

## MÔ HÌNH CƠ SỞ DỮ LIỆU HƯỚNG ĐỐI TƯỢNG MỜ DỰA TRÊN NGŨ NGHĨA ĐẠI SỐ GIA TỬ\*

NGUYỄN CÔNG HÀO<sup>1</sup>, TRƯƠNG THỊ MỸ LÊ<sup>2</sup>

<sup>1</sup>Trung tâm công nghệ thông tin, Đại học Huế

<sup>2</sup>Trường Đại học Quang Trung, Qui Nhơn

**Tóm tắt.** Trong thời gian gần đây, mô hình cơ sở dữ liệu hướng đối tượng mờ đã được nhiều tác giả quan tâm nghiên cứu theo nhiều cách tiếp cận khác nhau như lý thuyết tập mờ, lý thuyết khả năng. Tuy nhiên, các cách tiếp cận này vẫn còn nhiều hạn chế trong biểu diễn và đối sánh dữ liệu. Vì vậy, trong bài báo này, chúng tôi sử dụng một hướng tiếp cận mới có thể khắc phục được các hạn chế của các cách tiếp cận khác đó là dựa trên đại số gia tử để xây dựng mô hình cơ sở dữ liệu hướng đối tượng mờ. Một số phép toán đại số được đề xuất phù hợp với mô hình mới này. Cuối cùng, chúng tôi đưa ra một phương pháp mới xử lý truy vấn hướng đối tượng mờ một cách linh hoạt.

**Abstract.** In recent times, fuzzy object-oriented databases model has been studied in many different approaches such as fuzzy set theory, possibility theory,... However, the mentioned approaches are still limited in data performance and matching. In this paper, we propose a new approach to overcome the limitations of the approach that is based on hedge algebra structure for building fuzzy object-oriented databases model. Some operators fuzzy object-oriented relation algebra are proposed corresponding with the model. Finally, we propose a method of fuzzy object-oriented query processing flexibly.

### 1. ĐẶT VẤN ĐỀ

Mô hình cơ sở dữ liệu hướng đối tượng mờ được các tác giả trong và ngoài nước quan tâm nghiên cứu và đã có nhiều kết quả đáng kể [5,6,7,9]. Các mô hình cơ sở dữ liệu hướng đối tượng mờ đã được nghiên cứu chủ yếu dựa theo cách tiếp cận lý thuyết tập mờ, quan hệ tương tự và lý thuyết khả năng... Tuy đã có nhiều cách tiếp cận để xử lý thông tin mờ nhưng việc tính toán, xử lý và đối sánh các đối tượng mờ trong mô hình nhằm xây dựng các thao tác dữ liệu vẫn còn phức tạp, khó khăn và hạn chế. Hầu hết việc biểu diễn và đối sánh dữ liệu vẫn phức tạp và mang tính chủ quan, phụ thuộc vào nhiều yếu tố làm ảnh hưởng đến hiệu quả của việc thao tác dữ liệu. Chẳng hạn như theo cách tiếp cận lý thuyết tập mờ, yếu tố ảnh hưởng vào việc biểu diễn ngữ nghĩa là việc xây dựng hàm thuộc và chọn ngưỡng lát cắt  $\alpha$  của tập mờ, theo cách tiếp cận quan hệ tương tự là việc chọn ngưỡng tương tự giữa hai giá trị, ngưỡng của mỗi thuộc tính và ngưỡng của bộ dữ liệu... Vì vậy, cần có một cách tiếp cận

---

\*Nghiên cứu này được hoàn thành dưới sự hỗ trợ từ Quỹ phát triển khoa học và Công nghệ quốc gia (NAFOSTED)

để xử lý thông tin mờ một cách hiệu quả hơn, đơn giản và trực quan hơn. Với ưu điểm của đại số gia tử (ĐSGT) trong quá trình xây dựng mô hình cơ sở dữ liệu mờ [1], chúng tôi sử dụng cách tiếp cận mới này để xây dựng mô hình cơ sở dữ liệu hướng đối tượng mờ. Trước hết một số khái niệm cơ bản trong cơ sở dữ liệu hướng đối tượng như đối tượng, lớp, quan hệ lớp đối tượng, lớp con, lớp cha, và đa thừa kế được mở rộng trong cơ sở dữ liệu hướng đối tượng mờ. Tiếp theo, một số phép toán và thao tác dữ liệu được đề xuất hiệu quả, phù hợp với mô hình mới.

Bài báo gồm 5 mục. Mục 2 trình bày một số kiến thức cơ bản; Mục 3 trình bày mô hình cơ sở dữ liệu hướng đối tượng mờ và một số phép toán; Mục 4 trình bày phương pháp xử lý truy vấn; Và cuối cùng là một số nhận xét kết luận cho bài báo.

## 2. CÁC KHÁI NIỆM CƠ SỞ

### 2.1. Đại số gia tử

Cho một ĐSGT tuyến tính đầy đủ  $\underline{X} = (\mathbf{X}, G, H, \Sigma, \Phi, \leq)$ , trong đó  $Dom(\mathbf{X}) = \mathbf{X}$  là miền các giá trị ngôn ngữ của thuộc tính ngôn ngữ  $\mathbf{X}$  được sinh tự do từ tập các phần tử sinh  $G = \{1, c^+, W, c^-, 0\}$  bằng việc tác động tự do các phép toán một ngôi trong tập  $H, \Sigma$  và  $\Phi$  là hai phép tính với ngữ nghĩa là cận trên đúng và cận dưới đúng của tập  $H(x)$ , tức là  $\Sigma x = \sup H(x)$  and  $\Phi x = \inf H(x)$ , 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. Ví dụ, nếu ta có thuộc tính *Luong* là “Lương thu nhập của nhân viên trong một tháng”, thì  $Dom(Luong) = \{high, low, veryhigh, morehigh, possiblyhigh, verylow, possiblylow, lesslow, \dots\}$ ,  $G = \{1, high, W, low, 0\}$ ,  $H = \{very, more, possibly, less\}$  và  $\leq$  một quan hệ thứ tự cảm sinh từ ngữ nghĩa của các từ trong  $Dom(Luong)$ , chẳng hạn ta có  $veryhigh > high$ ,  $morehigh > high$ ,  $possiblyhigh < high$ ,  $lesshigh < high, \dots$  Cho tập các gia tử  $H = H^- \cup H^+$ , trong đó  $H^+ = \{h_1, \dots, h_p\}$  và  $H^- = \{h_{-1}, \dots, h_{-q}\}$ , với  $h_1 < \dots < h_p$  và  $h_{-1} < \dots < h_{-q}$ , trong đó  $p, q > 1$ . Ký hiệu  $fm : \mathbf{X} \rightarrow [0, 1]$  là độ đo tính mờ của ĐSGT  $\underline{X}$ . Khi đó ta có:

**Định nghĩa 2.1.** Với mỗi  $x \in \underline{X}$ , độ dài của  $x$  được ký hiệu  $|x|$  và xác định như sau:

- (a) Nếu  $x = c^+$  hoặc  $x = c^-$  thì  $|x| = 1$ .
- (b) Nếu  $x = hx'$  thì  $|x| = 1 + |x'|$ , với mọi  $h \in H$ .

**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:

- (a)  $fm(hx) = \mu(h)fm(x), \forall x \in \underline{X}$ ;
- (b)  $fm(c^-) + fm(c^+) = 1$ ;
- (c)  $\Sigma_{-q \leq i \leq p, i \neq 0} fm(h_i c) = fm(c)$  trong đó  $c \in \{c^-, c^+\}$ ;
- (d)  $\Sigma_{-q \leq i \leq p, i \neq 0} fm(h_i x) = fm(x), x \in \underline{X}$ ;
- (e)  $\Sigma\{\mu(h_i) : -q \leq i \leq -1\} = \alpha$  và  $\Sigma\{\mu(h_i) : 1 \leq i \leq p\} = \beta$ , trong đó  $\alpha, \beta > 0$  và  $\alpha + \beta = 1$ .

**Định nghĩa 2.2.** (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  $c \in \{c^-, c^+\}$ :

$$(a) Sgn(c^-) = -1, Sgn(c^+) = +1;$$

(b)  $Sgn(h'hx) = 0$ , 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  $h$  (hoặc  $c$ , nếu  $h = I$  và  $x = c$ );

$Sgn(h'hx) = +Sgn(hx)$ , nếu  $h'hx \neq hx$  và  $h'$  dương tính đối với  $h$  (hoặc  $c$ , nếu  $h = I$  và  $x = c$ ).

**Định nghĩa 2.3.** Giả sử  $X$  là một ĐSGT tuyến tính đầy đủ,  $fm(x)$  và  $\mu(h)$  tương ứng là các độ đo tính mờ của ngôn ngữ  $x$  và của gia tử  $h$  thỏa mãn các tính chất trong Mệnh đề 2.1. Khi đó, ta nói  $\nu$  là ánh xạ cảm sinh bởi độ đo tính mờ  $fm$  của ngôn ngữ nếu nó được xác định như sau:

$$(a) \nu(W) = \kappa = fm(c^-), \nu(c^-) = \kappa - \alpha fm(c^-) = \beta fm(c^-), \nu(c^+) = \kappa + \alpha fm(c^+);$$

$$(b) \nu(h_jx) = \nu(x) + Sgn(h_jx) \{ \sum_{i=Sgn(j)}^j \mu(h_i) fm(x) - \omega(h_jx) \mu(h_j) fm(x) \},$$

trong đó

$$\omega(h_jx) = \frac{1}{2} [1 + Sgn(h_jx) Sgn(h_p h_jx) (\beta - \alpha)] \in \{\alpha, \beta\} \text{ với mọi } j, -q \leq j \leq p \text{ và } j \neq 0;$$

(c)  $\nu(\Phi c^-) = 0$ ,  $\nu(\Sigma c^-) = \kappa = \nu(\Phi c^+)$ ,  $\nu(\Sigma c^+) = 1$ , và với mọi  $j, -q \leq j \leq p$  và  $j \neq 0$ , ta có

$$\nu(\Phi h_jx) = \nu(x) + Sgn(h_jx) \{ \sum_{i=Sgn(j)}^{j-1} \mu(h_i) fm(x) \}$$

và

$$\nu(\Sigma h_jx) = \nu(x) + Sgn(h_jx) \{ \sum_{i=Sgn(j)}^j \mu(h_i) fm(x) \}.$$

## 2.2. Đối tượng mờ

Các thực thể trong thế giới thực hay các khái niệm trừu tượng thường là các đối tượng phức tạp. Các đối tượng này chứa một tập nhất định các thông tin về đối tượng và các hành vi dựa trên các thông tin đó. Thông tin về đối tượng được gọi là thuộc tính đối tượng và được xác định bởi giá trị cụ thể, giá trị này có thể là giá trị rõ hoặc vì một lý do nào đó mà ta không xác định được giá trị chính xác của nó. Chẳng hạn, thuộc tính tuổi của một đối tượng được cho là “khoảng 18”, “từ 20 đến 22”, hoặc có thể là một giá trị ngôn ngữ “rất trẻ”. Hoặc trong một ngữ cảnh khác, thuộc tính lương của một đối tượng là “cao”, “khả năng thấp”, “khoảng 5.000.000”, ... Những thông tin không chính xác, không rõ ràng như vậy gọi là thông tin mờ. Như vậy, một đối tượng là mờ vì có một hoặc nhiều thuộc tính có chứa thông tin mờ (gọi là thuộc tính mờ). Không mất tính tổng quát, về mặt hình thức, đối tượng có ít nhất một thuộc tính mờ gọi là đối tượng mờ.

## 2.3. Lớp mờ

Các đối tượng có những thuộc tính giống nhau được đưa vào các lớp và tổ chức thành hệ thống phân cấp. Về mặt lý thuyết, một lớp có thể được xem xét từ hai quan điểm khác nhau:

thứ nhất, một lớp mở rộng được định nghĩa bởi danh sách các đối tượng của nó. Thứ hai, một lớp khái niệm, được xác định bởi một tập các thuộc tính và các giá trị chấp nhận được của nó (và các phương thức để thao tác trên các thuộc tính này). Ngoài ra, một lớp con được xác định từ lớp cha bằng cách thừa kế trong cơ sở dữ liệu hướng đối tượng (CSDL HDT) có thể được xem như là trường hợp đặc biệt của trường hợp thứ hai.

Vì vậy, một lớp được xem là mờ vì những lý do sau: Thứ nhất, một số đối tượng của một lớp được xác định là đối tượng mờ, khi đó, những đối tượng này thuộc về lớp với độ thuộc nhất định. Thứ hai, khi một lớp được định nghĩa, miền trị của một thuộc tính nào đó có thể là mờ và như vậy một lớp mờ được hình thành. Ví dụ, một lớp *Hình ảnh* là mờ vì miền giá trị thuộc tính *năm* của nó sử dụng yếu tố thời gian là một tập hợp các giá trị mờ như *lâu*, *rất lâu* và *khoảng 50 năm*. Thứ ba, một lớp con được kế thừa một hoặc nhiều lớp cha, trong đó có ít nhất một lớp cha là lớp mờ.

Sự khác biệt chính giữa các lớp mờ và các lớp rõ đó là ranh giới của các lớp mờ không rõ ràng. Sự thiếu chính xác trong ranh giới giữa các lớp mờ là do sự mơ hồ của những giá trị trong miền trị thuộc tính. Trong CSDL HDT mờ, các lớp là mờ vì miền trị thuộc tính của chúng chứa các giá trị mờ. Một đối tượng mờ thuộc về một lớp xảy ra vì lớp hoặc đối tượng đó có thể là mờ. Tương tự như vậy, một lớp là lớp con của một lớp khác với độ thuộc  $k (k \in Z^+)$  nào đó vì đó là lớp mờ. Do vậy, các đánh giá của mối quan hệ lớp đối tượng mờ và phân cấp thừa kế mờ là quan trọng của mô hình CSDL HDT mờ.

#### 2.4. Quan hệ lớp đối tượng mờ

Theo [6], trong CSDL HDT mờ, 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 (b), các đối tượng có thể thuộc về lớp với mức độ thuộc  $k$ . 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 mức độ thuộc  $k$ . 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. Rõ ràng sự đánh giá mức độ thành viên không chắc chắn của các đối tượng vào lớp là rất quan trọng trong quan hệ lớp đối tượng mờ.

Theo [1], đối với mỗi giá trị ngôn ngữ mờ  $x$ , 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  $0 < |x| \leq k^*$ ,  $\forall x \in \mathbf{X}$ . Với bất kỳ  $x \in \mathbf{X}$ , 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:

- Trường hợp  $k = j$  :  $O_{\min,k}(x) = I(h_{-1}x) \cup I(h_1x)$ .
- Trường hợp  $1 \leq k < j$  :  $O_{\min,k}(x) = I(x)$ .
- Trường hợp  $j+1 \leq k \leq k^*$  :  $O_{\min,k}(x) = I(h_ly) \cup I(h_{l'}y')$ , với  $l, l' \in \{-q, p\}$ ,  $y, y' \in H(x)$ ,

$$|h_ly| = |h_{l'}y'| = k + 1.$$

Từ đó, ta thống nhất cách biểu diễn dữ liệu ngôn ngữ mờ theo định nghĩa sau.

**Định nghĩa 2.4.** [1] Cho  $x \in X \cup C$ , một biểu diễn khoảng của  $x$  là một tập  $IRp(x)$  các khoảng được xác định:  $IRp(x) = \{O_{\min,k}(x) | 1 \leq k \leq n\}$ .

Cách biểu diễn dữ liệu ngôn ngữ mờ như trên có thể sử dụng để biểu diễn các dạng dữ liệu khác. Đối với giá trị số, đây là loại dữ liệu rõ, độ mờ của dữ liệu bằng 0, khi đó mỗi giá trị số  $a$  được biểu diễn bằng  $[a, a]$ , và  $O_{\min,k}(a) = \{[a, a]\}$ , với mọi  $1 \leq k \leq k^*$  và  $IRp(a) = \{[a, a]\}$ . Còn mỗi giá trị khoảng  $a$  được biểu diễn bằng  $[a - \varepsilon, a + \varepsilon]$ , với  $\varepsilon$  được xem là bán kính với tâm  $a$ . Vì  $[a - \varepsilon, a + \varepsilon]$  là dữ liệu rõ nên  $O_{\min,k}([a - \varepsilon, a + \varepsilon]) = \{[a - \varepsilon, a + \varepsilon]\}$ , với mọi  $1 \leq k \leq k^*$  và  $IRp([a - \varepsilon, a + \varepsilon]) = \{[a - \varepsilon, a + \varepsilon]\}$ .

Quan hệ gần nhau sẽ được xây dựng dựa trên các khoảng mờ của các phần tử trong  $\mathbf{X}$ . Chúng là một cơ sở tôpô ngữ nghĩa trên miền trị tham chiếu của thuộc tính mờ  $A$ . Giá trị thuộc tính  $A$  của đối tượng  $o$  ký hiệu  $o(A)$ . Vì tập các khoảng mờ thuộc  $P_k$  là một phân hoạch trên miền trị thuộc tính mờ nên nó xác định một quan hệ tương đương với các lớp tương đương là các khoảng mờ này  $A$ . Các giá trị nằm trong cùng khoảng sẽ được coi là gần nhau mức  $k$ . Tuy nhiên, như đã phân tích ở phần trên, khi  $x$  có độ dài bé hơn  $k$  thì giá trị  $\nu(x)$  là điểm đầu mút của một lớp tương đương  $I(u)$  trong  $P_k$ . Điều này dẫn đến có những giá trị trong lân cận của  $x$  lại không tương tự mức  $k$ . Vì vậy, ta sẽ xây dựng một phân hoạch khác sao cho  $\nu(x)$  là điểm tôpô của phân hoạch với mọi  $x, |x| \leq k$ , như sau:

Xét  $\mathbf{X}$  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:

$$H_1 = \{h_i, h_{-j} | 1 \leq i \leq [p/2], 1 \leq j \leq [q/2]\},$$

$$H_2 = \{h_i, h_{-j} | [p/2] \leq i \leq p, [q/2] \leq j \leq q\}.$$

Đặt  $P_{k+1}(H_n) = \{I(h_iy) | y \in X_k, h_i \in H_n\}$ , 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  $y \in \mathbf{X}_k$ ,  $P_{k+1}(H_1)$  được phân thành các cụm có dạng  $\{I(h_iy) | h_i \in H_1\}$ . Hơn nữa, do  $I(h_{-1}y) \leq \nu(y) \leq I(h_1y)$  hoặc là  $I(h_1y) \leq \nu(y) \leq I(h_{-1}y)$  nên bao giờ ta cũng có  $\nu(y) \in \{I(h_iy) | h_i \in H_1\}$ .

Bây giờ ta phân cụm các khoảng mờ của  $P_{k+1}(H_2)$ . Giả sử  $\mathbf{X}_k = \{x_s | s = 0, \dots, m-1\}$  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 sau đây:

- (a) Cụm nằm bên trái  $x_0$  :  $\{I(h_ix_0) | h_i \in H_2^+\}$ .

(b) Cụm nằm bên phải  $x_{m-1} : \{I(h_i x_{m-1}) | h_i \in H_2^+\}$ .

(c) Các 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:

(c1)  $\mathbf{P} = \{I(h_i x_s), I(h'_j x_{s+1}) | h_i \in H_2^+, h'_j \in H_2^-\}$ ,

nếu  $Sgn(h_p x_s) = +1$  và  $Sgn(h_p x_{s+1}) = +1$ .

(c2)  $\mathbf{P} = \{I(h_i x_s), I(h'_j x_{s+1}) | h_i \in H_2^+, h'_j \in H_2^+\}$ ,

nếu  $Sgn(h_p x_s) = +1$  và  $Sgn(h_p x_{s+1}) = -1$ .

(c3)  $\mathbf{P} = \{I(h_i x_s), I(h'_j x_{s+1}) | h_i \in H_2^-, h'_j \in H_2^-\}$ ,

nếu  $Sgn(h_p x_s) = -1$  và  $Sgn(h_p x_{s+1}) = +1$ .

(c4)  $\mathbf{P} = \{I(h_i x_s), I(h'_j x_{s+1}) | h_i \in H_2^-, h'_j \in H_2^+\}$ ,

nếu  $Sgn(h_p x_s) = -1$  và  $Sgn(h_p x_{s+1}) = -1$ .

Tập tất cả các cụm được ký hiệu là  $\mathbf{C}$  và ta định nghĩa khoảng tương tự mức  $k$  như sau:

**Định nghĩa 2.5.** [1] Mỗi  $\mathbf{P}$  thuộc  $\mathbf{C}$ , ta gọi khoảng tương tự mức  $k$  ứng với  $\mathbf{P}$  là  $S_k(\mathbf{P}) = \cup \{I(u) | I(u) \in \mathbf{P}\}$ .

Với cách định nghĩa này, mỗi khoảng  $S_k(\mathbf{P})$  sẽ không quá lớn để phủ bất kỳ một khoảng  $I(u)$  thuộc  $P_k$  nhưng lại không quá nhỏ để nằm gọn trong một khoảng  $I(u)$  thuộc  $P_{k+1}$  nào. Vì  $\{S_k(\mathbf{P}) | \mathbf{P} \in \mathbf{C}\}$  là một phân hoạch trên miền trị tham chiếu nên nó xác định một quan hệ tương đương và 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  $\mathbf{P}$  sao cho  $\nu(x) \in S_k(\mathbf{P})$ . Vì vậy, ta có thể định nghĩa  $S_k(x) = S_k(\mathbf{P})$ .

**Mệnh đề 2.2.** [1] Cho  $\underline{\mathbf{X}}$  là ĐSGT tuyến tính đầy đủ, trong đó  $H^+$  và  $H^-$  có ít nhất hai phần tử. Khi đó:

(a) Với mỗi  $k, \{S_k(u) | u \in \mathbf{X} \cup C\}$  được xác định duy nhất và là một phân hoạch của đoạn  $[0, 1]$ .

(b) Với mọi  $x, u \in \mathbf{X} \cup C$ , nếu  $\nu(x) \in S_k(u)$  thì lân cận bé nhất mức  $k$  của  $x$  nằm trong  $S_k(u)$ , tức là  $O_{\min, k}(x) \in S_k(u)$ .

**Định nghĩa 2.6.** Cho một đối tượng bất kỳ  $o$  trên tập thuộc tính  $\{A_1, A_2, \dots, A_n\}$  của lớp  $C, \underline{\mathbf{X}}$  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  $u \in \mathbf{X}$ , 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 2.7.** Cho hai đối tượng bất kỳ  $o_1, o_2$  trên tập thuộc tính  $\{A_1, A_2, \dots, A_n\}$  của lớp  $C, \underline{\mathbf{X}}$  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:

(a) 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  $O_{\min, k}(o_2(A_i)) \in S_k(u)$ .

(b) 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  $O_{\min,k}(o_2(A_i)) \in S_k(u)$ .

**Bổ đề 2.1.** [1] *Quan hệ bằng nhau theo mức  $k(=_k)$  là một quan hệ tương đương.*

**Hệ quả 2.1.** *Cho  $o_1, o_2$  là hai đối tượng bất kỳ trên tập thuộc tính  $\{A_1, A_2, \dots, A_n\}$  của lớp  $C$ ,  $S_k$  là quan hệ tương tự mức  $k(0 < k \leq k^*)$  trên miền trị thuộc tính  $A_i$  của lớp  $C$ ,*

(a) *Nếu  $o_1(A_i) =_k o_2(A_i)$  thì  $o_1(A_i) =_{k'} o_2(A_i), \forall k' < k$ ;*

(b) *Nếu  $o_1(A_i) \neq_k o_2(A_i)$  thì  $o_1(A_i) \neq_{k'} o_2(A_i), \forall k' > k$ .*

### 3. MÔ HÌNH CƠ SỞ HƯỚNG ĐỐI TƯỢNG MỜ VÀ MỘT SỐ THAO TÁC

#### 3.1. Định nghĩa lớp mờ

Các lớp trong CSDL HDT mờ có thể mờ. Theo đó, một đối tượng thuộc một lớp tùy theo mức  $k$  và một lớp là lớp con của một lớp khác cũng theo mức  $k(k \in Z^+)$ . Trong CSDL HDT, một lớp được định nghĩa bao gồm mối quan hệ kế thừa, thuộc tính và phương thức. Để xác định một lớp mờ, cần thiết bổ sung thêm một số định nghĩa, khi khai báo mối quan hệ kế thừa cần chỉ ra mức mà lớp này là lớp con của lớp cha. Và trong định nghĩa của một lớp mờ, các thuộc tính mờ có thể được chỉ ra một cách rõ ràng. Về mặt hình thức, định nghĩa của một lớp mờ được thể hiện như sau:

```

CLASS tên lớp
  INHERITES
    tên lớp cha thứ 1 WITH LEVEL OF mức_1
    ...
    tên lớp cha thứ n WITH LEVEL OF mức_n
  ATTRIBUTES
    tên thuộc tính thứ 1: [FUZZY] DOMAIN dom_1: TYPE OF kiểu_1
    ...
    tên thuộc tính thứ m: [FUZZY] DOMAIN dom_m: TYPE OF kiểu_m
  METHODS
    ...
END

```

**Ví dụ 3.1.** Cho một lớp “Nhân viên trẻ” như sau:

```

Class NhanVienTre {
  Oid: allID
  Họ tên : string
  Tuổi : [fuzzy] domain [18 .. 60]: int
  Quê quán : string
  Hệ số lương: [fuzzy] domain [0..7,5]: float
  Số lượng sản phẩm: [fuzzy] domain [0..30]: int
}

```

Cho các đối tượng trên tập thuộc tính của lớp “Nhân viên trẻ”:

- $o_1(oid_1, \text{Hải}, 27, \text{Huế}, 2.67, 15),$
- $o_2(oid_2, \text{Nam}, \text{“khoảng 30”}, \text{Phú Yên}, \text{“ít thấp”}, \text{“rất cao”}),$
- $o_3(oid_3, \text{thái}, \text{“khá trẻ”}, \text{Cần Thơ}, \text{“khả năng ít thấp”}, \text{“khả năng cao”}),$
- $o_4(oid_4, \text{Quốc}, \text{“ít khá trẻ”}, \text{Hà Nội}, \text{“khoảng 3.0”}, \text{“khoảng 17”}).$

Khi cần xác định mức độ thuộc của các đối tượng vào lớp thì ta chỉ cần xác định mức trong quan hệ bằng nhau của thuộc tính tuổi.

Trước hết, ta sẽ xem miền trị của thuộc tính “Tuổi” là một đại số gia tử và được xác định như sau:  $G = \{ \text{trẻ}, \text{già} \}, H^- = \{ \text{gần}, \text{ít} \}$  và  $H^+ = \{ \text{khá}, \text{rất} \}$ . Các tham số mờ:  $fm(\text{trẻ}) = 0.42; fm(\text{già}) = 0.58; \mu(\text{rất}) = 0.2; \mu(\text{khá}) = 0.28; \mu(\text{gần}) = 0.27; \mu(\text{ít}) = 0.25$ . Giả sử  $k^* = 3$ .

Giả sử miền trị tham chiếu của biến Tuổi của những người đang công tác là  $cdom(\text{Tuổi}) = [18, 60]$ , nên ta sẽ dùng hệ số  $r = 42$  để chuyển đổi từ  $[0,1]$  qua  $[18,60]$ , các kí hiệu có kèm  $r$  chỉ cho sự chuyển đổi này.

Với  $k = 1$ , ta có:  $O_{\min,1}(\text{khá trẻ}) = I_r(\text{khá trẻ}) = (21.5280, 26.4672],$   
 $O_{\min,1}(\text{ít khá trẻ}) = I_r(\text{ít khá trẻ}) = (25.2324, 26.4672].$

Với  $k = 2$ , ta có:  $O_{\min,2}(\text{ít khá trẻ}) = I_r(\text{ít khá trẻ}) = (25.2324, 26.4672]$   
 $O_{\min,2}(\text{khá trẻ}) = I_r(\text{khá khá trẻ}) \cup I_r(\text{gần khá trẻ}) = (22.51584, 25.2324]$

Với  $k = 3$ , ta có:  
 $O_{\min,3}(\text{khá trẻ}) = I_r(\text{ít khá khá trẻ}) \cup I_r(\text{rất gần khá trẻ}) = (23.55308, 24.16554].$   
 $O_{\min,3}(\text{ít khá trẻ}) = I_r(\text{khá ít khá trẻ}) \cup I_r(\text{gần ít khá trẻ}) = (25.4794, 26.1585].$

Như vậy,  $IRp(\text{khá trẻ}) = \{(21.5280, 26.4672], (22.51584, 25.2324], (23.55308, 24.23222]\}$   
 $IRp(\text{ít khá trẻ}) = \{(25.2324, 26.4672], (25.2324, 26.4672], (25.4794, 26.1585]\}$

Đối với giá trị thuộc tính đối tượng  $o_1(\text{Tuổi}) = 26$ , ta có  $O_{\min,k}(27) = [27, 27], \forall 1 \leq k \leq k^*$  và  $IRp(27) = \{[27, 27]\}$ . Còn  $o_2(\text{Tuổi}) = \text{‘khoảng 30’}$  được biểu diễn bằng khoảng  $[29, 31]$ , nên  $O_{\min,k}([29, 31]) = [29, 31], \forall 1 \leq k \leq k^*$  và  $IRp([29, 31]) = \{[29, 31]\}$ .

Với  $u = \text{“trẻ”}$ , ta có các lớp tương đương  $S_k(\text{trẻ})$  của quan hệ tương tự  $S_k$  như sau:

$S_{1,r}(\text{trẻ}) = I_r(\text{khá trẻ}) \cup I_r(\text{gần trẻ}) = (21.5280, 26.4672] \cup (26.4672, 31.23] = (21.5280, 31.23]$   
 $S_{2,r}(\text{trẻ}) = I_r(\text{ít khá trẻ}) \cup I_r(\text{rất gần trẻ}) = (25.2324, 27.91476].$   
 $S_{3,r}(\text{trẻ}) = I_r(\text{ít ít khá trẻ}) \cup I_r(\text{rất rất gần trẻ}) = (26.1585, 26.2767]$

Như vậy, sử dụng các định nghĩa trên ta được các đối tượng thuộc vào lớp “Nhân viên trẻ” như sau:

Khi  $k = 1$ , các đối tượng  $o_1, o_2, o_3$  và  $o_4$  đều thuộc lớp “Nhân viên trẻ” vì:

$O_{\min,1}(o_1(\text{tuổi})) = [27, 27] \subseteq S_{1,r}(\text{trẻ}), O_{\min,1}(o_2(\text{tuổi})) = [29, 31] \subseteq S_{1,r}(\text{trẻ}),$

$O_{\min,1}(o_3(\text{tuổi})) = (21.5280, 26.4672] \subseteq S_{1,r}(\text{trẻ}),$  và  $O_{\min,1}(o_4(\text{tuổi})) = (25.2324, 26.4672] \subseteq S_{1,r}(\text{trẻ}).$

Khi  $k = 2$ , chỉ có các đối tượng  $o_1$  và  $o_4$  thuộc lớp “Nhân viên trẻ” vì  $O_{\min,2}(o_1(\text{tuổi})) = [27, 27] \subseteq S_{2,r}(\text{trẻ}), O_{\min,2}(o_4(\text{tuổi})) = (25.2324, 26.4672] \subseteq S_{2,r}(\text{trẻ}).$

Và khi  $k = 3$ , không có đối tượng nào thuộc lớp “Nhân viên trẻ”.



### 3.2. Một số phép toán đại số quan hệ

Do xây dựng mô hình cơ sở dữ liệu theo ngữ nghĩa mới nên các thao tác trong mô hình này cần được nghiên cứu. Bài báo đề xuất các phép toán đại số cơ bản: phép chọn mờ ( $\tilde{\delta}$ ), phép chiếu mờ ( $\tilde{\Pi}$ ), phép tích mờ ( $\tilde{\times}$ ), kết nối mờ ( $\tilde{\bowtie}$ ) và phép hợp mờ ( $\tilde{\cup}$ ). Cho  $C_1$  và  $C_2$  là các lớp mờ,  $Attr(C_1)$  và  $Attr(C_2)$  là các tập thuộc tính tương ứng của chúng. Giả sử một lớp mới  $C$  được tạo ra bằng cách kết hợp  $C_1$  và  $C_2$ . Khi đó,

$$\begin{aligned} C &= C_1 \tilde{\times} C_2, \text{ nếu } Attr(C_1) \cap Attr(C_2) = \phi, \\ C &= C_1 \tilde{\bowtie} C_2, \text{ nếu } Attr(C_1) \cap Attr(C_2) \neq \phi \text{ và } Attr(C_1) \neq Attr(C_2), \\ C &= C_1 \tilde{\cup} C_2, \text{ nếu } Attr(C_1) = Attr(C_2). \end{aligned}$$

#### 3.2.1. Phép chọn mờ ( $\tilde{\delta}$ )

Cho lớp  $C$  gồm tập các thuộc tính  $\{A_1, A_2, \dots, A_n\}$ ,  $f = (A_i =_k fvalue_i)$  là một biểu thức điều kiện mà mỗi đối tượng của lớp  $C$  có thể thỏa mãn hoặc không. Khi đó, phép chọn trên  $C$  các đối tượng thỏa mãn điều kiện  $f$ , ký hiệu  $\tilde{\delta}_f(C)$ , được xác định như sau:

$$\tilde{\delta}_f(C) = \{o | o(A_i) =_k fvalue_i\}$$

#### 3.2.2. Phép chiếu mờ ( $\tilde{\Pi}$ )

Cho lớp  $C$  gồm tập thuộc tính  $Attr(C) = \{A_1, A_2, \dots, A_n\}$  và tập các đối tượng  $O(Attr(C)) = \{o_1, o_2, \dots, o_m\}$ ,  $X \subseteq Attr(C)$ ,  $\alpha \in Z^+$  là mức cho trước. Khi đó, phép chiếu trên tập  $X$  của lớp  $C$ , ký hiệu  $\tilde{\alpha}_X(C)$ , sẽ cho ra một lớp mới  $C'$  có tập thuộc tính là  $X$  và các đối tượng được xác định như sau:

$$C' = \tilde{\prod}_{\alpha X} (C) = \{o_i | o_i \in O(X) \wedge \forall o_i, o_j \in C', i \neq j, \forall A_t \in X : o_i(A_t) \neq_{\alpha} o_j(A_t)\}.$$

#### 3.2.3. Phép hợp mờ ( $\tilde{\cup}$ )

Phép hợp mờ của  $C_1$  và  $C_2$  sẽ cho ra một lớp mới  $C$ . Khi đó các đối tượng của lớp  $C$  gồm ba loại đối tượng như sau: Hai loại đầu tiên là các đối tượng trực tiếp đến từ các lớp thành phần (lớp  $C_1, C_2$ ) khác nhau theo mức cho trước. Loại thứ ba bao gồm các đối tượng là kết quả của việc kết hợp các đối tượng còn lại từ hai lớp bằng nhau theo mức cho trước. Với  $\alpha \in Z^+$  là mức cho trước, ta có:

$$\begin{aligned} C &= C_1 \tilde{\cup}_{\alpha} C_2 = \{o | (\forall o'' \in C_2 \wedge o \in C_1 : o(C_1) \neq_{\alpha} o''(C_2)) \vee (\forall o' \in C_1 \wedge o \in C_2 : ((o(C_2) \neq_{\alpha} \\ & o''(C_1))) \\ & \vee (\exists o' \in C_1 \wedge \exists o'' \in C_2 : o'(C_1) =_{\alpha} o''(C_2) \wedge o = \Psi(o', o''))\}. \end{aligned}$$

Ở đây,  $\Psi$  là một thao tác hợp nhất hai đối tượng còn lại để tạo thành một đối tượng mới của lớp. Cho  $o'$  và  $o''$  là hai đối tượng của lớp  $C_1$  và  $C_2$ ,  $o$  là đối tượng mới của lớp  $C$ ,  $o = \Psi(o', o'')$ . Khi đó,  $o(C) = o'(C_1)$  hoặc  $o(C) = o''(C_2)$ .

#### 3.2.4. Phép tích mờ ( $\tilde{\times}$ )

Phép tích mờ của hai lớp  $C_1$  và  $C_2$  sẽ cho kết quả là một lớp mới  $C$  có các thuộc tính bao gồm các thuộc tính của  $C_1, C_2$  và *bổ sung thêm một thuộc tính định danh mới*. Các đối tượng

của lớp  $C$  được tạo ra bởi các đối tượng của lớp  $C_1$  và  $C_2$  như sau:

$$C = C_1 \tilde{\times} C_2 = \{o | \forall o' \in C_1 \wedge \forall o'' \in C_2 : o(Attr(C_1)) = o'(C_1) \wedge o(Attr(C_2)) = o''(C_2)\}.$$

Giả sử các đối tượng thuộc vào lớp  $C_1$  theo mức  $k_1$  và các đối tượng thuộc vào lớp  $C_2$  theo mức  $k_2$ . Khi đó, theo Bổ đề 2.1, các đối tượng này thuộc về lớp mới  $C$  với mức  $k = \min\{k_1, k_2\}$ .

### 3.2.5. Phép kết nối mờ ( $\tilde{\times}$ )

Cho hai lớp  $C_1$  và  $C_2$ , với  $Attr(C_1) \cap Attr(C_2) \neq \phi$  và  $Attr(C_1) \neq Attr(C_2)$ . Khi đó, phép kết nối mờ của  $C_1$  và  $C_2$  cho kết quả là một lớp mới  $C$ , có tập thuộc tính là  $Attr(C_1) \cup (Attr(C_2) - (Attr(C_1) \cap Attr(C_2)))$  và *bổ sung thêm một thuộc tính định danh mới*. Còn các đối tượng của  $C$  được tạo ra bởi các thành phần đối tượng từ  $C_1$  và  $C_2$ , trong đó, giá trị của các đối tượng trên các thuộc tính chung phải bằng nhau theo mức  $\alpha$  cho trước,  $\alpha \in Z^+$ . Khi đó,

$$C = C_1 \tilde{\times}_\alpha C_2 = \{o | \exists o' \in C_1 \wedge \exists o'' \in C_2 : \\ o'(Attr(C_1) \cap Attr(C_2)) =_\alpha o''(Attr(C_1) \cap Attr(C_2)) \\ \wedge o(Attr(C_1)) = o'(C_1) \\ \wedge o(Attr(C_2) - (Attr(C_1) \cap Attr(C_2))) = o''(Attr(C_2) \setminus (Attr(C_1) \cap Attr(C_2)))\}.$$

Các đối tượng này thuộc về lớp mới  $C$  với mức  $k = \min\{k_1, k_2\}$ .

Dựa vào Đại số quan hệ, người ta có thể xây dựng các biểu thức đại số quan hệ để trả lời các câu truy vấn.

**Ví dụ 3.2.** Cho biết oid, họ tên và tuổi của các đối tượng thuộc lớp “Nhân viên trẻ” có hệ số lương ít thấp với mức 2.

Sử dụng các phép toán đại số trên, ta có thể trả lời cho câu hỏi này:

$$\prod_2 \text{Oid, Họ tên, Tuổi}(\tilde{\delta} \text{Hệ số lương} = 2 \text{ “ít thấp” (Nhân viên trẻ)}).$$

Đối với thuộc tính Hệ số lương: gọi  $\underline{X}_{HSL}$  là một ĐSGT với  $G = \{ \text{thấp, cao} \}$ ,  $H^+ = \{ \text{rất, khá} \}$ ,  $H^- = \{ \text{khả năng, ít} \}$ ,  $\text{rất} > \text{khá}$ ,  $\text{ít} > \text{khả năng}$ . Chọn  $w = 0.4$ ,  $fm(\text{thấp}) = 0.4$ ;  $fm(\text{cao}) = 0.6$ ;  $\mu(\text{rất}) = 0.3$ ;  $\mu(\text{khá}) = 0.25$ ;  $fm(\text{khả năng}) = 0.2$ ;  $\mu(\text{ít}) = 0.25$ ;  $cdom_{HSL} = [0, 7.5]$ .

$$O_{\min,2}(2.67) = [2.67, 2.67];$$

$$O_{\min,2}(\text{ít thấp}) = I_r(\text{khả năng ít thấp}) \cup I_r(\text{khá ít thấp}) = (3.55, 3.963];$$

$$O_{\min,2}(\text{khả năng ít thấp}) = I_r(\text{khả năng ít thấp}) = (3.55, 3.775];$$

$$O_{\min,2}(\text{khoảng 3.0}) = [2.67, 3.33];$$

$$S_{2,HSL,r}(\text{ít thấp}) = I_r(\text{khả năng ít thấp}) \cup I_r(\text{khá ít thấp}) = (3.55, 3.963];$$

Kết quả nhận được:  $o_2(oid_2, \text{Nam, “khoảng 30”})$ ,  $o_3(oid_3, \text{Thái, “khá trẻ”})$ .

## 4. XỬ LÝ TRUY VẤN TRONG CƠ SỞ DỮ LIỆU

Xử lý truy vấn dựa trên đề xuất mô hình cơ sở dữ liệu hướng đối tượng mờ liên quan đến thao tác lựa chọn các đối tượng thuộc lớp theo một mức nhất định và đáp ứng các điều kiện truy vấn cũng theo một mức xác định. Như vậy, truy vấn trong cơ sở dữ liệu hướng đối tượng

mờ có liên quan đến sự lựa chọn các mức, và vì vậy, một câu lệnh truy vấn trong cơ sở dữ liệu hướng đối tượng mờ có cấu trúc như sau:

```
SELECT < Danh sách thuộc tính >
FROM < Lớpp1 WITH mứck1, ..., Lớppm WITH mứckm >
WHERE < Điều_kiện_truy_vấn WITH mứck >
```

Ở đây, <điều kiện truy vấn> là một điều kiện mờ hoặc liên kết các điều kiện mờ có sử dụng các phép toán tuyển và hội. Tất cả các mức thuộc số nguyên dương. Sử dụng câu lệnh truy vấn hướng đối tượng mờ, chúng ta có thể lấy được các đối tượng thuộc về các lớp theo các mức cho trước và đồng thời đáp ứng các điều kiện truy vấn theo mức cho trước.

Theo phân tích trên ở mục 2.4 về mối quan hệ lớp đối tượng mờ, ta đã xác định được các đối tượng thuộc vào lớp theo mức xác định. Như vậy, vấn đề quan trọng còn lại trong câu truy vấn hướng đối tượng mờ chính là xác định các đối tượng thỏa điều kiện mờ theo mức cho trước.

**Thuật toán.** Xử lý truy vấn hướng đối tượng mờ

Vào:

- Lớp  $C$  cùng với các thuộc tính  $\{A_1, A_2, \dots, A_n\}$ , tập các đối tượng thuộc lớp  $C : \{o_t, t = 1, \dots, m\}$ .

- Câu truy vấn dạng select... from... where ( $A_i =_k \text{fvalue}_i \varphi A_j =_k \text{fvalue}_j$ ), trong đó, fvalue là giá trị mờ,  $\varphi$  là phép toán hội (and) hoặc tuyển (or).

Ra: Tập các đối tượng  $O = \{o_t : o_t(A_i) =_k \text{fvalue}_i \varphi o_t(A_j) =_k \text{fvalue}_j\}$

Phương pháp:

(1) Xây dựng các ĐSGT cho các thuộc tính có trong điều kiện truy vấn:

- Xác định  $G_{A_i}, H_{A_i}$ , chọn độ đo tính mờ cho các phần tử sinh và gia tử
- Xác định  $G_{A_j}, H_{A_j}$ , chọn độ đo tính mờ cho các phần tử sinh và gia tử

(2) Xác định miền trị kinh điển:  $D_{A_i} = [\min_{A_i}, \max_{A_i}]$  và  $D_{A_j} = [\min_{A_j}, \max_{A_j}]$

(3)  $O = \phi$ .

(4) Xây dựng lân cận tối thiểu mức  $k$  của giá trị thuộc tính  $A_i$  và  $A_j$  của mỗi đối tượng, ký hiệu là  $O_{\min, k}(o_t(A_i)), O_{\min, k}(o_t(A_j))$ .

(5) Xây dựng lớp tương đương  $S_k(\text{fvalue}_i)$  và  $S_k(\text{fvalue}_j)$

Duyệt lần lượt các đối tượng ban đầu của lớp để tìm các đối tượng thỏa điều kiện mờ:

(6) For each  $o_t (t = 1, \dots, m)$  do

(7) If  $O_{\min, k}(o_t(A_i)) \subseteq S_k(\text{fvalue}_i) \varphi O_{\min, k}(o_t(A_j)) \subseteq S_k(\text{fvalue}_j)$  then

(8)  $O = O \cup o_t$

(9) Return  $O$

**Ví dụ 4.1.** Cho các đối tượng của lớp “Nhân viên trẻ” theo Ví dụ 3.1:

Yêu cầu truy vấn mờ:

“Tìm những nhân viên thuộc lớp Nhân viên trẻ với mức  $\alpha = 1$ , có hệ số lương ít thấp và số lượng sản phẩm khả năng cao với mức  $k = 1$ ”.

Theo Ví dụ 3.1, các đối tượng của lớp “Nhân viên trẻ” đã cho đều thuộc lớp theo mức 1. Bây giờ chúng ta sẽ dựa vào thuật toán trên để xác định các đối tượng thỏa điều kiện truy vấn mờ.

Trước hết, ta xem miền trị của thuộc tính “Tuổi”, “Hệ số lương” và “Số lượng sản phẩm” là các đại số gia tử được xác định như các ví dụ trên:

Đối với thuộc tính Hệ số lương:

$$O_{\min,k}(\text{khoảng } 3.0) = [2.67, 3.33]; O_{\min,k}(2.67) = [2.67, 2.67], \text{ với mọi } k \leq k^*$$

$$O_{\min,1}(\text{khả năng thấp}) = I_r(\text{khả năng thấp}) = (2.35, 3.25]$$

$$O_{\min,1}(\text{ít thấp}) = I_r(\text{ít thấp}) = (3.25, 4]$$

$$O_{\min,1}(\text{khả năng ít thấp}) = I_r(\text{khả năng ít thấp}) = (3.55, 3.775]$$

$$S_{1,r}(\text{ít thấp}) = I_r(\text{rất thấp}) \cup I_r(\text{ít thấp}) = [0.4];$$

Đối với thuộc tính Số lượng sản phẩm, Gọi  $\underline{X}_{SLSP}$  là một ĐSGT của thuộc tính số lượng sản phẩm (SLSP), với  $G = \{\text{thấp, cao}\}$ ,  $H^+ = \{\text{rất, khá}\}$ ,  $H^- = \{\text{khả năng, ít}\}$ ,  $\text{rất} > \text{khá}$ ,  $\text{ít} > \text{khả năng}$ . Chọn  $w = 0.4$ ,  $f_m(\text{thấp}) = 0.4$ ;  $f_m(\text{cao}) = 0.6$ ;  $\mu(\text{rất}) = 0.2$ ;  $\mu(\text{khá}) = 0.3$ ;  $\mu(\text{khả năng}) = 0.3$ ;  $\mu(\text{ít}) = 0.2$ . Cho  $\text{Dom}(\text{Số lượng SP}) = [0, 30]$ , nên ta sẽ dùng hệ số  $r = 30$  để chuyển đổi từ  $[0, 1]$  qua  $[0, 30]$ .

$$O_{\min,k}(\text{khoảng } 17) = [16, 18] \text{ và } O_{\min,k}(15) = [15, 15], \text{ với mọi } k \leq k^*.$$

$$O_{\min,1}(\text{rất cao}) = I_r(\text{rất cao}) = (26.4, 30];$$

$$O_{\min,1}(\text{khả năng cao}) = I_r(\text{khả năng cao}) = (15.6, 21].$$

$$S_{1,SLSP,r}(\text{khả năng cao}) = I_r(\text{khả năng cao}) \cup I_r(\text{khá cao}) = (15.6, 21] \cup (21, 26.4] = (15.6, 26.4]$$

Duyệt lần lượt các đối tượng như bước (6), ta được các đối tượng thỏa điều kiện truy vấn:

$o_2(o_{id_2}$ , Nam, “khoảng 30”, Phú Yên, “ít thấp”, “rất cao”),

$o_3(o_{id_3}$ , thái, “khá trẻ”, Cần Thơ, “khả năng ít thấp”, “khả năng cao”),

$o_4(o_{id_4}$ , Quốc, “ít khá trẻ”, Hà Nội, “khoảng 3.0”, “khoảng 17”).

**Ví dụ 4.2.** Nếu mức truy vấn trên được sửa lại thành  $k = 2$ , khi đó:

Đối với thuộc tính Hệ số lương:

$$O_{\min,2}(\text{khả năng thấp}) = I_r(\text{khả năng khả năng thấp}) \cup I_r(\text{khá khả năng thấp}) = (2.62, 2.845] \cup (2.845, 3.115] = (2.62, 3.115]$$

$$O_{\min,2}(\text{ít thấp}) = I_r(\text{khả năng ít thấp}) \cup I_r(\text{khá ít thấp}) = (3.55, 3.963]$$

$$O_{\min,2}(\text{khả năng ít thấp}) = I_r(\text{khả năng ít thấp}) = (3.55, 3.775]$$

$$S_{2,HSL,r}(\text{ít thấp}) = I_r(\text{khả năng ít thấp}) \cup I_r(\text{khá ít thấp}) = (3.55, 3.963].$$

Đối với thuộc tính Số lượng sản phẩm, ta có:

$$O_{\min,2}(\text{rất cao}) = I_r(\text{khả năng rất cao}) \cup I_r(\text{khá rất cao}) = (27.12, 28.2] \cup (28.2, 29.28] = (27.12, 29.28];$$

$$O_{\min,2}(\text{khả năng cao}) = I_r(\text{khả năng khả năng cao}) \cup I_r(\text{khá khả năng cao}) = (16.68, 18.3] \cup (18.3, 19.92] = (16.68, 19.92].$$

$$S_{2,SLSP,r}(\text{khả năng cao}) = I_r(\text{khả năng khả năng cao}) \cup I_r(\text{khá khả năng cao}) = (16.68, 18.3] \cup (18.3, 19.92] = (16.68, 19.92]$$

Các đối tượng thỏa điều kiện truy vấn chỉ còn:

$o_3(oid_3, \text{Thái}, \text{“khá trẻ”}, \text{Cần Thơ}, \text{“khả năng ít thấp”}, \text{“khả năng cao”})$

## 5. KẾT LUẬN

Bài báo đề xuất một mô hình mới cho cơ sở dữ liệu hướng đối tượng mờ dựa trên cấu trúc định lượng của ĐSGT. Mục đích chính là biểu diễn dữ liệu bằng tập các khoảng trên không gian tham chiếu của miền trị thuộc tính của đối tượng  $o$ . Các quan hệ đối sánh bằng nhau theo mức  $k$  được tương tự như trong cơ sở dữ liệu truyền thống. Một số phép toán đại số quan hệ được trình bày phù hợp với mô hình mới. Việc thực hiện truy vấn để tìm kiếm các đối tượng trong cơ sở dữ liệu hướng đối tượng mờ bước đầu được nghiên cứu. Các dạng truy vấn phức tạp như sử dụng truy vấn lồng nhau, sử dụng các lượng từ mờ và các dạng phụ thuộc dữ liệu sẽ được nghiên cứu trong những công trình sau.

## TÀI LIỆU THAM KHẢO

- [1] Nguyễn Cát Hồ, Lê Xuân Vinh, Nguyễn Công Hào, Thống nhất dữ liệu và xây dựng quan hệ tương tự trong cơ sở dữ liệu ngôn ngữ bằng đại số gia tử, *Tạp chí Tin học và Điều khiển học* **25** (4) (2009) 314–332.
- [2] N.C. Ho, A Topological completion of refined Hedge algebras and a model of fuzziness of linguistic terms, *Fuzzy Sets and Systems* **158** (4) (2007) 436–451.
- [3] S. Al-Hamouz and R. Biswas, Fuzzy functional dependencies in relational databases, intern. *J. of Computational Cognition* (<http://www.ijcc.us>) **Vol. 4** (1) (2006) 39–43.
- [4] T.K. Bhattarjee, A.K. Mazumdar, Axiomatisation of fuzzy multivalued dependencies in a fuzzy relational data model, *Fuzzy Sets and Systems* **96** (1998) 343–352.
- [5] Z. M. MA+ AND LI YAN, A Literature overview of fuzzy database models, *Journal of Information Science and Engineering* **24** (2008) 189–202.
- [6] Cristina-Maria Vladarean, Extending object-oriented databases for fuzzy information modeling, *ROMAI J.* **2** (1) (2006) 225–237.
- [7] Zongmin Ma, *Advances in fuzzy object-oriented databases: modeling and applications*, Published in the United States of America by Idea Group Publishing, 2005.
- [8] [8] R.De Caluwe, N. Van Gyseghem, V. Cross, Basic notions and rationale of the integration of uncertainty management and object-oriented databases, *Fuzzy and Uncertain Object-Oriented Databases* (1997) 1–20.
- [9] T.H. Cao, J.M. Rossiter, T.P. Martin and J.F. Baldwin, Inheritance and Recognition in Uncertain and Fuzzy Object-Oriented Models, *IEEE*, 2001 (2317–2322).

*Ngày nhận bài 10 - 8 - 2011*  
*Nhận lại sau sửa ngày 22 - 8 - 2012*