), với n là số đối tượng và m là số thuộc tính của lớp C_1. Độ phức tạp thuật toán **CTO** là max(O(p), max(O(n*m), O(n*n))), với p là số thuộc tính của lớp C.

Ví dụ 4: *Kiểm tra hai đối tượng được cho ở ví dụ 1 có xấp xỉ mức k=1 hay không?.*

Bước (1)-(21): Chọn độ đo tính mờ và xây dựng mức phân hoạch k=1 cho các thuộc tính sau: tuoi, chieuCao ở lớp SinhVien, và dienTich ở lớp PhongHoc.

Đối với thuộc tính chieuCao: $D_{chieuCao}$ = [1.0, 2.0], các phần tử sinh là {0, thấp, W, cao, 1}, tập các gia tử là {ít, khả năng, hơn, rất}, $FD_{chieuCao} = H_{chieuCao}$ (cao) ∪ $H_{chieuCao}$ (thấp).

Chọn *fm(thấp)*= 0.65, *fm(cao)*= 0.35, *fm(rất)*= 0.2, *fm(hơn)*= 0.25, *fm(khả năng)*= 0.25, *fm(ít)*= 0.3. Đoạn [1.0, 2.0] được phân hoạch thành 5 khoảng tương tự mức 1 và tương tự như cách tính trong ví dụ 4, chúng ta cũng có các lớp tương tự như sau: $FRN_1^{chieuCao}$(0)= [1,1.06], $FRN_1^{chieuCao}$(thấp)= (1.06, 1.385], $FRN_1^{chieuCao}$(W)= (1.385, 1.685], $FRN_1^{chieuCao}$ (cao) = (1.685, 1.93] và $FRN_1^{chieuCao}$(1)= (1.93, 2].

Đối với thuộc tính tuoi: D_{tuoi} = [0, 100], các phần tử sinh là {0, trẻ, W, già, 1}, tập các gia tử là {ít, khả năng, hơn, rất}, $FD_{tuoi} = H_{tuoi}$(trẻ) ∪ H_{tuoi}(già).

Chọn fm(trẻ) = 0.4, fm(già) = 0.6, μ(khả năng) = 0.25, μ(ít) = 0.2, μ(hơn) = 0.15 và μ(rất) = 0.4. Tương tự như cách tính ở ví dụ 4, chúng ta cũng có các lớp tương tự như sau: FRN_1^{tuoi}(0)= [0, 16), FRN_1^{tuoi}(trẻ)= [16, 32), FRN_1^{tuoi}(W)= [32, 52), FRN_1^{tuoi}(già)= [52, 76) và FRN_1^{tuoi}(1)= [76, 100].

Đối với thuộc tính dienTich: $D_{dienTich}$ = [0, 50], các phần tử sinh là {0, nhỏ, W, lớn, 1}, tập các gia tử là {ít, khả năng, hơn, rất}, $FD_{dienTich} = H_{dienTich}$(nhỏ) ∪ $H_{dienTich}$(lớn).

Chọn *fm(nhỏ)*= 0.6, *fm(lớn)*= 0.4, *fm(rất)*= 0.35, *fm(hơn)*= 0.25, *fm(khả năng)*= 0.2, *fm(ít)*= 0.2. Đoạn [0, 50] được phân hoạch thành 5 khoảng tương tự mức 1 và tương tự như cách tính trong ví dụ 4, chúng ta cũng có các lớp tương tự như sau: $FRN_1^{dienTich}$(0)= [0,10.5], $FRN_1^{dienTich}$(nhỏ)= (10.5, 24], $FRN_1^{dienTich}$(W)= (24, 34], $FRN_1^{dienTich}$(lớn) = (34, 43] và $FRN_1^{dienTich}$(1)= (43, 50].

Bước (22)-(28): Đối sánh các giá trị thuộc tính của hai đối tượng để xác định hai đối tượng o_1 và o_2 có *xấp xỉ* nhau mức *k=1* hay không.

(1). Đối sánh o_1.dienTich và o_2.dienTich: o_1.dienTich = 35 ∈ $FRN_1^{dienTich}$(lớn), o_2.dienTich = lớn ∈ $FRN_1^{dienTich}$(lớn), nên $o_1.dienTich \approx_1 o_2.dienTich$

(2). Đối sánh o1.sV và o2.sV:

- o_1.sV = {sv1, sv2, sv3, sv4}, với các lớp tương đương được xây dựng cho các thuộc tính của các đối tượng thuộc lớp SinhVien ở trên, ta có tập các lớp tương đương của o_1.sV là $\Omega_1(o_1.sV)$={ FRN_1^{tuoi} (trẻ), $FRN_1^{chieuCao}$ (thấp), $FRN_1^{chieuCao}$ (cao)}.

- o2.sV = {sv1, sv3, sv5, sv6}, và tập các lớp tương đương của o_2.sV là $\Omega_2(o_2.sV)$ ={ FRN_1^{tuoi} (trẻ), $FRN_1^{chieuCao}$ (thấp), $FRN_1^{chieuCao}$ (cao)}.

Dễ dàng nhận thấy tập tương đương $\Omega_1(o_1.sV)$ của o_1.sV và $\Omega_2(o_2.sV)$ của o_2.sV là hoàn toàn như nhau, vậy $o_1.sV \approx_1 o_2.sV$.

Từ (1) và (2) suy ra hai đối tượng o_1 và o_2 xấp xỉ nhau mức k = 1, nghĩa là $o_1 \approx_1 o_2$

5. Kết luận

Bài báo đã trình bày cách tiếp cận để đối sánh hai đối tượng mờ thuộc cùng một lớp trong mô hình CSDL HĐT mờ dựa trên cách tiếp cận của đại số gia tử. Cách tiếp cận này làm cho việc xử lý dữ liệu trở nên đơn giản và thuần nhất dữ liệu. Dựa trên định lượng ngữ nghĩa để xác định *lân cận mức k* của giá trị ngôn ngữ và thực hiện việc đối sánh dữ liệu trên *lân cận mức k* này. Hai thuật toán thực hiện việc kiểm tra xem các đối tượng có *xấp xỉ với nhau mức k* hay không: (1) thuật toán **CTAC** thực hiện việc đối sánh thuộc tính nhận giá trị sưu tập; (2) thuật toán **CTO** thực hiện việc đối sánh hai đối tượng thuộc cùng một lớp. Kết quả sẽ là sơ sở để thực hiện hướng tiếp cận giải pháp đối sánh hai đối tượng mà giá trị thuộc tính của nó tham chiếu đến đối tượng khác (đối tượng phức hợp) hoặc xử lý những dữ liệu lớn trong vấn đề máy học ❑

TÀI LIỆU THAM KHẢO

Biazzo V., Giugno R, Lukasiewiez T., Subrahmanian V.S (2003). Temporal probabillistic object bases. *IEEE Transaction on Knowledge and Engineering*, 15, pp 921-939.

D.V.Thang, D.V.Ban (2011). *Query Data With Fuzzy Information In Object-Oriented Databases An Approach The Semantic Neighborhood Of Hedge Algebras*. International Journal of Computer Science and Information Security (IJCSIS), Vol9.No5, pp 37-42.

M.Koyuncu, A.Yazici (2001). *A fuzzy database and knowledge base environment for intelligent retrieval*, Proceedings of the IFSA/NAFIPS World Congress.

Nguyễn Cát Hồ (2002). *Lý thuyết tập mờ và công nghệ tính toán mềm, Hệ mờ, mạng nơron và ứng dụng.* NXB Khoa học kỹ thuật.

Ho Cam Ha, Vu Duc Quang (2011). *Fuzzy function dependencies in fuzzy object-oriented databases*, Journal of HNUE 7, 23-31.

Nguyễn Công Hào, Trương Thị Mỹ Lệ (2012). 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* **20 (**3) 129-140.

Nguyễn Cát Hồ, Nguyễn Công Hào (2008). Một phương pháp xử lý truy vấn 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*, T.24, S.4, pp 281-294.

N.Marín, J.M.Medina, O.Pons, D.Sánchez, M.A.Vila (2003). *Complex object comparison in a fuzzy context*, Information and Software Technology 45, 431–444.