

TRUY VẤN DỮ LIỆU VỚI THÔNG TIN MỜ VÀ KHÔNG CHẮC CHẴN TRONG CƠ SỞ DỮ LIỆU HƯỚNG ĐỐI TƯỢNG

QUERY DATA WITH FUZZY INFORMATION AND UNCERTAINTIES IN OBJECT-ORIENTED DATABASES

Đoàn Văn Thắng

Trường CĐ Hữu Nghị CNTT Việt-Hàn

TÓM TẮT

Ngày nay, có nhiều hướng tiếp cận khác nhau để xử lý giá trị thuộc tính của lớp đối tượng được nhiều tác giả trong và ngoài nước quan tâm. Trong bài báo này, chúng tôi đề xuất một phương pháp xử lý giá trị thuộc tính của lớp đối tượng trong cơ sở dữ liệu hướng đối tượng với thông tin mờ và không chắc chắn dựa trên định lượng ngữ nghĩa của đại số gia tử. Bài báo sử dụng phương pháp này để chuyển giá trị thuộc tính của lớp đối tượng về đoạn con $[0, 1]$ tương ứng. Ngoài ra, vì tính mờ của các phần tử trong đại số gia tử là các đoạn con của $[0, 1]$. Từ đó, đưa ra một thuật toán thực hiện việc so sánh hai đoạn con $[0, 1]$ này nhằm đáp ứng yêu cầu của việc truy vấn dữ liệu.

ABSTRACT

In this paper, a method of handling attribute values of object classes in object-oriented database with fuzzy information and uncertainties based on quantitatively semantics-based hedge algebras will be presented. The article also deals with this method to transfer the attribute values of objects into sub intervals in $[0, 1]$ respectively. In addition, because the fuzziness of the elements in the hedge algebra is the subintervals in $[0, 1]$, an algorithm that allows the comparison of two subintervals $[0, 1]$ will be presented to meet the requirements of the query data.

1. Giới thiệu

Trong những năm gần đây, khi mà mô hình cơ sở dữ liệu hướng đối tượng (CSDL HĐT) truyền thống không phù hợp với thực tế bởi vì thông tin về các đối tượng trong thế giới thực thường có tính mơ hồ, không chắc chắn, không đầy đủ. Để giải quyết vấn đề đặt ra trong thực tế mô hình CSDL HĐT mờ đã được đề xuất nhằm biểu diễn và xử lý được các đối tượng mà các thông tin về chúng có thể là mơ hồ và không chắc chắn.

Các giá trị thuộc tính của đối tượng trong CSDL HĐT mờ rất phức tạp: bao gồm các giá trị ngôn ngữ, giá trị số, giá trị khoảng, tham chiếu đến đối tượng (đối tượng này có thể mờ), tuyển tập, ... Vì vậy, khi truy vấn dữ liệu trong CSDL HĐT với thông tin mờ và không chắc chắn thì vấn đề quan trọng nhất là làm thế nào tìm ra một phương pháp xử lý các giá trị mờ để từ đó xây dựng phương pháp so sánh giữa chúng. Có nhiều cách tiếp cận để xử lý các giá trị mờ được các tác giả trong và ngoài nước quan tâm nghiên cứu: Eiter và cộng sự (2001) đã giới thiệu một mô hình cơ sở đối tượng xác suất gọi là POB (Probabilistic Object Base). Đây là mô hình mở rộng mô hình CSDL HĐT xác suất của Kornatzky và Shimony (1994) với các chiến lược kết hợp xác suất để biểu

diễn và xử lý các thuộc tính đối tượng được thể hiện bởi các phân bố xác suất cận dưới và cận trên của một tập giá trị. Không chỉ mở rộng phép chọn trong mô hình của Kornatzky và Shimony, các tác giả còn xây dựng một đại số đối tượng hoàn chỉnh trên POB. Tuy nhiên, thiếu sót chính trong mô hình POB là không cho phép giá trị thuộc tính mờ và chưa thể hiện được phương thức lớp; Nguyễn Cát Hồ và cộng sự (2006) đã đề nghị một mô hình trong đó giá trị tập mờ của mỗi thuộc tính được biểu diễn bởi một nhãn ngôn ngữ và ngữ nghĩa của nó được xác định bởi một đại số gia tử. Trong mô hình này, tác giả đã định nghĩa độ đo tính mờ, ngữ nghĩa của các nhãn ngôn ngữ, tính tương tự và các quan hệ đối sánh giữa chúng, làm cơ sở cho một ngôn ngữ truy vấn dữ liệu mờ trên mô hình này.

Dựa theo cách tiếp cận của Đại số gia tử (ĐSGT), ngữ nghĩa ngôn ngữ có thể biểu thị bằng một lân cận các khoảng được xác định bởi độ đo tính mờ các giá trị ngôn ngữ của một thuộc tính được xem như là biến ngôn ngữ. Trên cơ sở đó, trong bài báo này xem miền giá trị thuộc tính đối tượng là đại số gia tử, chúng tôi đề xuất một phương pháp biến đổi về các giá trị khoảng và sau đó chuyển các giá trị khoảng này về đoạn con của $[0, 1]$. Từ đó, việc truy vấn và xử lý dữ liệu của các đối tượng mờ với thông tin mờ hồ, và không chắc chắn trở nên hiệu quả.

Trong phần 2 chúng tôi 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 xử lý giá trị khoảng và phương pháp so sánh trên các đoạn con của $[0, 1]$. Phần 4 trình bày phương pháp xử lý truy vấn với thông tin mờ và không chắc chắn

2. Đại số gia tử

Để xây dựng cách tiếp cận đại số gia tử, trong phần này chúng tôi trình bày tổng quan về một số nét cơ bản của đại số gia tử và khả năng biểu thị ngữ nghĩa dựa vào cấu trúc của đại số gia tử [1].

Chúng ta xét miền ngôn ngữ của biến chân lý TRUTH gồm các từ sau: $\text{Dom}(\text{TRUTH}) = \{\text{true, false, very true, very false, more-or-less true, more-or-less false, possibly true, possibly false, approximately true, approximately false, little true, little false, very possibly true, very possibly false.....}\}$, trong đó true, false là các từ nguyên thủy, các từ nhân (mordifier hay intensifier) very, more-or-less, possibly, approximately true, little gọi là các gia tử (hedges). Khi đó miền ngôn ngữ $T = \text{dom}(\text{TRUTH})$ có thể biểu thị như một đại số $\underline{X} = (X, C, H, \leq)$, trong đó C là tập các từ nguyên thủy được xem là các phần tử sinh. H là tập các gia tử được xem như là các phép toán một ngôi, quan hệ \leq trên các từ (các khái niệm mờ) là quan hệ thứ tự được "cảm sinh" từ ngữ nghĩa tự nhiên. Ví dụ dựa trên ngữ nghĩa, các quan hệ thứ tự sau là đúng: $\text{false} \leq \text{true}$, $\text{more true} \leq \text{very true}$ nhưng $\text{very false} \leq \text{more false}$, $\text{possibly true} \leq \text{true}$ nhưng $\text{false} \leq \text{possibly false}$, ... Tập X được sinh ra từ C bởi các phép tính trong H . Như vậy, mỗi phần tử của X sẽ có dạng biểu diễn $x = h_n h_{n-1} \dots h_1 x$, $x \in C$. Tập tất cả các phần tử được sinh ra từ một phần tử x được ký hiệu là $H(x)$. Nếu C có đúng hai từ **nguyên thủy mờ**, thì

một được gọi là phần tử sinh dương ký hiệu là c^+ , một gọi là phần tử sinh âm ký hiệu là c^- và ta có $c^- < c^+$. Trong ví dụ trên **True** là dương còn **False** là âm.

Như vậy, cho $X = (X, G, H, \leq)$ với $G = \{c^-, W, c^+\}$, $H = H^- \cup H^+$, 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 $p, q > 1$ là dãy các gia tử.

Định nghĩa 2.1 [1]. $f: X \rightarrow [0, 1]$ gọi là hàm ngữ nghĩa định lượng của X nếu $\forall h, k \in H^+$ hoặc $\forall h, k \in H^-$, và $\forall x, y \in X$ ta có:

$$\frac{|f(hx) - f(x)|}{|f(kx) - f(x)|} = \frac{|f(hy) - f(y)|}{|f(ky) - f(y)|} \#$$

Với đại số gia tử và hàm định lượng ngữ nghĩa ta có thể định nghĩa *tính mờ* của một khái niệm mờ. cho trước hàm định lượng ngữ nghĩa f của X, xét bất kỳ $x \in X$. Tính mờ của x khi đó được đo bằng đường kính của tập $f(H(x)) \subseteq [0, 1]$.

Định nghĩa 2.2 [1]: Hàm $fm: X \rightarrow [0, 1]$ 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

$$\sum_{-q \leq i \leq p, i \neq 0} fm(h_i u) = fm(u) \#$$

(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 $\forall x, y \in X, \forall h \in H$, 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 $\mu(h)$ gọi là độ đo tính mờ (fuzziness measure) của gia tử h .

Định nghĩa 2.3 [1]: Gọi fm là độ đo tính mờ trên đại số gia tử X, $f: X \rightarrow [0, 1]$. Với mọi $x \in X$, ta ký hiệu $I(x) \subseteq [0, 1]$ và $|I(x)|$ là độ dài của $I(x)$.

Một họ $J = \{I(x): x \in X\}$ 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 $\{I(h_i x): I = 1 \dots p+q\}$ 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.

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 [1]: 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 , được ký hiệu $u =_k v$, khi và chỉ

khí $I(u)$ và $I(v)$ cùng bao hàm trong một khoảng mờ mức k . Có nghĩa $\forall u, v \in X$, $u =_k v \Leftrightarrow \exists \Delta^k \in P^k : I(u) \subseteq \Delta^k$ và $I(v) \subseteq \Delta^k$.

3. Phương pháp xử lý giá trị thuộc tính

3.1 Giá trị thuộc tính

Giá trị thuộc tính của đối tượng có thể là một trong bốn trường hợp sau:

1. **Giá trị chính xác:** Giá trị có thể là các giá trị của các kiểu dữ liệu nguyên thủy như các kiểu số hoặc kiểu xâu ký tự, hoặc là tập hợp các giá trị nguyên thủy. Miền giá trị trong trường hợp này chúng ta có thể dễ dàng thao tác bằng việc sử dụng các phép toán ($\leq, \geq, =$) trong biểu thức điều kiện của câu hỏi truy vấn; hoặc chúng ta có thể xây dựng các điều kiện mờ để thực hiện truy vấn dữ liệu, ví dụ “cho biết tất cả đối tượng nhân viên có thu nhập thấp hơn lương trung bình”
2. **Giá trị không chính xác (hoặc mờ):** Trường hợp với giá trị không chính xác (hoặc mờ) rất phức tạp, thường thì nhân ngôn ngữ được sử dụng để biểu diễn cho những loại giá trị này [9]. Những kiểu giá trị không chính xác phải được xem xét theo ngữ nghĩa của các giá trị không chính xác. Chẳng hạn, một thực vật có tên là *húng tây*, đất thích hợp cho nó phát triển là *mùn* nhưng không chắc chắn nó cần mức độ ánh sáng là *nhẹ* hay *trung bình*, hoặc chiều cao của anh ta khoảng 2m, hoặc khoảng [0, 35] để biểu diễn khái niệm *trẻ* của con người.
3. **Đối tượng:** Trong trường hợp này giá trị thuộc tính có thể tham chiếu đến một đối tượng khác (đối tượng phức). Đối tượng mà nó tham chiếu đến có thể mờ. [9]
4. **Tuyển tập:** Thuộc tính có thể là tập các giá trị hoặc tập (set) các đối tượng [9]. Sự không chính xác của các thuộc tính này được chia thành 2 mức:
 - a. Tập (set) có thể mờ.
 - b. Thành viên của tập có thể là các giá trị mờ hoặc các đối tượng mờ

Trong bài báo này chúng tôi tập trung xét các giá trị thuộc tính trong trường hợp thứ hai và đặc biệt giá trị khoảng phục vụ cho việc xử lý và thao tác dữ liệu trong các phần sau.

3.2 Chuyển các giá trị khoảng về đoạn con [0,1]

Gọi $Dom(A_i) = [\min, \max]$ là miền giá trị thuộc tính đối tượng, trong đó \min và \max tương ứng là 2 giá trị nhỏ nhất và lớn nhất của $Dom(A_i)$.

Định nghĩa 3.1 [6]: $f: Dom(A_i) \rightarrow [0,1]$ và được xác định:

$$f(a) = \frac{a - \min}{\max - \min} \forall a \in Dom(A_i)$$

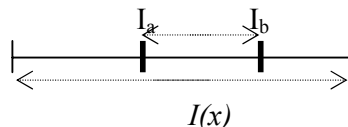
3.3 So sánh các giá trị khoảng với độ đo tính mờ trên đại số gia tử

Cho đại số gia tử $X = (X, G, H, \leq)$ và một giá trị khoảng $[a, b]$ sau khi đã

chuyển về đoạn con $[0, 1]$ (sử định nghĩa 3.1 để chuyển về đoạn $[0, 1]$ tương ứng). Vì tính mờ của x là một đoạn con $[0, 1]$ do đó để thực hiện việc so sánh chúng ta chỉ cần tìm phần giao của hai đoạn con $[0, 1]$ tương ứng này.

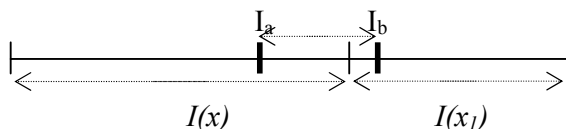
Đoạn con thứ nhất với $x \in X$, ta ký hiệu $I(x) \subseteq [0,1]$ và $|I(x)| = fm(x)$. Đoạn con thứ hai là $[I_a, I_b] = [f(a), f(b)] \subseteq [0,1]$. Khi đó phương pháp so sánh và tìm phần giao của hai đoạn con $[0,1]$ như sau:

Trường hợp 1: với mỗi đoạn $[I_a, I_b]$, nếu tồn tại $x \in X$ sao cho $[I_a, I_b] \subseteq I(x)$, thì $|a, b| =_{|x|} x$

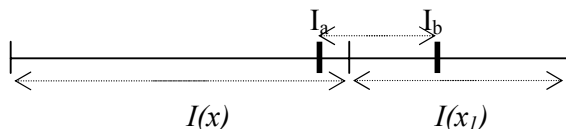


Trường hợp 2: với mỗi đoạn $[I_a, I_b]$, nếu tồn tại $x \in X$ sao cho $[I_a, I_b] \not\subseteq I(x)$. Trong trường hợp này chúng tôi chia thành 2 trường hợp nhỏ như sau:

- *Trường hợp a:* $\forall x, x_1 \in X$, và giả sử $x < x_1$ thì nếu $|[I_a, I_b] \cap I(x)| \geq |[I_a, I_b]|/c$, thì $|a, b| =_{|x|} x$

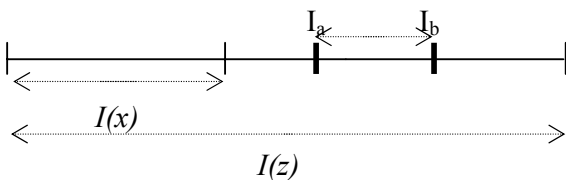


- *Trường hợp b:* $\forall x, x_1 \in X$, và giả sử $x < x_1$ thì nếu $|[I_a, I_b] \cap I(x_1)| \geq |[I_a, I_b]|/c$, thì $|a, b| =_{|x_1|} x_1$



Với c là số đoạn con $I(x_i) \subseteq [0,1]$, sao cho $([I_a, I_b] \cap I(x)) \neq \emptyset$

Trường hợp 3: với mỗi đoạn $[I_a, I_b]$ và nếu tồn tại $x \in X$ sao cho $([I_a, I_b] \cap I(x)) = \emptyset$. Nếu tồn tại $z \in X$ sao cho $[I_a, I_b] \subseteq I(z)$ và $I(x) \subseteq I(z)$ thì $|a, b| =_{|z|} z$



4. Truy vấn dữ liệu với thông tin mờ và không chắc chắn

4.1 Ví dụ: chúng ta xét cơ sở dữ liệu với 6 đối tượng sinh viên như sau

<u>SV1</u>	<u>SV2</u>	<u>SV3</u>
ten: Anh tuoi: 30 chieucao:[1.65,1.69]	ten: thanh tuoi: khoảng 25 chieucao:1.72	ten: nhàn tuoi: [25,30] chieucao:[1.7,1.8]
<u>SV4</u>	<u>SV5</u>	<u>SV6</u>
ten: Hoa tuoi: [35,45] chieucao:1.67	ten: Phuong tuoi: 45 chieucao: [1.8,1.9]	ten: Hồng tuoi: [27,32] chieucao: 1.6

4.2 Thuật toán xác định giá trị chân lý thuộc tính

Ý tưởng: Để xác định giá trị chân lý cho giá trị thuộc tính đối tượng. Chúng tôi chuyển đổi giá trị thuộc tính của đối tượng về đoạn con $[0, 1]$ và so sánh $x \in X$ (vì tính mờ x là đoạn con $[0, 1]$) với đoạn con này. Thuật toán dừng khi tìm đoạn con thỏa mãn điều kiện mờ.

Thuật toán 4.2 xác định giá trị chân lý của điều kiện mờ.

Vào: Tập các đối tượng $O(U)$, với $U = \{A_1, A_2, \dots, A_n\}$ là tập các thuộc tính của đối tượng

Ra: $O[A_i] =_k x$ là đúng

Phương pháp:

b1: Đặt $Dom(A_i) = [\min, \max]$

b2: **For** mỗi O **do**

if $O[A_i]$ không phải là giá trị khoảng **then** chuyển $O[A_i]$ thành khoảng $[a, b]$ tương ứng.

b3: **For** mỗi O **do**

// sử dụng hàm f để chuyển đoạn $[a, b]$ thành đoạn con $[0, 1]$

$$O[A_i] = [f(a), f(b)];$$

b4: **For** mỗi O **do**

if $O[A_i] \subseteq I(x)$ **then** ($O[A_i] =_k x$);

if $O[A_i] \not\subseteq I(x)$ **then**

if $(\exists x_1 \in X)$ và $(x < x_1)$ và $(|O[A_i] \cap I(x)| \geq |O[A_i]| / c)$ **then** ($O[A_i] =_k x$);

if $(\exists x_1 \in X)$ và $(x < x_1)$ và $(|O[A_i] \cap I(x_1)| \geq |O[A_i]| / c)$ **then** ($O[A_i] =_k x_1$)

if $(O[A_i] \cap I(x) = \emptyset)$ **then**

if $(\exists z \in X)$ và $(O[A_i] \subseteq I(z))$ và $(I(x) \subseteq I(z))$ **then** ($O[A_i] =_k z$)

4.3 Truy vấn

Truy vấn 1: Cho biết những sinh viên có *tuổi khả năng trẻ* trong ví dụ ở 4.1.

Vậy để trả lời được câu truy vấn 1 chúng ta thực hiện như sau:

Bước 1:

Gọi $\underline{X}_{tuoi} = (\underline{X}_{tuoi}, G_{tuoi}, H_{tuoi}, \leq)$ là đại số gia tử, với $G_{tuoi} = \{già, trẻ\}$, $H^+_{tuoi} = \{rất, hơn\}$, $H^-_{tuoi} = \{khả năng, ít\}$, $rất > hơn$ và $ít > khả năng$.

Chọn $W_{tuoi} = 0.6$, $fm(trẻ) = 0.6$, $fm(già) = 0.4$, $fm(rất) = 0.25$, $fm(hơn) = 0.25$, $fm(khả năng) = 0.25$, $fm(ít) = 0.25$, và $Dom(TUOI) = [0,100]$.

Ta có $fm(rất trẻ) = 0.15$, $fm(hơn trẻ) = 0.15$, $fm(ít trẻ) = 0.15$, $fm(khả năng trẻ) = 0.15$. Do $rất trẻ < hơn trẻ < trẻ < khả năng trẻ < ít trẻ$ nên ta có $I(rất trẻ) = [0, 0.15]$, $I(hơn trẻ) = [0.15, 0.30]$, $I(khả năng trẻ) = [0.30, 0.45]$, $I(ít trẻ) = [0.45, 0.60]$.

Bước 2: Áp dụng thuật toán 4.2

Sau khi thực hiện **b1** và **b2** của thuật toán 4.2 ta có kết quả như sau với $Dom(TUOI) = [0, 100]$:

<u>SV1</u>	<u>SV2</u>	<u>SV3</u>
ten: Anh tuoi: [0.30,0.30] chieucao:[1.65,1.69]	ten: thanh tuoi: [0.23,0.27] chieucao:1.72	ten: nhân tuoi: [0.25,0.30] chieucao:[1.7,1.8]
<u>SV4</u>	<u>SV5</u>	<u>SV6</u>
ten: Hoa tuoi: [0.35,0.45] chieucao:1.67	ten: Phuong tuoi: [0.45,0.45] chieucao: [1.8,1.9]	ten: Hồng tuoi: [0.27,0.32] chieucao: 1.6

b3: xác định các đối tượng thỏa mãn điều kiện

Vì $[0.35, 0.45] \subseteq I(khả năng trẻ)$, mà $[f(a), f(b)] = [0.35, 0.45]$ nên $[0.35, 0.45] = {}_2 khả năng trẻ$, và vì $[0.45, 0.45] \subseteq I(khả năng trẻ)$, mà $[f(a), f(b)] = [0.45, 0.45]$ nên $[0.45, 0.45] = {}_2 khả năng trẻ$.

Vậy có hai sinh viên có *tuổi khả năng trẻ* là SV4 và SV5

Truy vấn 2: Cho biết những sinh viên có *chiều cao ít cao*.

Vậy để trả lời được câu truy vấn 2 chúng ta thực hiện như sau:

Bước 1:

Gọi $\underline{X}_{chieucao} = (\underline{X}_{chieucao}, G_{chieucao}, H_{chieucao}, \leq)$ là đại số gia tử, với $G_{chieucao} = \{cao, thấp\}$, $H^+_{trongluong} = \{rất, hơn\}$, $H^-_{trongluong} = \{khả năng, ít\}$, $rất > hơn$ và $ít > khả năng$.

Chọn $W_{\text{chieucao}} = 0.6$, $f_m(\text{thấp}) = 0.6$, $f_m(\text{cao}) = 0.4$, $f_m(\text{rất}) = 0.35$, $f_m(\text{hơn}) = 0.25$, $f_m(\text{khả năng}) = 0.2$, $f_m(\text{ít}) = 0.2$

Ta có $f_m(\text{rất cao}) = 0.14$, $f_m(\text{hơn cao}) = 0.1$, $f_m(\text{ít cao}) = 0.08$, $f_m(\text{khả năng cao}) = 0.08$. Do $\text{ít cao} < \text{khả năng cao} < \text{cao} < \text{hơn cao} < \text{rất cao}$ nên ta có $I(\text{ít cao}) = [0.60, 0.68]$, $I(\text{khả năng cao}) = [0.68, 0.76]$, $I(\text{hơn cao}) = [0.76, 0.86]$, $I(\text{rất cao}) = [0.86, 1]$.

Bước 2: Áp dụng thuật toán 4.2

Sau khi thực hiện **b1** và **b2** của thuật toán 4.2 ta có kết quả như sau với $\text{Dom}(\text{CHIEUCAO}) = [1.0, 2.0]$:

<u>SV1</u>	<u>SV2</u>	<u>SV3</u>
ten: Anh tuoi: 30 chieucao:[0.65,0.69]	ten: thanh tuoi: khoảng 25 chieucao:[0.72,0.72]	ten: nhân tuoi: [25,30] chieucao:[0.7,0.8]
<u>SV4</u>	<u>SV5</u>	<u>SV6</u>
ten: Hoa tuoi: [35,45] chieucao:[0.67,0.67]	ten: Phuong tuoi: 45 chieucao: [0.8,0.9]	ten: Hồng tuoi: [27,32] chieucao: [0.6,0.6]

b3: xác định các đối tượng thỏa mãn điều kiện

Vì $[0.6, 0.6] \subseteq I(\text{ít cao})$, mà $[f(a), f(b)] = [0.6, 0.6]$ nên $[0.6, 0.6] =_2 \text{ít cao}$, và vì $[0.67, 0.67] \subseteq I(\text{ít cao})$, mà $[f(a), f(b)] = [0.67, 0.67]$ nên $[0.67, 0.67] =_2 \text{ít cao}$.

Ngoài ra, xét $[0.65, 0.69]$, ta có $[0.65, 0.69] \cap I(\text{ít cao}) = [0.65, 0.68]$ và $[0.65, 0.69] \cap I(\text{khả năng cao}) = [0.68, 0.69]$.

Vì $|[0.65, 0.69] \cap I(\text{ít cao})| \geq [0.65, 0.69]/2$, nên $[0.65, 0.69] =_2 \text{ít cao}$.

Vậy có ba sinh viên có chiều cao *ít cao* là SV1, SV4 và SV6

5. Kết luận

Trong bài báo này, chúng tôi đề xuất một phương pháp mới cho việc thao tác dữ liệu dạng khoảng trong CSDL HDT với thông tin mờ và không chắc chắn, đây là phương pháp tiếp cận ngữ nghĩa định lượng đại số gia tử. Với cách tiếp cận này, việc thao tác dữ liệu trở nên đơn giản vì các giá trị khoảng được chuyển thành các đoạn con tương ứng của $[0,1]$ và tính mờ của các phần tử trong đại số gia tử cũng là các đoạn con của $[0,1]$ nên việc so sánh này trở thành việc so sánh trên hai đoạn con của $[0,1]$.

Cuối cùng, chúng tôi xây dựng thuật toán cho việc xác định giá trị chân lý của điều kiện mờ trong câu truy vấn.

TÀI LIỆU THAM KHẢO

- [1] F. Bancilhon, R. Ramakrishnan, An Amateur's Introduction to Recursive Query Processing Strategies, *SIGMOD* 1986.
- [2] Nguyễn Cát Hồ, Lý thuyết tập mờ và công nghệ tính toán mềm. Hệ mờ, mạng nơron và ứng dụng, *NXB Khoa học kỹ thuật 2001*, trang 37-74.
- [3] N.C. Ho, Quantifying Hedge Algebras and Interpolation Methods in Approximate Reasoning, *Proc. of the 5th Inter. Conf. on Fuzzy Information Processing, Beijing, March 1-4 (2003)*, p105-112.
- [4] N.C. Ho, H.V. Nam, T.D Khang and L.H. Chau, Hedge Algebras, Linguistic-valued Logic and their Application to Fuzzy Reasoning, *Inter.J. of Uncertainty, Fuzziness and Knowledge-Based System*, 7 (1999) 347-361.
- [5] N. C. Ho, A model of relational databases with linguistic data of hedge algebras-based semantics, *Hội thảo quốc gia lần thứ ba về Nghiên cứu phát triển và ứng dụng CNTT và Truyền thông ICT.rda*, 2006.
- [6] Nguyễn Công Hào, Truy vấn trong cơ sở dữ liệu mờ theo cách tiếp cận đại số gia tử. *Báo cáo toàn văn hội thảo khoa học một số vấn đề thời sự trong công nghệ thông tin và ứng dụng toán học, Học viện kỹ thuật quân sự 2006*, 218-229.
- [7] Nguyễn Công Hào, Một phương pháp xử lý giá trị khoảng trong cơ sở dữ liệu mờ. *Tạp chí BCVT & CNTT kỳ 3 10/2007*, trang 67-73.
- [8] Z.Ma, Fuzzy Database Modeling of Imprecise and Uncertain Engineering Information, www.springerlink.com. © Springer-Verlag Berlin Heidelberg 2006.
- [9] Z.Ma, *Advances in Fuzzy Object-Oriented Databases: Modeling and Application*, Copyright © 2005, Idea Group Publishing.
- [10] L. Cuevasa, N. Marínb, O. Ponsb, M.A. Vilab. A fuzzy object-relational system, *Fuzzy Sets and Systems 159 (2008)*, p1500 – 1514.
- [11] Bordogna G., Pasi G., and Lucarella D., A Fuzzy object-oriented data model managing vague and uncertain information, *International Journal of Intelligent Systems 14 (1999)*, p623-651.
- [12] Cross V.V., Fuzzy set theory + object-oriented model= fuzzy object model, <http://www.sel.iit.nrc.ca/iitcolloq/season1998-99/cross.html>
- [13] Cao, T.H., Rosssiter, J.M. (2003). A deductive probabilistic and fuzzy object-oriented database language. *Fuzzy Sets and System*, 140, 129-150.
- [14] Cross, V.V. (2003). Defining fuzzy relationships in object models: Abstraction and interpretation. *International Journal of Fuzzy Sets and Systems*, 140, 5-27.
- [15] Na S. and Park S., Fuzzy object oriented data model and fuzzy association algebra, 2000.

- [16] Bosc P, Pivert O, *SQLf: a relational database language for fuzzy query*, *IEEE Transaction on Fuzzy Systems*, Vol. 3, No.1, 1995, 1-19.
- [17] Patrick Bosc, *Subqueries in SQLf, a Fuzzy Database Query Language*, *1995 IEEE*, 3636-3641.
- [18] David W. Embley, 1998, *Object Database Development: Concepts and Principles*, *Addison Wesley Longman*, [http:// osm7.cs.byu.edu](http://osm7.cs.byu.edu).
- [19] Fred R.McFadden, Jeffrey a.Hoffer, 1999, *Modern Database Management*, *Addision – Wesley*.
- [20] Michael B., William P., 1998, *Object - Oriented Modeling and Design for Database Applications*, *Prentice Hall, New Jersey*.