

TRUY VẤN TRONG CƠ SỞ DỮ LIỆU HƯỚNG ĐỐI TƯỢNG MỜ DỰA TRÊN NGỮ NGHĨA ĐẠI SỐ GIA TỬ*

Nguyễn Công Hào¹, Trương Thị Mỹ Lê²

¹ Trung tâm Công nghệ thông tin, Đại học Huế

² Trường Đại học Quang Trung, Qui Nhơn

Tóm tắt. Trong cơ sở dữ liệu hướng đối tượng mờ, ngôn ngữ truy vấn được các tác giả nghiên cứu với nhiều cách tiếp cận khác nhau như tiếp cận lý thuyết tập mờ, lý thuyết khả năng... Tuy nhiên, việc đối sánh và biểu diễn dữ liệu vẫn còn nhiều khó khăn. Vì vậy, trong bài báo này, chúng tôi đề xuất một phương pháp mới để xử lý truy vấn trong cơ sở dữ liệu hướng đối tượng mờ là dựa trên ngữ nghĩa đại số gia tử. Cuối cùng, ngữ truy vấn với lượng từ ngôn ngữ được xem xét, nghiên cứu.

1. Giới thiệu

Thông thường, các quan sát và hiểu biết về thế giới thực là không đầy đủ, và như một hệ quả tất yếu, dữ liệu được lưu trữ bên trong mô hình của nó thường được mô tả một cách không chắc chắn, không hoàn chỉnh, không chính xác và gọi chung là thông tin mờ. Trong những năm qua, việc xử lý thông tin mờ trong môi trường cơ sở dữ liệu (CSDL) đã được nhiều tác giả quan tâm nghiên cứu, đặc biệt là trong mô hình cơ sở dữ liệu quan hệ [5, 6], và chỉ có một số hạn chế các nhóm tác giả nghiên cứu các vấn đề của thông tin mờ trong mô hình CSDL hướng đối tượng [1, 2, 4, 7]. Trên thực tế, chính sự kết hợp thông tin mờ vào CSDL hướng đối tượng sẽ cho thấy thêm ưu điểm mô hình dữ liệu hướng đối tượng đó là khả năng biểu diễn ngữ nghĩa của nó trở nên phong phú hơn.

Một số mô hình CSDL hướng đối tượng mờ cơ bản đã được nghiên cứu dựa vào lý thuyết xác suất, lý thuyết tập mờ, lý thuyết khả năng, quan hệ tương tự, ... Tuy nhiên, trong các mô hình đề xuất đó còn có nhiều hạn chế trong biểu diễn và đối sánh dữ liệu. Một hướng tiếp cận mới để xây dựng mô hình cơ sở dữ liệu hướng đối tượng với thông tin mờ đó là dựa trên đại số gia tử (ĐSGT), một mô hình toán học về cấu trúc thứ tự ngữ nghĩa của miền giá trị của biến ngôn ngữ. Dựa trên mô hình được đề xuất đó, chúng tôi tiếp tục nghiên cứu một phương pháp mới xử lý truy vấn hướng đối tượng mờ và xem xét việc đưa lượng từ vào trong câu truy vấn nhằm làm cho thao tác tìm kiếm dữ liệu

* Nghiên cứu được tài trợ bởi Quỹ hỗ trợ phát triển KHCN Quốc gia Nafosted

trên mô hình này được mềm dẻo và phù hợp với thực tế. Bài báo gồm 4 phần. Phần 2 trình bày những kiến thức cơ bản về CSDL hướng đối tượng mờ và đại số gia tử, phần 3 trình bày các phương pháp xử lý truy vấn trong CSDL hướng đối tượng mờ, phần 4 là một số nhận xét kết luận cho bài báo.

2. Một số kiến thức cơ sở

2.1. Các lớp và 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 của chính đối tượng 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 (giá trị chính xác) hoặc vì một lý do nào đó mà ta không xác định được giá trị chính xác của nó, ví dụ, thuộc tính “tuổi” của một đối tượng được cho là *khoảng 18*, hoặc có thể là một giá trị ngôn ngữ “*rất trẻ*”, đây là những thông tin không chính xác, không rõ ràng, mà ta gọi chung là thông tin mờ. Như vậy, một đối tượng là mờ vì thiếu thông tin. Về mặt hình thức, các đối tượng có ít nhất một thuộc tính có giá trị là một tập mờ là các đối tượng mờ.

Các đối tượng có các thuộc tính giống nhau được đưa vào các lớp được tổ chức thành hệ thống phân cấp. Về mặt lý thuyết, một lớp có thể được xem xét từ hai quan điểm khác nhau: (a) lớp mở rộng, được định nghĩa bởi danh sách các đối tượng, và (b) lớp khái niệm, được xác định bởi một tập các thuộc tính và các giá trị của thuộc tính.

Ngoài ra, một lớp con được xác định từ lớp cha bằng cách thừa kế trong CSDL hướng đối tượng có thể được xem như là trường hợp đặc biệt (b) ở trên.

Vì vậy, một lớp được coi là mờ vì một số lý do sau đây: Trước tiên, một số đối tượng của một lớp được xác định có thể là mờ. Những đối tượng này thuộc về lớp với độ thuộc trong đoạn $[0,1]$. 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 *Picture* 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 *xưa*, *rất xưa* 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, thì lớp con này cũng là lớp mờ.

Sự khác biệt chính giữa các lớp mờ và các lớp rõ đó là ranh giới giữa 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 hướng đối tượng mờ, các lớp là mờ vì miền trị thuộc tính của chúng chứa các tập mờ. Vấn đề một đối tượng thuộc về một lớp với mức độ không chắc chắn 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 cũng với một mức độ không chắc chắn nào đó vì đó là lớp mờ. Các đánh giá của *mối quan hệ lớp đối tượng mờ* và *phân cấp thừa kế mờ* là cốt lõi của mô hình thông tin mờ trong CSDL hướng đối tượng mờ.

2.2. Các quan hệ trên các lớp đối tượng mờ

Trong CSDL hướng đối tượng, việc xác định một đối tượng thuộc về một lớp phụ thuộc vào giá trị thuộc tính của đối tượng có nằm trong miền trị thuộc tính tương ứng của lớp đó hay không. Tương tự, để đánh giá độ thuộc của một đối tượng vào một lớp trong quan hệ lớp đối tượng mờ, cần thiết phải đánh giá mức độ tương đương nhau giữa miền trị thuộc tính của lớp và giá trị thuộc tính của đối tượng.

Cho C là một lớp với các thuộc tính $\{A_1, A_2, \dots, A_n\}$, mỗi thuộc tính A_i có miền giá trị $dom(A_i)$, o là một đối tượng trên C , và $o(A_i)$ biểu thị giá trị thuộc tính của o trên thuộc tính $A_i, 1 \leq i \leq n$. Như ta đã biết, $dom(A_i)$ là tập giá trị rõ trong CSDL hướng đối tượng và có thể là một tập mờ trong CSDL mờ. Vì vậy, trong một mô hình CSDL hướng đối tượng thống nhất cho cả thông tin rõ và mờ, $dom(A_i)$ nên được hợp nhất của cả hai thành phần, $dom(A_i) = cdom(A_i) \cup fdom(A_i)$, với $cdom(A_i)$ và $fdom(A_i)$ tương ứng là các giá trị rõ và các tập con mờ. Tương tự như vậy, $o(A_i)$ có thể là một giá trị rõ hoặc một giá trị mờ. Đối với các thuộc tính rõ, các kiểu dữ liệu bao gồm các loại đơn giản như số nguyên, thực, logic, chuỗi, và các loại phức tạp như kiểu tập hợp và kiểu đối tượng. Đối với thuộc tính mờ, có thể nhận nhiều kiểu dữ liệu đơn giản hoặc phức tạp nhưng chúng ta có thể thống nhất chuyển về cùng kiểu dữ liệu [5].

Theo [5], đối với mỗi giá trị ngôn ngữ mờ x , chúng ta sẽ định nghĩa một biểu diễn khoảng cho x . Xét ĐSGT đầy đủ, tuyến tính $\underline{X} = (X, G, C, H, \Sigma, \Phi, \leq)$, với $H = \{h_{-q}, h_{-q+1}, \dots, h_{-1}\}$, $H^+ = \{h_1, \dots, h_p\}$. Giả thiết $h_{-q} > h_{-q+1} > \dots > h_{-1}$ và $h_1 < \dots < h_p$, ở đây $p, q \geq 2$. 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 $0 < |x| \leq k^*$, với mọi $x \in X$. 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:

- Trường hợp $k=j$: $O_{min,k}(x) = I_{k+1}(h_{-1}x) \cup I_{k+1}(h_1x)$
- Trường hợp $1 \leq k < j$: $O_{min,k}(x) = I_j(x)$
- Trường hợp $j+1 \leq k \leq k^*$: $O_{min,k}(x) = I_{k+1}(h_{-1}y) \cup I_{k+1}(h_1y)$

Từ đó, trong bài báo thống nhất cách biểu diễn dữ liệu ngôn ngữ mờ theo định nghĩa sau:

Định nghĩa 2.1.[5] Cho $x \in X \cup C$, một biểu diễn khoảng của x là một tập $IRp(x)$ các khoảng được xác định:

$$IRp(x) = \{O_{min,k}(x) \mid 1 \leq k \leq n\}$$

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 $[a-\epsilon, a+\epsilon]$, với ϵ được xem là bán kính với tâm a . Vì $[a-\epsilon, a+\epsilon]$ là dữ liệu rõ nên $O_{min,k}([a-\epsilon, a+\epsilon]) = \{[a-$

$\varepsilon, a+\varepsilon\}$ }, với mọi $1 \leq k \leq k^*$ và $\text{IRp}([a-\varepsilon, a+\varepsilon]) = \{[a-\varepsilon, a+\varepsilon]\}$. Với phương pháp biểu diễn khoảng nêu trên, chúng ta đã xem xét các kiểu dữ liệu khác nhau trên một quan điểm thống nhất. Mỗi thông tin đều được biểu diễn bởi một tập các khoảng trên không gian tham chiếu. Như vậy, vấn đề đặt ra là định nghĩa quan hệ bằng nhau như thế nào để so sánh hai thông tin mờ đã được biểu diễn dưới dạng khoảng theo một cách thống nhất. 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 $P_{k+1}(H_n) = \{I_{k+1}(h_i y) \mid y \in H_n\}$, với $n = 1, 2$. Hai khoảng $I_{k+1}(x)$ và $I_{k+1}(y)$ 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 $I_{k+1}(x)$ đến $I_{k+1}(y)$. 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 $\{I_{k+1}(h_i y) \mid h_i \in H_1\}$. Hơn nữa, do $I_{k+1}(h_{-l} y) \leq v(y) \leq I_{k+1}(h_l y)$ hoặc là $I_{k+1}(h_l y) \leq v(y) \leq I_{k+1}(h_{-l} y)$ nên bao giờ ta cũng có $v(y) \in I_{k+1}(h_i y) \mid h_i \in H_1$, với v là hàm định lượng ngữ nghĩa trên \underline{X} .

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 $H_2^+ = H_2 \cap H^+$, do $h_{-q} \in H_2^-$ và $h_p \in H_2^+$ nên các cụm được sinh ra từ các khoảng mờ thuộc $P_{k+1}(H_2)$ có ba loại sau đây:

- Cụm nằm bên trái x_0 : $\{I_{k+1}(h_i x_0) \mid h_i \in H_2^+\}$
- Cụm nằm bên phải x_{m-1} : $\{I_{k+1}(h_i x_{m-1}) \mid h_i \in H_2^+\}$
- Các cụm nằm giữa x_s và x_{s+1} với $s = 0, \dots, m-2$: phụ thuộc vào $\text{Sgn}(h_p x_s)$ và $\text{Sgn}(h_p x_{s+1})$ như sau:

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

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

Định nghĩa 2.2. [5] Mỗi \mathbf{C} thuộc \mathbf{C} , ta gọi khoảng tương tự mức k ứng với \mathbf{C} là:

$$S_k(\mathbf{C}) = \cup \{I_{k+1} \mid I_{k+1} \in \mathbf{C}\}$$

Với cách định nghĩa này, mỗi khoảng $S_k(\mathbf{C})$ sẽ không quá lớn để phủ bất kỳ một khoảng I_k nhưng lại không quá nhỏ để nằm gọn trong một khoảng I_{k+1} nào. Vì $\{S_k(\mathbf{C}) \mid \mathbf{C} \in \mathbf{C}\}$ 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 , tồn tại duy nhất một cụm \mathbf{C} sao cho $v(x) \in S_k(\mathbf{C})$. Vì vậy, chúng ta

có thể định nghĩa $S_k(x)=S_k(C)$.

Mệnh đề 2.1.[5] Cho \underline{X} là ĐSGT tuyến tính đầy đủ, trong đó H^+ và H có ít nhất hai phần tử. Khi đó:

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

(2) Với mọi $x,u \in X \cup C$, nếu $v(x) \in S_k(u)$ 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 2.3. Cho miền trị của thuộc tính A_i là một ĐSGT tuyến tính đầy đủ \underline{X} và một đối tượng bất kỳ o trên tập thuộc tính $\{A_1, A_2, \dots, A_n\}$ của lớp C , 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 miền trị của thuộc tính A_i là một ĐSGT tuyến tính đầy đủ \underline{X} và hai đối tượng bất kỳ o_1, o_2 trên tập thuộc tính $\{A_1, A_2, \dots, A_n\}$ của lớp C , 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, ta định nghĩa:

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

Bổ đề 2.1. 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 , ta có:

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

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

2.3. Lớp mờ trong CSDL hướng đối tượng mờ

Dựa trên các thảo luận trên, chúng ta thấy rằng các lớp trong CSDL hướng đối tượng mờ có thể mờ. Theo đó, trong CSDL hướng đối tượng mờ, 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 hướng đối tượng, một lớp được định nghĩa bao gồm mỗi quan hệ kế thừa, thuộc tính và phương thức. Để xác định một lớp mờ, cần bổ sung một số định nghĩa mới. Khi khai báo mỗi quan hệ kế thừa cần chỉ ra mức mà lớp này là lớp con của lớp cha,

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

```

Đối với các thuộc tính không phải là thuộc tính mờ, các kiểu dữ liệu bao gồm các loại đơn giản như số nguyên, thực, logic, chuỗi, và các loại phức tạp như kiểu tập hợp và kiểu đối tượng. Đối với thuộc tính mờ, các kiểu dữ liệu là kiểu mờ dựa trên các kiểu dữ liệu đơn giản hoặc phức tạp, cho phép biểu diễn thông tin không chính xác [5]. Do có sự thay đổi trong mô hình cơ sở dữ liệu sẽ tác động lên các thao tác trong mô hình mới. Vì vậy, trong phần tiếp theo, chúng tôi sẽ nghiên cứu vấn đề xử lý yêu cầu truy vấn trên các lớp mờ theo ngữ nghĩa mới.

3. Truy vấn trong cơ sở dữ liệu hướng đối tượng mờ

3.1. Xử lý truy vấn

Đối với truy vấn trong CSDL hướng đối tượng mờ như đã đề cập ở phần trên, mỗi đối tượng thuộc về một lớp theo mức k ($k \in Z^+$). Ngoài ra, một đối tượng đáp ứng các điều kiện truy vấn cũng theo mức k vì thông tin mờ xảy ra trong điều kiện truy vấn, hoặc trong đối tượng, hoặc cả hai. Do đó, việc xử lý truy vấn dựa trên đề xuất mô hình CSDL 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 CSDL hướng đối tượng mờ có liên quan đến sự lựa chọn các mức, và vì vậy, một câu truy vấn trong cơ sở dữ liệu hướng đối tượng mờ có cấu trúc như sau:

```

SELECT < danh sách thuộc tính >
FROM < Lớp1 WITH mứck1, ..., Lớpm WITH mứckm >
WHERE < điều_kiện_truy_vấn WITH mứck >

```

Trong đó, <điều kiện truy vấn> là một điều kiện mờ hoặc liên kết các điều kiện mờ có sử dụng các phép toán tuyến và hội, k là số nguyên dương.

Thuật toán 3.1: 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 $C: \{o_i, i=1, \dots, m\}$

Câu truy vấn dạng *select ... from ... where* ($A_i =_k fvalue_i \psi A_j =_k fvalue_j$), trong đó ψ là phép toán hội (and) hoặc tuyển (or).

Ra: Tập các đối tượng $O = \{o_t: o_t(A_i) =_k fvalue_i \psi o_t(A_j) =_k fvalue_j\}$

Phương pháp:

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

$X_{A_i} = \{0, c_{A_i}^-, W, c_{A_i}^+, 1\}$, $H_{A_i} = H_{A_i}^+ \cup H_{A_i}^-$. trong đó $H_{A_i}^+ = \{h_1, h_2\}$, $H_{A_i}^- = \{h_3, h_4\}$, với $h_1 < h_2$ và $h_3 > h_4$. Chọn độ đo tính mờ cho các phần tử sinh và gia tử.

$X_{A_j} = \{0, c_{A_j}^-, W, c_{A_j}^+, 1\}$, $H_{A_j} = H_{A_j}^+ \cup H_{A_j}^-$. trong đó $H_{A_j}^+ = \{h_1, h_2\}$, $H_{A_j}^- = \{h_3, h_4\}$, với $h_1 < h_2$ và $h_3 > h_4$. Chọn độ đo tính mờ cho các phần tử sinh và gia tử.

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

(3) $O = \emptyset$

(4) Xây dựng lân cận tối thiểu k của giá trị thuộc tính A_i và A_j của mỗi đối tượng, ký hiệu là $O_{\min,k}(o_t(A_i))$ và $O_{\min,k}(o_t(A_j))$, với $j = 1, \dots, m$,

(5) Xây dựng lớp tương đương $S_k(fvalue_i)$ và $S_k(fvalue_j)$

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

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

If $O_{\min,k}(o_t(A_i)) \subseteq S_k(fvalue_i) \psi O_{\min,k}(o_t(A_j)) \subseteq S_k(fvalue_j)$ then

$O = O \cup o_t$

(7) Return O

Ví dụ 3.1. Cho các đối tượng của lớp “Nhân viên trẻ” như sau:

Bảng 3.1. Lớp nhân viên trẻ

Đối tượng	Tên	Tuổi	Hệ số lương	Số lượng sản phẩm
O_1	Hải	27	Khả năng thấp	Khoảng 20
O_2	Nam	Khoảng 30	Ít thấp	Rất cao
O_3	Thái	Hơn trẻ	Khả năng ít thấp	Khả năng cao

O ₄	Quốc	Ít hơn trẻ	Khoảng 3.0	Khoảng 17
O ₅	Phú	Trẻ	2.67	15

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 $k = 1$ ”. Ta thấy các đối tượng của lớp “Nhân viên trẻ” đã cho đều thuộc lớp theo mức 1. Do đó chỉ cần sử dụng thuật toán 3.1. để xác định các đối tượng thỏa điều kiện mờ là “có hệ số lương không thấp lắm và số lượng sản phẩm khả năng cao với mức $k = 1$ ”.

Đối với thuộc tính Số lượng sản phẩm (SLSP)

Chúng ta sẽ chuyển giá trị thuộc tính SLSP của các đối tượng về biểu diễn khoảng: $\text{Irp}(O_1(\text{SLSP})) = \{[19,21]\}$; $\text{Irp}(O_4(\text{SLSP})) = \{[16,18]\}$; $\text{Irp}(O_5(\text{SLSP})) = \{[15,15]\}$. Gọi $\underline{X}_{\text{SLSP}} = (X, G, C, H, \leq)$ là một ĐSGT của thuộc tính SLSP, với $G = \{\text{thấp, cao}\}$, $H^+ = \{\text{rất, hơn}\}$, $H = \{\text{khả năng, ít}\}$, $\text{rất} > \text{hơn}$, $\text{ít} > \text{khả năng}$. Chọn $W = 0.4$, $fm(\text{thấp}) = 0.4$, $fm(\text{cao}) = 0.6$, $\mu(\text{rất}) = 0.2$, $\mu(\text{hơn}) = 0.3$, $\mu(\text{khả năng}) = 0.3$, $\mu(\text{ít}) = 0.2$. Chọn $D_{\text{SLSP}} = [0,30]$ nên ta sẽ dùng hệ số $r = 30$ để chuyển đổi từ $[0,1]$ thành $[0,30]$.

Ta có: $fm_r(\text{ít cao}) = \mu(\text{ít}) \times fm(\text{cao}) \times r = 0.2 \times 0.6 \times 30 = 3.6$, $fm_r(\text{khả năng cao}) = 5.4$, $fm_r(\text{hơn cao}) = 5.4$, $fm_r(\text{rất cao}) = 3.6$. $O_{\min,1}(\text{rất cao}) = I_{2,r}(\text{rất cao}) = (26.4,30]$; $O_{\min,1}(\text{hơn cao}) = I_{2,r}(\text{hơn cao}) = (21,26.4]$; $O_{\min,1}(\text{khả năng cao}) = I_{2,r}(\text{khả năng cao}) = (15.6,21]$. $O_{\min,2}(\text{rất cao}) = I_{3,r}(\text{khả năng rất cao}) \cup I_{3,r}(\text{hơn rất cao}) = (27.12,28.2] \cup (28.2,29.28] = (27.12,29.28]$, $O_{\min,2}(\text{hơn cao}) = I_{3,r}(\text{khả năng hơn cao}) \cup I_{3,r}(\text{hơn hơn cao}) = (22.08,23.7] \cup (23.7,25.32] = (22.08,25.32]$, $O_{\min,2}(\text{khả năng cao}) = I_{3,r}(\text{khả năng khả năng cao}) \cup I_{3,r}(\text{hơn khả năng cao}) = (16.68,18.3] \cup (18.3,19.92] = (16.68,19.92]$.

Ta có: $O_{\min,k}(\text{khoảng } 20) = [19,21]$, $O_{\min,k}(\text{khoảng } 17) = [16,18]$, $O_{\min,k}(15) = [15,15]$, với mọi $k \leq k^*$.

$O_{\min,1}(\text{rất cao}) = I_{2,r}(\text{rất cao}) = (26.4,30]$; $O_{\min,1}(\text{khả năng cao}) = I_{2,r}(\text{khả năng cao}) = (15.6,21]$.

$S_{1,r}(\text{khả năng cao}) = I_{2,r}(\text{khả năng cao}) \cup I_{2,r}(\text{hơn cao}) = (15.6,21] \cup (21,26.4] = (15.6, 26.4]$.

Đối với thuộc tính Hệ số lương (HSL)

Gọi $\underline{X}_{\text{HSL}} = (X, G, C, H, \leq)$ là một ĐSGT, với $G = \{\text{thấp, cao}\}$, $H^+ = \{\text{rất, hơn}\}$, $H = \{\text{khả năng, ít}\}$, $\text{rất} > \text{hơn}$, $\text{ít} > \text{khả năng}$. Chọn $W = 0.4$, $fm(\text{thấp}) = 0.4$, $fm(\text{cao}) = 0.6$, $\mu(\text{rất}) = 0.3$, $\mu(\text{hơn}) = 0.25$, $\mu(\text{khả năng}) = 0.3$, $\mu(\text{ít}) = 0.25$. Chọn $D_{\text{HSL}} = [0, 7.5]$. $O_{\min,k}(\text{khoảng } 3.0) = [2.67,3.33]$, $O_{\min,k}(2.67) = [2.67,2.67]$, với mọi $k \leq k^*$, $O_{\min,1}(\text{khả năng thấp}) = I_{2,r}(\text{khả năng thấp}) = (2.35,3.25]$, $O_{\min,1}(\text{ít thấp}) = I_{2,r}(\text{ít thấp}) = (3.25,4]$, $O_{\min,1}(\text{khả năng ít thấp}) = I_{3,r}(\text{khả năng ít thấp}) = (3.55,3.775]$, $S_{1,r}(\text{ít thấp}) = I_{2,r}(\text{rất thấp}) \cup I_{2,r}(\text{ít thấp}) = [0,4]$. Duyệt lần lượt các đối tượng như bước 6, ta được các đối tượng thỏa điều

kiện truy vấn:

Bảng 3.2: Kết quả thực hiện truy vấn (với $k=1$)

Đối tượng	Tên	Tuổi	Hệ số lương	Số lượng sản phẩm
O_1	Hải	27	Khả năng thấp	Khoảng 20
O_2	Nam	Khoảng 30	Ít thấp	Rất cao
O_3	Thái	Hơn trẻ	Khả năng ít thấp	Khả năng cao
O_4	Quốc	Ít hơn trẻ	Khoảng 3.0	Khoảng 17

Nếu mức truy vấn trên được sửa lại thành $k=2$, khi đó:

Đối với thuộc tính Số lượng sản phẩm (SLSP): $O_{\min,2}(\text{rất cao})=I_{3,r}(\text{khả năng rất cao}) \cup I_{3,r}(\text{hơn rất cao}) = (27.12,28.2] \cup (28.2,29.28] = (27.12,29.28]$, $O_{\min,2}(\text{khả năng cao})=I_{3,r}(\text{khả năng khả năng cao}) \cup I_{3,r}(\text{hơn khả năng cao}) = (16.68,18.3] \cup (18.3,19.92] = (16.68,19.92]$.

$S_{2,r}(\text{khả năng cao})=I_{3,r}(\text{khả năng khả năng cao}) \cup I_{3,r}(\text{khá khả năng cao}) = (16.68,18.3] \cup (18.3,19.92] = (16.68,19.92]$

Đối với thuộc tính Hệ số lương (HSL): $O_{\min,2}(\text{khả năng thấp})=I_{3,r}(\text{khả năng khả năng thấp}) \cup I_{3,r}(\text{hơn khả năng thấp}) = (2.62,2.845] \cup (2.845,3.115] = (2.62,3.115]$, $O_{\min,2}(\text{ít thấp})= I_{3,r}(\text{khả năng ít thấp}) \cup I_{3,r}(\text{hơn ít thấp}) = (3.55,3.963]$, $O_{\min,2}(\text{khả năng ít thấp})= I_{3,r}(\text{khả năng ít thấp}) = (3.55,3.775]$.

$S_{2,r}(\text{ít thấp})=I_{2,r}(\text{khả năng ít thấp}) \cup I_{2,r}(\text{hơn ít thấp}) = (3.55,3.775] \cup (3.775,3.963] = (3.55,3.963]$.

Khi đó, các đối tượng thỏa điều kiện truy vấn là:

Bảng 3.3. Kết quả thực hiện truy vấn (với $k=2$)

Đối tượng	Tên	Tuổi	Hệ số lương	Số lượng sản phẩm
O_3	Thái	Hơn trẻ	Khả năng ít thấp	Khả năng cao

3.2. Đưa lượng từ ngôn ngữ vào câu truy vấn

Truy vấn dữ liệu mờ đã giúp khai thác dữ liệu trong mô hình CSDL hướng đối tượng mờ một cách linh hoạt. Điểm mới của câu truy vấn đó chính là sử dụng các điều kiện mờ, như các ví dụ trên: “Cho biết những nhân viên có số lượng sản phẩm khá cao”, “Tìm những nhân viên có hệ số lương ít thấp và số lượng sản phẩm khả năng cao”,... Để xử lý các câu truy vấn dạng như vậy chúng ta chỉ cần tìm những đối tượng thỏa mãn những điều kiện mờ theo mức cho trước và kết quả của truy vấn là tất cả các đối tượng tìm được. Tuy nhiên, khi gặp những yêu cầu truy vấn có dạng như “cho biết

một vài nhân viên có số lượng sản phẩm khá cao”, “cho biết **ít nhất 3** nhân viên có hệ số lương khả năng thấp và số lượng sản phẩm cao”, ... thì vấn đề xử lý câu truy vấn là phức tạp, bởi vì, ngoài việc tìm những đối tượng thỏa điều kiện truy vấn mờ, kết quả của truy vấn còn phụ thuộc vào các lượng từ “một vài” và “ít nhất 3”.

Theo [4], Zadel chia lượng từ ngôn ngữ thành hai loại đó là: lượng từ tuyệt đối (*absolute quantifier*) và lượng từ tỉ lệ (*proportion quantifier*). Lượng từ tuyệt đối thường dùng trong các mệnh đề có số lượng xác định như “ít nhất 3”, “nhiều hơn 5”, ... Lượng từ tỉ lệ thể hiện những số lượng phụ thuộc vào số lượng tập các đối tượng đang xử lý, chẳng hạn như “một vài”, “khoảng một nửa”, “hầu hết”,...

Trong thực tế, để đáp ứng yêu cầu thao tác dữ liệu, việc xây dựng phương pháp đánh giá lượng từ tương đối và lượng từ tuyệt đối, và việc đưa lượng từ vào câu truy vấn là cần thiết và cần được quan tâm giải quyết.

Do truy vấn sử dụng lượng từ có thể xem là một sự mở rộng của truy vấn hướng đối tượng mờ, cho nên một câu truy vấn hướng đối tượng mờ sử dụng lượng từ có thể có cấu trúc như sau:

SELECT < danh sách thuộc tính >

FROM <Lớp₁ WITH mức_{k₁}, ..., Lớp_m WITH mức_{k_m}>

WHERE <Lượng từ> (điều_kiện_truy_vấn WITH mức_k>

Để đánh giá lượng từ trong câu truy vấn, trước hết chúng ta xác định giá trị chân lý của các điều kiện mờ đối với các lớp tham gia truy vấn. Có nghĩa là tìm những đối tượng thuộc lớp tham gia truy vấn thỏa điều kiện mờ theo mức k cho trước như đã trình bày ở trên. Tiếp theo, ta sẽ đánh giá lượng từ trong câu truy vấn dựa vào các đối tượng vừa tìm được so với số đối tượng của lớp ban đầu tham gia truy vấn.

Gọi Q là lượng từ trong câu truy vấn, n là tổng số đối tượng ban đầu của lớp C , miền trị $D_C = [0..n]$. Chúng ta có thể chia lượng từ Q thành hai trường hợp:

(a) Trường hợp Q là lượng từ tuyệt đối: Ký hiệu $|Q|$ là số lượng xác định của lượng từ Q ,

Nếu Q đơn điệu tăng: Ta xây dựng một hàm $f_Q^A : D_C \rightarrow \{0,1\}$ sao cho $\forall x \in D_C, f_Q^A(x) = 1$ nếu $x \geq |Q|$ và $f_Q^A(x) = 0$ nếu ngược lại.

Nếu Q đơn điệu giảm: Ta xây dựng một hàm $f_Q^D : D_C \rightarrow \{0,1\}$ sao cho $\forall x \in D_C, f_Q^D(x) = 1$ nếu $x = |Q|$ và $f_Q^D(x) = 0$ nếu ngược lại.

(b) Trường hợp Q là lượng từ tỉ lệ: Trước hết, ta có thể xét các tình huống như sau: Khi ta nói hầu hết các đối tượng của lớp thỏa mãn điều kiện truy vấn mờ, có nghĩa là số đối tượng tìm được phải xấp xỉ số đối tượng của lớp; Hoặc trong trường hợp khác, chỉ một số ít các đối tượng thỏa điều kiện, có nghĩa là tổng số đối tượng tìm được phải

xấp xỉ $\frac{1}{n}$. Hay một giả thiết ta thường gặp đó là khoảng một nửa các đối tượng của lớp thỏa mãn điều kiện, khi đó chắc chắn rằng tổng số đối tượng phải là xấp xỉ của $\frac{n}{2}$. Điều

này gợi ý cho chúng ta có thể đánh giá lượng từ tỷ lệ dựa trên phân hoạch của $[0, n]$. Để chuyển $[0, n]$ về $[0, 1]$ ta có thể sử dụng một số phép biến đổi tuyến tính, vì vậy, giả thiết mọi miền $D_c = [0, n]$ như vậy đều là khoảng $[0, 1]$. Khi đó ta xây dựng hai khoảng mờ của hai khái niệm nguyên thủy nhỏ và lớn, ký hiệu là $I(\text{nhỏ})$ và $I(\text{lớn})$ với độ dài tương ứng là $fm(\text{nhỏ})$ và $fm(\text{lớn})$ sao cho chúng tạo thành một phân hoạch của miền tham chiếu $[0, 1]$. Tiếp đến, ta đi xây dựng các lớp tương đương $S(1)$, $S(\text{lớn})$, $S(W)$, $S(\text{nhỏ})$, $S(0)$ dựa vào độ đo tính mờ của các gia tử và các khái niệm nguyên thủy trong ĐSGT. Do đó, nếu gọi t là tổng số đối tượng thỏa điều kiện mờ với lượng từ “hầu hết” thì $t \in S(1) \times n$, còn nếu t là tổng số đối tượng thỏa điều kiện mờ với lượng từ “một vài” thì $t \in S(0) \times n$.

Từ đó, ta có thể khẳng định rằng tổng số đối tượng của lớp C thỏa điều kiện mờ với lượng từ Q , ký hiệu là t , khi đó: $t \in S(1) \times n$, hoặc $t \in S(\text{lớn}) \times n$, hoặc $t \in S(W) \times n$, hoặc $t \in S(\text{nhỏ}) \times n$, hoặc $t \in S(0) \times n$, hay nói cách khác: $\frac{t}{n}$ phải thuộc về một trong các khoảng $S(1)$, $S(\text{lớn})$, $S(W)$, $S(\text{nhỏ})$, $S(0)$.

Như vậy, chúng ta đã xây dựng được phương pháp đánh lượng từ trong truy vấn, nó sẽ được dùng để xác định kết quả của truy vấn sau khi đã tìm được các đối tượng thỏa điều kiện truy vấn. Gọi O là tập các đối tượng thỏa điều kiện truy vấn và $|O|$ là tổng số đối tượng của tập O , Q là lượng từ tham gia vào truy vấn. Vì số lượng các lượng từ không nhiều nên ta có thể liệt kê một số các lượng từ tương đối và tuyệt đối. Các lượng từ tuyệt đối thường ở hai dạng là “ít nhất m ” và “nhiều nhất m ”, còn các lượng từ tương đối thường bao gồm: “một ít (*few*)” hay “một vài (*several*)”, “kha khá (*quite a few*)”, “khoảng một nửa (*about half*)”, “nhiều (*many*)”, “hầu hết (*most*)”, “với mọi (*all, every*)”. Kết quả truy vấn được xác định như sau:

Nếu Q là lượng từ tuyệt đối, theo (a) ta sẽ đi xác định hàm $f_Q^A(|O|)$ hoặc $f_Q^D(|O|)$. Kết quả truy vấn là tập O nếu $f_Q^A(|O|) = 1$ (hoặc $f_Q^D(|O|) = 1$), ngược lại, kết quả là rỗng.

Nếu Q là lượng từ tương đối, khi đó kết quả truy vấn là tập O nếu $\frac{|O|}{n} \in S(0)$ khi Q là “một ít”, hoặc $\frac{|O|}{n} \in S(\text{nhỏ})$ khi Q là “kha khá”, hoặc $\frac{|O|}{n} \in S(W)$ khi Q là “khoảng một nửa”, hoặc $\frac{|O|}{n} \in S(\text{lớn})$ khi Q là “nhiều”, hoặc $\frac{|O|}{n} \in S(1)$ khi Q là “hầu hết”, hoặc $\frac{|O|}{n} = 1$ khi Q là “với mọi”, ngược lại, kết quả truy vấn là rỗng.

Lượng từ tương đối Q có thể không phải là các lượng từ đã liệt kê như trên, nhưng ta có thể dựa vào cấu trúc thứ tự trong ngôn ngữ tự nhiên để đưa nó vào một trong 6 trường hợp như đã trình bày.

Thuật toán 3.2: Xử lý lượng từ trong truy vấn

Vào: Lớp C cùng với các thuộc tính $\{A_1, A_2, \dots, A_n\}$, tập các đối tượng thuộc lớp C : $\{o_i, i=1, \dots, m\}$

Câu truy vấn dạng: *select * from C where $Q(A_i =_k \text{fvalue}_i \psi A_j =_k \text{fvalue}_j)$* , trong đó ψ là phép toán hội (and) hoặc tuyển (or), $Q \in Q_{ABS} \cup Q_{PRO}$ với $Q_{ABS} = \{\text{ít nhất } l, \text{ nhiều nhất } l\}$, $Q_{PRO} = \{\text{một ít, kha khá, khoảng một nửa, hầu hết, với mọi}\}$,

Ra: Tập các đối tượng $O_Q = \{o_i: Q(o_i(A_i) =_k \text{fvalue}_i \psi o_i(A_j) =_k \text{fvalue}_j)\}$

Phương pháp:

- (1) $O_Q = \emptyset$
- (2) Sử dụng thuật toán 3.1 ta được kết quả là tập các đối tượng O
- (3) if $Q \in Q_{ABS}$ then
- (4) if $f_Q^A(|O|) = 1$ or $f_Q^D(|O|) = 1$ then $O_Q = O$
- (5) if $Q \in Q_{PRO}$ then
- (6) Xây dựng các khoảng $S_I(1), S_I(\text{lớn}), S_I(W), S_I(\text{nhỏ}), S_I(0)$
- (7) Case Q of
- (8) “Một ít”: if $\frac{|O|}{n} \in S_I(0)$ then $O_Q = O$
- (9) “Kha khá”: if $\frac{|O|}{n} \in S_I(\text{nhỏ})$ then $O_Q = O$
- (10) “Khoảng một nửa”: if $\frac{|O|}{n} \in S_I(W)$ then $O_Q = O$
- (11) “Nhiều”: if $\frac{|O|}{n} \in S_I(\text{lớn})$ then $O_Q = O$
- (12) “Hầu hết”: if $\frac{|O|}{n} \in S_I(1)$ then $O_Q = O$
- (13) “Với mọi”: if $\frac{|O|}{n} = 1$ then $O_Q = O$
- (14) Return O_Q

Ví dụ 3.2. Sử dụng các đối tượng của lớp “Nhân viên trẻ” như ví dụ 3.1. Cho các yêu cầu truy vấn mờ có sử dụng lượng từ như sau:

(a) “Tìm ít nhất 3 nhân viên có số lượng sản phẩm khả năng cao với mức $k = 2$ ”,

Dựa vào ví dụ 3.1, ta có các kết quả tính toán thuộc tính SLSP

$O_{\min, k}(\text{khoảng } 20) = [19, 21]$, $O_{\min, k}(\text{khoảng } 17) = [16, 18]$, $O_{\min, k}(15) = [15, 15]$, với mọi $k \leq k^*$.

$O_{\min,2}(\text{rất cao})=(27.12,29.28]$, $O_{\min,2}(\text{khả năng cao})=(16.68,19.92]$. $S_{2,r}(\text{khả năng cao})=(16.68,19.92]$. Sử dụng quan hệ bằng nhau theo mức $k = 2$, ta được các đối tượng thỏa điều kiện:

Bảng 3.4: Kết quả “Truy vấn số lượng sản phẩm khả năng cao” (với $k = 2$)

Đối tượng	Tên	Tuổi	Hệ số lương	Số lượng sản phẩm
O_3	Thái	Hơn trẻ	Khả năng ít thấp	Khả năng cao
O_4	Quốc	Ít hơn trẻ	Khoảng 3.0	Khoảng 17

Vì lượng từ “ít nhất 3” là lượng từ tuyệt đối và đơn điệu tăng, nên ta có $f_{\text{ít nhất } 3}^A(|O|) = f_{\text{ít nhất } 3}^A(3) = 0$, vì vậy kết quả truy vấn không có đối tượng nào thỏa mãn.

(b) “Cho biết nhiều nhân viên có hệ số lương khả năng thấp và số lượng sản phẩm cao với mức $k=1$ ”

Dựa vào ví dụ 3.1, ta có được các đối tượng thỏa điều kiện truy vấn:

Bảng 3.5. Kết quả “Truy vấn nhân viên có hệ số lương khả năng thấp và số lượng sản phẩm cao với $k = 1$)

Đối tượng	Tên	Tuổi	Hệ số lương	Số lượng sản phẩm
O_1	Hải	27	Khả năng thấp	Khoảng 20
O_2	Nam	Khoảng 30	Ít thấp	Rất cao
O_3	Thái	Hơn trẻ	Khả năng ít thấp	Khả năng cao
O_4	Quốc	Ít hơn trẻ	Khoảng 3.0	Khoảng 17

Vì lượng từ “nhiều” là lượng từ tương đối nên ta đi xây dựng các khoảng mờ $S_1(1)$, $S_1(\text{lớn})$, $S_1(W)$, $S_1(\text{nhỏ})$, $S_1(0)$. Chọn $fm(\text{lớn}) = 0.35$, $fm(\text{nhỏ}) = 0.65$, $\mu(\text{rất}) = 0.4$, $\mu(\text{hơn}) = 0.15$, $\mu(\text{khả năng}) = 0.25$, $\mu(\text{ít}) = 0.2$. Ta phân hoạch đoạn $[0,1]$ thành 5 khoảng tương tự mức 1 là:

$fm(\text{rất lớn}) = 0.35 \times 0.35 = 0.1225$. Vậy $S_1(1) = (0.8775, 1]$. $fm(\text{khả năng lớn}) + fm(\text{hơn lớn}) = 0.25 \times 0.35 + 0.15 \times 0.35 = 0.14$. Vậy $S_1(\text{lớn}) = (0.7375, 0.8775]$. $fm(\text{ít nhỏ}) + fm(\text{ít lớn}) = 0.2 \times 0.65 + 0.2 \times 0.35 = 0.20$. Vậy $S_1(W) = (0.5375, 0.7375]$; $fm(\text{khả năng nhỏ}) + fm(\text{hơn nhỏ}) = 0.25 \times 0.65 + 0.15 \times 0.65 = 0.26$.

Vậy $S_1(\text{nhỏ}) = (0.2775, 0.5375]$ và $S_1(0) = [0, 0.2775]$.

Vì $\frac{|O|}{n} = \frac{4}{5} = 0.8 \in S_1(\text{lớn}) = (0.7375, 0.8775]$ nên kết quả của truy vấn là:

Bảng 3.6. Kết quả “Truy vấn nhiều nhân viên có hệ số lương khả năng thấp và số lượng sản phẩm cao với $k = 1$)

Đối tượng	Tên	Tuổi	Hệ số lương	Số lượng sản phẩm
O_1	Hải	27	Khả năng thấp	Khoảng 20

O ₂	Nam	Khoảng 30	Ít thấp	Rất cao
O ₃	Thái	Hơn trẻ	Khả năng ít thấp	Khả năng cao
O ₄	Quốc	Ít hơn trẻ	Khoảng 3.0	Khoảng 17

4. Kết luận

Trong bài báo này chúng tôi đã đề xuất một cách tiếp cận mới để xây dựng mô hình CSDL hướng đối tượng mờ theo cách tiếp cận ĐSGT. Dựa trên cấu trúc thứ tự ngữ nghĩa của ĐSGT mỗi giá trị ngôn ngữ được biểu diễn một khoảng $I_p(x)$, từ đó, xây dựng biểu diễn dạng khoảng cho các kiểu dữ liệu khác nhau một cách thống nhất. Tiếp theo, các quan hệ đối sánh theo mức k được đề xuất nhằm phục vụ cho việc xác định mối quan hệ lớp đối tượng mờ và phân cấp thừa kế mờ. Cuối cùng, chúng tôi đề xuất cách xử lý truy vấn và xử lý truy vấn có sử dụng lượng từ ngôn ngữ trong mô hình mới này nhằm đáp ứng cho việc tìm kiếm đối tượng trong mô hình CSDL hướng đối tượng mờ. Các dạng ràng buộc dữ liệu là hướng nghiên cứu trong các bài báo tiếp theo.

TÀI LIỆU THAM KHẢO

- [1]. 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.
- [2]. G. Bordogna, G. Pasi, and D. Lucarella, *A fuzzy object-oriented data model for managing vague and uncertain information*, International Journal of Intelligent Systems, Vol. 14, (1999), 623-651.
- [3]. L.A. Zadeh, *A Computational approach to fuzzy quantifiers in natural languages*, Computers and Mathematics with Applications, Vol 9. No. 1, (1983), 149-184.
- [4]. M. Umamo, T. Imada, I. Hatono, and H. Tamura, *Fuzzy object-oriented databases and implementation of its SQL-type data manipulation language*, in Proceedings of the 7th IEEE International Conference on Fuzzy Systems, Vol. 2, (1998), 1344- 1349.
- [5]. 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, tập 25, số 4, (2009), 314-332.
- [6]. Nguyễn Cát Hồ, Nguyễn Công Hào, *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, tập 24, số 4 (2008), 281-294.
- [7]. Z. M. Ma, *Advances in Fuzzy Object-Oriented Databases: Modeling and Applications*, Idea Group Publishing, 2004.
- [8]. Z. M. Ma and Li Yan: *A Literature Overview of Fuzzy Database Models*, Journal Of Information Science And Engineering 24, (2008), 189-202.

QUERY DATA IN FUZZY OBJECT-ORIENTED DATABASES BASED ON THE SEMANTICS HEDGE ALGEBRA

Nguyen Cong Hao¹, Truong Thi My Le²

¹ *Information Technology Center, Hue University*

² *Quang Trung University, Qui Nhon*

Abstract: In fuzzy object-oriented databases, query language is examined by authors with many different approaches such as fuzzy-set approaching theory, *possibility theory*... However, there has been difficulty in describing and comparing data. Thus, in this paper, we propose a method for processing query in fuzzy object-oriented databases based on the semantics hedge algebra structure with much benefit. Finally, the query language with fuzzy quantifiers is considered.