

## CÂU TRẢ LỜI GẦN ĐÚNG CHO TRUY VẤN NULL TRÊN MÔ HÌNH CƠ SỞ DỮ LIỆU HƯỚNG ĐỐI TƯỢNG MỜ THIỂU THÔNG TIN

ĐOÀN VĂN BAN<sup>1</sup>, TRƯƠNG CÔNG TUẤN<sup>2</sup>, ĐOÀN VĂN THẮNG<sup>3</sup>

<sup>1</sup>*Viện Công nghệ Thông tin, Viện Khoa Học và Công Nghệ Việt Nam*

<sup>2</sup>*Đại Học Khoa Học - Đại Học Huế*

<sup>3</sup>*Cao Đẳng CNTT Hữu Nghị Việt Hàn*

**Tóm tắt.** Lập luận tương tự là một công cụ suy diễn quan trọng trong nghiên cứu trí tuệ nhân tạo. Truy vấn Null là những truy vấn mà nó nhận một câu trả lời Null từ hệ thống cơ sở dữ liệu, thường là vì thông tin không đầy đủ trong cơ sở dữ liệu (CSDL). Lập luận tương tự thật sự rất là hữu ích để tạo ra một câu trả lời gần đúng cho truy vấn Null. Lập luận tương tự sử dụng phụ thuộc hàm mờ để trả lời các truy vấn Null trên cơ sở dữ liệu quan hệ mờ đã được nghiên cứu. Bài báo đề xuất một cách tiếp cận để có được một câu trả lời gần đúng cho truy vấn Null trên mô hình CSDL hướng đối tượng (HDT) với thông tin mờ và không chắc chắn dựa trên đại số gia tử (ĐSGT). Phương pháp tiếp cận ở đây dựa trên lân cận ngữ nghĩa của ĐSGT để xây dựng các xác định mờ thuộc tính và xác định mờ phương thức.

**Abstract.** Reasoning by analogy is an important inference tool in the study of artificial intelligence. Null queries are queries that find an answer to Null from the database system, often because of the incompleteness of information in the database. Analogical reasoning are really useful to generate an approximate answers for Null queries. Analogical reasoning using fuzzy functional dependencies to answer null queries on fuzzy relational databases have been studied. In this paper, we present an approach to obtain an approximate answer for Null queries on Object-oriented database model with fuzzy information and uncertainty based on hedge algebra. Our approach was based on the concept semantic neighborhood of hedge algebra to build attribute fuzzy determination and method fuzzy determination.

### 1. MỞ ĐẦU

Mô hình CSDL quan hệ và hướng đối tượng mờ chứa thông tin không đầy đủ, không chính xác hoặc không chắc chắn đã được nghiên cứu rộng rãi trong những năm gần đây [1-11]. Để biểu diễn những thông tin mờ trong mô hình dữ liệu, có nhiều hướng tiếp cận cơ bản: mô hình dựa trên quan hệ tương tự [5] và mô hình phân bố khả năng [1], ...

Một câu truy vấn được xem là truy vấn Null khi nó nhận được câu trả lời Null từ CSDL. Câu trả lời Null có thể được sinh ra do nhiều nguyên nhân khác nhau. Nguyên nhân thứ nhất, là khi dữ liệu trong CSDL không thỏa mãn điều kiện của câu truy vấn. Một nguyên nhân tiếp theo, hoặc là do dữ liệu không tồn tại trong CSDL, chẳng hạn ta truy vấn các bộ (tuple) dữ liệu từ quan hệ  $R$ , nhưng chúng không tồn tại trong CSDL. Một nguyên nhân nữa, đó là do thông tin không đầy đủ trong CSDL, nguyên nhân này thường là do giá trị thuộc tính của

một số đối tượng bị thiếu. Và ngoài ra còn có rất nhiều nguyên nhân khác nhau. Bài báo sẽ tập trung nghiên cứu truy vấn Null được sinh ra do nguyên nhân thông tin không đầy đủ trong CSDL. Hầu hết trong các CSDL truyền thống khi truy vấn dữ liệu trong trường hợp này thì câu trả lời Null luôn được tạo ra. Tuy nhiên, việc có được câu trả lời gần đúng thì bao giờ cũng tốt hơn khi nhận được một câu trả lời Null.

Các tác giả trong [5] đã đưa ra mô hình lý thuyết sử dụng lập luận tương tự để trả lời các truy vấn Null cho mô hình CSDL quan hệ mở dựa vào phân bố khả năng. Tuy nhiên, trong mô hình này, Dutta không xem xét dữ liệu với các miền tương tự rời rạc, nơi dữ liệu có thể được mô hình hóa thông qua các mối quan hệ tương tự.

Dựa vào những ưu điểm của cấu trúc đại số gia tử (ĐSGT) [7], các tác giả đã nghiên cứu mô hình CSDL quan hệ [8-10] và hướng đối tượng [2, 3] mở dựa trên cách tiếp cận của ĐSGT, trong đó ngữ nghĩa ngôn ngữ được lượng hóa bằng các ánh xạ định lượng của ĐSGT. Theo cách tiếp cận của ĐSGT, ngữ nghĩa ngôn ngữ có thể được biểu thị bằng một lân cận các khoảng được xác định bởi độ đo tính mờ của các giá trị ngôn ngữ của một thuộc tính với vai trò là biến ngôn ngữ. Bài báo sử dụng lý thuyết lập luận tương tự [5] để trả lời các truy vấn Null cho mô hình CSDL HDT mở dựa theo cách tiếp cận ĐSGT.

Bài báo được trình bày như sau: Mục 2 trình bày một số khái niệm cơ bản liên quan đến ĐSGT làm cơ sở cho các mục tiếp theo. Mục 3 trình bày các định nghĩa về xác định mờ thuộc tính, xác định mờ phương thức và cơ chế lập luận tương tự cho việc trả lời những truy vấn Null. Mục 4 trình bày các qui trình sử dụng xác định mờ thuộc tính, xác định mờ phương thức và một số ví dụ minh họa cho ý tưởng của hướng tiếp cận này, và cuối cùng là kết luận.

## 2. MỘT SỐ KHÁI NIỆM CƠ BẢN

Sau đây là một số khái niệm về ánh xạ định lượng [7] và cách thức xác định các hệ lân cận ngữ nghĩa định lượng [8, 10].

### 2.1. Đại số gia tử

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

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

1.  $fm$  là độ đo mờ đầy đủ trên  $\mathbf{X}$ , 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  $\forall x, y \in \mathbf{X}, \forall h \in H$ , 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 2.2.** (Hàm định lượng ngữ nghĩa  $v$ )

Cho  $fm$  là độ đo tính mờ trên  $\mathbf{X}$ , hàm định lượng ngữ nghĩa  $v$  trên  $\mathbf{X}$  được định nghĩa như sau:

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

2. Nếu  $1 \leq j \leq p$  thì:  $v(h_jx) = v(x) + Sign(h_jx) \times \left[ \sum_{i=1}^j fm(h_ix) - \omega(h_jx)fm(h_jx) \right]$ ,

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

trong đó:

$$\omega(h_jx) = \frac{1}{2} [1 + Sign(h_jx)Sign(h_qh_jx)(\beta - \alpha)] \in \{\alpha, \beta\}.$$

*Phân hoạch dựa trên độ đo tính mờ của các giá trị ngôn ngữ trong đại số gia tử*

Vì độ đo tính mờ của các từ là một khoảng của đoạn  $[0, 1]$  và họ các khoảng như vậy của các từ có cùng độ dài sẽ tạo thành phân hoạch của  $[0, 1]$ . Phân hoạch ứng với độ dài từ lớn hơn sẽ mịn hơn và khi độ dài lớn vô hạn thì độ dài của các khoảng phân hoạch giảm dần về 0.

**Định nghĩa 2.3.** Gọi  $fm$  là độ đo tính mờ trên ĐSGT  $\mathbf{X}$ , Với mỗi  $x \in \mathbf{X}$ , 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  $[0, 1]$  gắn với  $x$  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  $\{I(h_ix) : i = 1 \dots p + q\}$  được định nghĩa là phân hoạch của  $I(x)$  sao cho thỏa mãn điều kiện:  $|I(h_ix)| = fm(h_ix)$  và  $\{I(h_ix) : i = 1 \dots p + q\}$  là tập sắp thứ tự tuyến tính, tức là: hoặc  $I(h_1x) < I(h_2x) < I(h_3x) < \dots < I(h_{p+q}x)$ , hoặc  $I(h_1x) > I(h_2x) > I(h_3x) > \dots > I(h_{p+q}x)$ .

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

**Định nghĩa 2.4.** Cho  $P^k = \{I(x) : x \in X_k\}$  với  $X_k = \{x \in \mathbf{X} : |x| = k\}$  là một phân hoạch của  $[0, 1]$ . Ta nói rằng  $u$  bằng  $v$  theo mức  $k$  trong  $P^k$ , được ký hiệu  $u \approx_k v$ , khi và chỉ khi  $I(u)$  và  $I(v)$  cùng thuộc một khoảng trong  $P^k$ . Có nghĩa  $\forall x, y \in \mathbf{X}, u \approx_k v \Leftrightarrow \exists \Delta^k \in P^k : I(u) \subseteq \Delta^k$  và  $I(v) \subseteq \Delta^k$ .

## 2.2. Lân cận mức $k$

Xét một CSDL  $\{U; Const\}$ , trong đó  $U = \{A_1, A_2, \dots, A_n\}$  là tập vũ trụ các thuộc tính,  $Const$  là một tập các ràng buộc dữ liệu của CSDL. Mỗi thuộc tính  $A$  được gắn với một miền giá trị thuộc tính, ký hiệu là  $Dom(A)$ , trong đó một số thuộc tính cho phép nhận các giá trị ngôn ngữ trong lưu trữ trong CSDL hay trong các câu hỏi truy vấn và được gọi là thuộc tính

ngôn ngữ. Những thuộc tính còn lại được gọi là thuộc tính thực hay kinh điển. Thuộc tính thực  $A$  được gắn với một miền giá trị kinh điển, ký hiệu là  $D_A$ . Thuộc tính ngôn ngữ  $A$  sẽ được gắn một miền giá trị kinh điển  $D_A$  và một miền giá trị ngôn ngữ  $LD_A$  hay là tập các phần tử của một ĐSGT. Để bảo đảm tính nhất quán trong xử lý ngữ nghĩa dữ liệu trên cơ sở thông nhất kiểu dữ liệu của thuộc tính ngôn ngữ, mỗi thuộc tính ngôn ngữ sẽ được gắn với một ánh xạ định lượng  $v_A : LD_A \rightarrow D_A$  được xác định bởi một bộ tham số định lượng của  $A$ . Như vậy, mỗi giá trị ngôn ngữ  $x$  của  $A$  sẽ được gán một nhân giá trị thực  $v_A \in D_A$  được xem như giá trị đại diện của  $x$ . Việc đánh giá độ tương tự giữa các dữ liệu của một thuộc tính  $A$  được dựa trên khái niệm lân cận mức  $k$  của một giá trị ngôn ngữ, với  $k$  là số nguyên dương.

Các tác giả trong [8, 10] đã lấy các khoảng mờ của các phần tử độ dài  $k$  làm độ tương tự giữa các phần tử, nghĩa là các phần tử mà các giá trị đại diện của chúng thuộc cùng một khoảng mờ mức  $k$  là tương tự mức  $k$ . Tuy nhiên, theo cách xây dựng các khoảng mờ mức  $k$ , giá trị đại diện của các phần tử  $x$  có độ dài nhỏ hơn  $k$  luôn là đầu mút của các khoảng mờ mức  $k$ . Do vậy, khi xác định lân cận mức  $k$  chúng ta mong muốn các giá trị đại diện như vậy phải là điểm trong của lân cận mức  $k$ .

Chúng ta luôn luôn giả thiết rằng mỗi tập  $H^-$  và  $H^+$  chứa ít nhất 2 gia tử. Xét  $\mathbf{X}_k$  là tập tất cả các phần tử độ dài  $k$ . Dựa vào khoảng mờ mức  $k$  và mức  $k+1$  các tác giả [8, 10] đã xây dựng một phân hoạch của miền  $[0, 1]$  như sau:

(1) *Độ tương tự mức 1*: Với  $k = 1$ , các khoảng mờ mức 1 gồm  $I(c^-)$  và  $I(c^+)$ . Các khoảng mờ mức 2 trên khoảng  $I(c^+)$  là  $I(h_{-q}c^+) \leq I(h_{-q+1}c^+) \dots \leq I(h_{-2}c^+) \leq I(h_{-1}c^+) \leq v_A(c^+) \leq I(h_1c^+) \leq I(h_2c^+) \leq \dots \leq I(h_{p-1}c^+) \leq I(h_pc^+)$ . Khi đó, ta xây dựng phân hoạch về độ tương tự mức 1 gồm các lớp tương đương sau:  $S(\mathbf{0}) = I(h_pc^-)$ ;  $S(c^-) = I(c^-) \setminus [I(h_{-q}c^-) \cup I(h_pc^-)]$ ;  $S(\mathbf{W}) = I(h_{-q}c^-) \cup I(h_{-q}c^+)$ ;  $S(c^+) = I(c^+) \setminus [I(h_{-q}c^+) \cup I(h_pc^+)]$  và  $S(\mathbf{1}) = I(h_pc^+)$ .

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

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

Bằng cách tương tự như vậy, có thể xây dựng các phân hoạch các lớp tương tự mức  $k$  bất kỳ. Tuy nhiên, trong thực tế ứng dụng theo [6] thì  $k \leq 4$ , tức có tối đa 4 gia tử tác động liên tiếp lên phần tử nguyên thủy  $c^-$  và  $c^+$ . Các giá trị rõ và các giá trị mờ gọi là có độ tương tự mức  $k$  nếu các giá trị đại diện của chúng cùng nằm trong một lớp tương tự mức  $k$ .

**Lân cận mức  $k$  của khái niệm mờ**: Giả sử phân hoạch các lớp tương tự mức  $k$  là các khoảng  $S(x_1), S(x_2), \dots, S(x_m)$ . Khi đó, mỗi giá trị ngôn ngữ  $fu$  chỉ và chỉ thuộc về một lớp tương tự, chẳng hạn đó là  $S(x_i)$  và nó gọi là lân cận mức  $k$  của  $fu$  và ký hiệu là  $FRN_k(fu)$ .

### 3. LẬP LUẬN XẤP XỈ ĐỂ TRẢ LỜI TRUY VẤN NULL

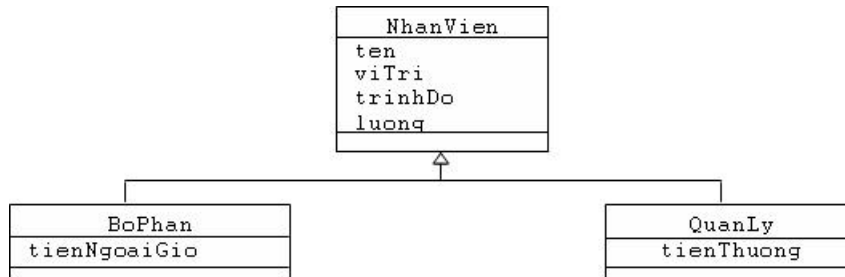
Cho một lớp mờ  $C$  bao gồm tập các thuộc tính và phương thức  $C = (\{a_1, a_2, \dots, a_n\}, \{M_1, M_2, \dots, M_m\})$ .

Lập luận tương tự là nhằm nhận được mức độ tương tự giữa các đối tượng nguồn và đối tượng đích đồng thời suy ra được một số các tính chất của đối tượng đích dựa vào các tính

chất tương tự của các đối tượng nguồn. Trong CSDL mờ, thông tin không đầy đủ, không chắc chắn thường câu trả lời cho truy vấn Null là câu trả lời Null do thiếu các giá trị thuộc tính của một số đối tượng. Quan hệ tương tự sẽ rất phù hợp để có được các câu trả lời gần đúng cho câu truy vấn Null.

Gọi  $S$  là đối tượng nguồn và  $T$  là đối tượng đích, tập đối tượng nguồn và đích có tính chất tương tự nhau là  $P$ . Theo lập luận tương tự, nếu  $S$  có tính chất  $P'$  thì chúng ta suy ra  $T$  có thể cũng có  $P'$  dựa vào tính chất  $P$  có trong cả  $S$  và  $T$  [11].

**Ví dụ 3.1.** Xét mô hình CSDL HDT mờ bao gồm một tập các đối tượng của lớp *NhanVien* như sau



Hình 3.1. Cơ sở dữ liệu hướng đối tượng

Sau đây là một số thể hiện của lớp *BoPhan*, lớp *QuanLy* để đơn giản ta giới hạn bảng dữ liệu chỉ gồm những thuộc tính.

Bảng 1. Thể hiện của lớp *BoPhan* và *QuanLy*

(a) Thể hiện của lớp *BoPhan*

BoPhan					
iDBP	ten	viTri	trinhDo	tienNgoaiGio	luong
iD1	An	ketoan	CD	30	90
iD2	Binh	ketoan	CD	27	80
iD3	Hoa	vanhanh	TC	24	70
iD4	Hue	vanhanh	TH	23	65

(b) Thể hiện của lớp *QuanLy*

QuanLy					
iDQL	ten	viTri	trinhDo	tienThuong	luong
iD1	Hai	Gd	CD	30	90
iD2	Minh	Qlvh	TC	27	
iD3	Tam	Qisp	CD	15	25
iD4	Nam	Qlht	CH	22	30

Ví dụ, xét các đối tượng được cho trong Ví dụ 3.1. Giả sử rằng giá trị thuộc tính *luong* của đối tượng *Binh* là thiếu và câu truy vấn như sau: “Cho biết lương của *Binh*” ( $luong(Binh)?$ ). Khi đó câu trả lời là Null được tạo ra. Ta có thể giả thiết rằng có một câu trả lời gần đúng cho câu trả lời Null đó là lương của *Binh* vào khoảng 90. Bởi vì công việc của *Binh* là kế toán và có thể tìm thấy lương của công việc kế toán từ tập các đối tượng trong Ví dụ 3.1 là 90, đây chính là lương của đối tượng *An* (*Binh* và *An* có thuộc tính giống nhau là *kế toán*, thuộc tính này là một yếu tố xác định lương). Tất nhiên, câu trả lời xấp xỉ có thể không đúng vì do nhiều yếu tố ảnh hưởng đến lương của nhân viên. Tuy nhiên, việc có được trả lời xấp xỉ có thể tốt hơn là không có câu trả lời khi truy vấn dữ liệu. Ngoài ra, trong mô hình CSDL HDT mờ ta giả thiết rằng các giá trị thuộc tính đối tượng là đầy đủ và không rỗng. Nếu các giá trị thuộc tính đối tượng bị thiếu (hoặc rỗng) thì trong trường hợp này sẽ không có câu trả lời khi truy vấn.

Có hai vấn đề sẽ được giải quyết trong quá trình suy diễn này: Thứ nhất đó là xác định tính chất tương tự nhau  $P$  giữa tập đối tượng nguồn và đích, chẳng hạn thuộc tính *viTri* trong Ví dụ 3.1; vấn đề thứ hai đó là thuộc tính  $Q$  trong cả  $S$  và  $T$  có thể xác định thuộc tính không rõ  $Q'$ , chẳng hạn  $Q$  là *viTri* và  $Q'$  là *luong*. Vấn đề này liên quan đến phụ thuộc hàm mờ (xác định mờ) giữa các thuộc tính.

Trong mô hình CSDL HDT mờ, giá trị thuộc tính của đối tượng rất phức tạp. Trong [3], đã trình bày giá trị thuộc tính có 4 trường hợp: giá trị rõ, giá trị mờ, tham chiếu đến đối tượng (đối tượng này có thể mờ) và tuyển tập (collection). Mục tiêu của bài báo là đưa ra giải pháp để có câu trả lời gần đúng, vì vậy sẽ tập trung xem xét giá trị thuộc tính trong trường hợp thứ nhất và hai, đó là: giá trị chính xác (giá trị rõ) và giá trị không chính xác (giá trị mờ), trên cơ sở xem giá trị rõ là trường hợp riêng của giá trị mờ. Giá trị mờ thường rất phức tạp và nhãn ngôn ngữ thường được sử dụng để biểu diễn cho những loại giá trị này. Miền giá trị thuộc tính mờ là hợp của hai thành phần:

$$Dom(a_i) = CDom(a_i) \cup FDom(a_i) (1 \leq i \leq n)$$

Trong đó:

- $CDom(a_i)$ : Miền giá trị rõ của thuộc tính  $a_i$ .
- $FDom(a_i)$ : Miền giá trị mờ của thuộc tính  $a_i$ .

### 3.1. Xấp xỉ mức $k$

Trên cơ sở khái niệm lân cận, chúng ta đưa ra định nghĩa về xấp xỉ mức  $k$  của các thuộc tính đối tượng. Xấp xỉ mức  $k$  được định nghĩa như sau:

**Định nghĩa 3.1.** Cho lớp mờ  $C$  xác định trên tập thuộc tính  $A$  và tập phương thức  $M$ ,  $a_i \subseteq A (1 \leq i \leq n)$ ,  $o_1, o_2 \in C$ . Ta nói rằng  $o_1.a_i$  xấp xỉ bậc  $k$   $o_2.a_i$  được ký hiệu  $o_1.a_i \approx_k o_2.a_i$  nếu  $o_1.a_i$  và  $o_2.a_i$  cùng thuộc một về một lớp tương tự  $FRN_k(fu)$ . Trong đó  $FRN_k(fu)$  là một khoảng phân hoạch các lớp tương tự mức  $k$ .

**Ví dụ 3.2.** Dựa trên mô hình CSDL HDT mờ đã cho ở Ví dụ 3.1, ta xét ĐSGT của biến ngôn ngữ *tienNgoaiGio* là thuộc tính của lớp *BoPhan*, trong đó  $D_{tienNgoaiGio} = [0, 30]$ , tập các phần tử sinh là  $\{0, \text{thấp}, W, \text{cao}, 1\}$ , tập các gia tử là  $\{\text{ít}, \text{khá}, \text{hơn}, \text{rất}\}$ , và  $FD_{tienNgoaiGio} = H_{tienNgoaiGio}(\text{cao}) \cup H_{tienNgoaiGio}(\text{thấp})$ .

Chọn  $fm(\text{cao}) = 0.35$ ,  $fm(\text{thấp}) = 0.65$ ,  $\mu(\text{khá}) = 0.25$ ,  $\mu(\text{ít}) = 0.20$ ,  $\mu(\text{hơn}) = 0.15$  và  $\mu(\text{rất}) = 0.40$ . Ta phân hoạch đoạn  $[0, 30]$  thành 5 khoảng tương tự mức 1 như sau:

$$fm(\text{rất cao}) * 30 = 0.35 * 0.35 * 30 = 3.675. \text{ vậy } S(1) * 30 = (26.325, 30].$$

$$(fm(\text{khá cao}) + fm(\text{hơn cao})) * 30 = (0.25 * 0.35 + 0.15 * 0.35) * 30 = 4.2,$$

$$\text{vậy } S(\text{cao}) * 30 = (22.125, 26.325].$$

$$(fm(\text{ít thấp}) + fm(\text{ít cao})) * 30 = (0.25 * 0.65 + 0.25 * 0.35) * 30 = 7.5,$$

$$\text{vậy } S(W) * 30 = (14.625, 22.125].$$

$$(fm(\text{khá thấp}) + fm(\text{hơn thấp})) * 30 = (0.25 * 0.65 + 0.15 * 0.65) * 30 = 7.8,$$

$$\text{vậy } S(\text{thấp}) * 30 = (6.825, 14.625], \text{ và } S(0) * 30 = [0, 6.825].$$

Từ đó, ta có lân cận mức 1 của các lớp tương tự như sau:  $FRN_1(0) = [0, 6.825]$ ,  $FRN_1(\text{thấp}) = (6.825, 14.625]$ ,  $FRN_1(W) = (14.625, 22.125]$ ,  $FRN_1(\text{cao}) = (22.125, 26.325]$  và  $FRN_1(1) = (26.325, 30]$ .

Vậy ta nói  $o_1.a_4 \approx_1 o_2.a_4$  vì  $o_1.a_4 = 30 \in FRN_1(0)$  và  $o_2.a_4 = 27 \in FRN_1(0)$ ; hoặc  $o_3.a_4 \approx_1 o_4.a_4$  vì  $o_3.a_4 = 24 \in FRN_1(\text{cao})$  và  $o_4.a_4 = 23 \in FRN_1(\text{cao})$ .

### 3.2. Xác định mờ thuộc tính

Trong phần này, sẽ đề xuất cách xác định mờ thuộc tính dựa trên lân cận ngữ nghĩa của ĐSGT. Khái niệm xác định mờ thuộc tính này là sự mở rộng của khái niệm phụ thuộc hàm mờ của mô hình CSDL quan hệ mờ dựa trên ĐSGT [9]. Ta định nghĩa về xác định mờ thuộc tính (**A**tttribute **F**uzzy **D**etermination, viết tắt là *AFD*) như sau.

**Định nghĩa 3.2.** Cho lớp mờ  $C$  với tập thuộc tính  $U, X, Y \subseteq U$  và  $X$  là tập các thuộc tính nhận giá trị đơn (giá trị rõ hoặc mờ),  $k$  là mức phân hoạch. Chúng ta định nghĩa  $X$  xác định mờ thuộc tính  $Y$  theo mức  $k$ , ký hiệu  $X \sim_{>k}^{afd} Y$  khi và chỉ khi  $\forall o_1, o_2 \in C$  nếu  $o_1.X \approx_k o_2.X$  thì  $o_1.Y \approx_k o_2.Y$  ( $o_1.X$  – giá trị của  $o_1$  trên  $X$ ).

$U$  là tập các thuộc tính của lớp  $C$ ;  $X, Y, W \subseteq U$ ,  $k$  là mức phân hoạch.

Dựa vào các kết quả [4, 8, 10], ta có các luật suy dẫn *AFD* như sau:

**Luật 1.1:** Phản xạ. Nếu  $Y \subseteq X$  thì  $X \sim_{>k}^{afd} Y$ .

**Luật 1.2:** Tăng trưởng. Nếu  $X \sim_{>k}^{afd} Y$  và  $Y \sim_{>k}^{afd} Z$  thì  $XZ \sim_{>k}^{afd} YZ$ .

**Luật 1.3:** Bắc cầu. Nếu  $X \sim_{>k}^{afd} Y$  và  $Y \sim_{>k}^{afd} Z$  thì  $X \sim_{>k}^{afd} Z$ .

**Luật 1.4:** Gia tăng. Nếu  $X \sim_{>k}^{afd} Y$  và  $Z \subseteq W$  thì  $XW \sim_{>k}^{afd} YZ$ .

**Ví dụ 3.3.** Trong Ví dụ 3.2 đã xây dựng ngữ nghĩa định lượng cho thuộc tính ngôn ngữ *tienNgoaiGio*. Trong ví dụ này, sẽ xây dựng ngữ nghĩa định lượng cho thuộc tính ngôn ngữ *luong* của lớp *BoPhan*. Xét ĐSGT của biến ngôn ngữ *luong*, trong đó  $D_{luong} = [0, 100]$ , các phần tử sinh là  $\{0, \text{thấp}, W, \text{cao}, 1\}$ , tập các gia tử là  $\{\text{ít}, \text{khá}, \text{hơn}, \text{rất}\}$ ,  $FD_{luong} = H_{luong}(\text{cao}) \cup H_{luong}(\text{thấp})$ .

Chọn  $fm(\text{cao}) = 0.60$ ,  $fm(\text{thấp}) = 0.40$ ,  $\mu(\text{khá}) = 0.15$ ,  $\mu(\text{ít}) = 0.25$ ,  $\mu(\text{hơn}) = 0.25$  và  $\mu(\text{rất}) = 0.35$ . Ta phân hoạch đoạn  $[0, 100]$  thành 5 khoảng tương tự mức 1 như sau:

$$fm(\text{rất cao}) * 100 = 0.35 * 0.60 * 100 = 21. \text{ vậy } S(1) * 21 = (79, 100].$$

$$(fm(\text{khá cao}) + fm(\text{hơn cao})) * 100 = (0.25 * 0.60 + 0.15 * 0.60) * 100 = 24,$$

$$\text{vậy } S(\text{cao}) * 100 = (55, 79].$$

$$(fm(\text{ít thấp}) + fm(\text{ít cao})) * 100 = (0.25 * 0.60 + 0.25 * 0.40) * 100 = 25,$$

$$\text{vậy } S(W) * 100 = (30, 55].$$

$$(fm(\text{khá thấp}) + fm(\text{hơn thấp})) * 100 = (0.25 * 0.40 + 0.15 * 0.40) * 100 = 16,$$

$$\text{vậy } S(\text{thấp}) * 100 = (14, 30], \text{ và } S(0) * 100 = [0, 14].$$

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

Dễ dàng nhận thấy rằng *tienNgoaiGio* xác định mờ thuộc tính *luong* theo mức  $k=1$ , nghĩa là  $tienNgoaiGio \sim_{>k}^{afd} luong$ .

### 3.3. Xác định mờ phương thức

Việc xác định sự gắn kết giữa phương thức đối tượng đối với các thuộc tính mờ được xác định thông qua việc phương thức sử dụng các thuộc tính đó để đọc hay sửa đổi. Việc xác định như vậy được gọi là xác định mờ phương thức (**M**ethod **F**uzzy **D**etermination, viết tắt là *MF*D).

**Định nghĩa 3.3.** Cho  $M_j$  là một phương thức của một lớp  $C$ ,  $X$  là tập các thuộc tính nhận giá trị đơn (giá trị rõ hoặc mờ) (có thể đọc hoặc sửa đổi),  $k$  là mức phân hoạch. Ta định nghĩa  $X$  xác định mờ phương thức  $M_j$  theo mức  $k$ , được ký hiệu  $X \sim_{>k}^{mfd} M_j$ , nếu  $X$  được sử dụng bởi  $M_j$  theo mức phân hoạch  $k$ .

Dựa vào các kết quả [4], ta có các luật suy dẫn *MFD* như sau:

**Luật 2.1:** Nếu  $X \sim_{>k}^{mfd} M_j$  và  $Y \sim_{>k}^{mfd} M_j$  thì  $XY \sim_{>k}^{mfd} M_j$ .

**Luật 2.2:** Nếu  $X \sim_{>k}^{mfd} M_i$  và  $X \sim_{>k}^{mfd} M_j$  thì  $X \sim_{>k}^{mfd} M_i M_j$ .

**Luật 2.3:** Nếu  $X \sim_{>k}^{mfd} M_j$  và  $Y \subseteq X$  thì  $Y \sim_{>k}^{mfd} M_j$ .

**Ví dụ 3.4.** Giả sử lớp NhanVien có phương thức *thucLinh()* thì việc xác định mờ phương thức *thucLinh()* được thể hiện trong hai lớp kế thừa là *BoPhan* và *QuanLy* như sau:

Đối với lớp *BoPhan*:  $\{luongNgoaiGio, luong\} \sim_{>k}^{mfd} thucLinh()$ .

Đối với lớp *QuanLy*:  $\{tienThuong, luong\} \sim_{>k}^{mfd} thucLinh()$ .

#### 4. THUẬT TOÁN TRẢ LỜI GẦN ĐÚNG TRUY VẤN NULL

Trong phần này, sẽ nêu giải pháp để trả lời gần đúng cho những truy vấn Null dựa trên xác định mờ thuộc tính và xác định mờ phương thức.

**a. Đối với truy vấn thuộc tính, ta có lời giải thuật toán như sau:**

**Thuật toán NQAO** (Null Queries for Attributes Objects)

**Vào:** Một lớp  $C$  với  $m$  thuộc tính và  $p$  phương thức;  $O = \{o_1, o_2, \dots, o_n\} \in C$ .

Trong đó:  $(o_1, o_2, \dots, o_{n-1}).[a_1, a_2, \dots, a_m]$  và  $o_n.[a_1, a_2, \dots, a_{m-1}]$  là được định nghĩa (có nghĩa là giá trị dữ liệu tồn tại) và  $o_n.a_m$  là chưa được định nghĩa (có nghĩa là giá trị dữ liệu bị thiếu).

**Ra:** Trả về giá trị tương tự gần đúng cho  $o_n.a_m$ .

**Phương pháp:**

// Khởi tạo các giá trị.

(1) **For**  $i = 1$  **to**  $m$  **do**

(2) **Begin**

(3)  $G_{a_i} = \{0, c_{a_i}^-, W, c_{a_i}^+, 1\}$ ;  $H_{a_i} = H_{a_i}^+ \cup H_{a_i}^-$ . 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.

(4)  $D_{a_i} = [\min_{a_i}, \max_{a_i}] // \min_{a_i}, \max_{a_i}$ : giá trị nhỏ nhất và lớn nhất miền trị giá trị  $a_i$ .

(5)  $FD_{a_i} = H_{a_i}(c_{a_i}^+) \cup H_{a_i}(c_{a_i}^-)$

(6) **End**

// Xây dựng các khoảng tương tự mức  $k = 1$

(7) **For**  $i = 1$  **to**  $m$  **do**

(8) **For**  $j = 1$  **to**  $8$  **do**

(9) *Xây dựng các khoảng tương tự mức  $k$ :  $S_{a_i}^k(x_j)$ ;*

// Xây dựng lân cận mức  $k = 1$  của  $o.a_i$

(10) **For**  $i = 1$  **to**  $n$  **do**

(11) **For**  $j = 1$  **to**  $m$  **do**

(12) **Begin**

(13)  $t = 0$ ;



(14)       **Repeat**  
(15)            $t = t + 1$ ;  
(16)       **Until**  $o_i.a_j \in S_{a_i}^k(x_t)$  or  $t > 8$   
(17)            $FRN_i^k(o_i.a_j) = FRN_i^k(o_i.a_j) \cup S_{a_i}^k(x_t)$   
(18)       **End**  
//Xây dựng thuộc tính chưa định nghĩa  $o_n.a_m$   
(19)  $result = \emptyset$ ;  
(20) **For**  $i = 1$  **to**  $n-1$  **do**  
(21)       **If**  $o_n.a_{m-1} = o_i.a_{m-1}$  **then**  $result = result \cup FRN_i^k(o_i.a_m)$ ;  
(22) **Return**  $result$ ;

**Ví dụ 4.1.** Cho biết lương của tất cả nhân viên quản lý

Từ CSDL HDT đã cho ở Ví dụ 3.1 ta nhận thấy lương của đối tượng *Minh* là chưa định nghĩa (bị thiếu), vì vậy ta cần tìm giá trị tương tự gần chúng cho lương của đối tượng này. Sử dụng thuật toán **NQAO**, như sau:

**Bước (1)–(6):** Chọn độ đo tính mờ cho các phần tử sinh và gia tử cho 2 thuộc tính *tienThuong* và *luong*.

$D_{tienThuong} = [0, 30]$  và  $D_{luong} = [0, 100]$ .  $FD_{tienThuong}$  và  $FD_{luong}$  có cùng tập xâu giống nhau với tập các phần tử sinh là  $\{0, \text{thấp}, W, \text{cao}, 1\}$  và tập các gia tử là  $\{\text{ít}, \text{khá}, \text{hơn}, \text{rất}\}$ .

**Bước (7)–(18):** Xây dựng mức phân hoạch cho 2 thuộc tính *tienThuong* và *luong*.

Đối với thuộc tính *tienThuong*: Chọn  $fm(\text{cao}) = 0.35$ ,  $fm(\text{thấp}) = 0.65$ ,  $\mu(\text{khá}) = 0.25$ ,  $\mu(\text{ít}) = 0.20$ ,  $\mu(\text{hơn}) = 0.15$  và  $\mu(\text{rất}) = 0.40$ . Ta phân hoạch đoạn  $[0, 30]$  thành 5 khoảng tương tự mức 1 như sau:

$$\begin{aligned} fm(\text{rất cao}) * 30 &= 0.35 * 0.35 * 30 = 3.675. \text{ vậy } S(1) * 30 = (26.325, 30]. \\ (fm(\text{khá cao}) + fm(\text{hơn cao})) * 30 &= (0.25 * 0.35 + 0.15 * 0.35) * 30 = 4.2, \\ \text{vậy } S(\text{cao}) * 30 &= (22.125, 26.325]. \\ (fm(\text{ít thấp}) + fm(\text{ít cao})) * 30 &= (0.25 * 0.65 + 0.25 * 0.35) * 30 = 7.5, \\ \text{vậy } S(W) * 30 &= (14.625, 22.125]. \\ (fm(\text{khá thấp}) + fm(\text{hơn thấp})) * 30 &= (0.25 * 0.65 + 0.15 * 0.65) * 30 = 7.8, \\ \text{vậy } S(\text{thấp}) * 30 &= (6.825, 14.625], \text{ và } S(0) * 30 = [0, 6.825]. \end{aligned}$$

Từ đó, ta có lần cận mức 1 của các lớp tương tự như sau:  $FRN_1(0) = [0, 6.825]$ ,  $FRN_1(\text{thấp}) = (6.825, 14.625]$ ,  $FRN_1(W) = (14.625, 22.125]$ ,  $FRN_1(\text{cao}) = (22.125, 26.325]$  và  $FRN_1(1) = (26.325, 30]$ .

Đối với thuộc tính *luong*:  $fm(\text{cao}) = 0.60$ ,  $fm(\text{thấp}) = 0.40$ ,  $\mu(\text{khá}) = 0.15$ ,  $\mu(\text{ít}) = 0.25$ ,  $\mu(\text{hơn}) = 0.25$  và  $\mu(\text{rất}) = 0.35$ . Ta phân hoạch đoạn  $[0, 100]$  thành 5 khoảng tương tự mức 1 như sau:

$$\begin{aligned} fm(\text{rất cao}) * 100 &= 0.35 * 0.60 * 100 = 21, \text{ vậy } S(1) * 21 = (79, 100]. \\ (fm(\text{khá cao}) + fm(\text{hơn cao})) * 100 &= (0.25 * 0.60 + 0.15 * 0.60) * 100 = 24, \\ \text{vậy } S(\text{cao}) * 100 &= (55, 79]. \\ (fm(\text{ít thấp}) + fm(\text{ít cao})) * 100 &= (0.25 * 0.60 + 0.25 * 0.40) * 100 = 25, \\ \text{vậy } S(W) * 100 &= (30, 55]. \\ (fm(\text{khá thấp}) + fm(\text{hơn thấp})) * 100 &= (0.25 * 0.40 + 0.15 * 0.40) * 100 = 16, \\ \text{vậy } S(\text{thấp}) * 100 &= (14, 30], \text{ và } S(0) * 100 = [0, 14]. \end{aligned}$$

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

**Bước (19)–(22):** Dựa vào định nghĩa 3.2 xác định mờ thuộc tính, nhận thấy rằng việc xác định mờ thuộc tính  $tienThuong \overset{afd}{\sim} >_1 luong$  là thỏa mãn.

Vậy giá trị tương tự đúng cho thuộc tính lương của đối tượng  $Minh \in FRN_1(1) = (79,100]$ .

**b. Đối với truy vấn phương thức, ta có lời giải thuật toán như sau:**

**Thuật toán NQMO** (Null Queries for Method Objects)

**Vào:** Một lớp  $C$  với  $m$  thuộc tính và  $p$  phương thức;  $O = o_1, o_2, \dots, o_n \in C$ .

Trong đó:  $(o_1, o_2, \dots, o_{n-1}).[a_1, a_2, \dots, a_m]$  và  $o_n.[a_1, a_2, \dots, a_{m-1}]$  là được định nghĩa (có nghĩa là giá trị dữ liệu tồn tại) và  $o_n.a_m$  là chưa được định nghĩa (có nghĩa là giá trị dữ liệu bị thiếu).

**Ra:** Trả về giá trị tương tự gần đúng cho phương thức.

**Phương pháp:**

// Khởi tạo các giá trị.

(1) **For**  $i = 1$  **to**  $m$  **do**

(2) **Begin**

(3)  $G_{a_i} = \{0, c_{a_i}^-, W, c_{a_i}^+, 1\}$ ;  $H_{a_i} = H_{a_i}^+ \cup H_{a_i}^-$ . 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.

(4)  $D_{a_i} = [min_{a_i}, max_{a_i}] // min_{a_i}, max_{a_i}$ : giá trị nhỏ nhất và lớn nhất miền trị giá trị  $a_i$ .

(5)  $FD_{a_i} = H_{a_i}(c_{a_i}^+) \cup H_{a_i}(c_{a_i}^-)$

(6) **End**

//Xây dựng các khoảng tương tự mức  $k = 1$

(7) **For**  $i = 1$  **to**  $m$  **do**

(8) **For**  $j = 1$  **to**  $8$  **do**

(9) *Xây dựng các khoảng tương tự mức  $k$ :  $S_{a_i}^k(x_j)$ ;*

//Xây dựng lân cận mức  $k = 1$  của  $o.a_i$

(10) **For**  $i = 1$  **to**  $n$  **do**

(11) **For**  $j = 1$  **to**  $m$  **do**

(12) **Begin**

(13)  $t = 0$ ;

(14) **Repeat**

(15)  $t = t + 1$ ;

(16) **Until**  $o_i.a_j \in S_{a_i}^k(x_t)$  or  $t > 8$

(17)  $FRN_i^k(o_i.a_j) = FRN_i^k(o_i.a_j) \cup S_{a_i}^k(x_t)$

(18) **End**

//Xây dựng thuộc tính chưa định nghĩa  $o_n.a_m$

(19) **For**  $i = 1$  **to**  $n-1$  **do**

(20) **If**  $o_n.a_{m-1} = o_i.a_{m-1}$  **then**  $FRN_i^k(o_n.a_m) = FRN_i^k(o_i.a_m)$ ;

(21) *Chọn hàm kết nhập DSGT  $F_{HA}$  cho phương thức;*

(22)  $result = \emptyset$ ;

(23) **For**  $i = 1$  **to**  $n$  **do**

(24)  $result = result \cup \bigcup_{j=1}^m F_{HA}(FRN_i^k(o_i.a_j))$ ;

(25) **Return**  $result$ ;

**Ví dụ 4.2.** Cho biết thực lĩnh của tất cả nhân viên quản lý.

Sử dụng thuật toán **NQMO**, như sau:

- **Bước (1)-(20)** tính toán giống như Ví dụ 4.1 ở trên. Kết quả được cho ở Bảng 2(a).
- **Bước (21)-(25)**: Hàm tích hợp ĐSGT trong trường hợp này phép hợp của hai ĐSGT *tienThuong* và *luong*. Kết quả cho ở Bảng 2(b).

Bảng 2. Các kết quả thực hiện ví dụ 4.2

(a) Kết quả thực hiện bước (1) – (20)				(b) Kết quả thực hiện bước (21) – (25)			
QuanLy				QuanLy			
iDQL	tienThuong	luong	thucLinh()	iDQL	tienThuong	luong	thucLinh()
iD1	30 (26.325, 30]	90 (79, 100]		iD1	30 (26.325, 30]	90 (79, 100]	(79, 130]
iD2	25 (26.325, 30]	(79, 100]		iD2	25 (26.325, 30]	(79, 100]	(79, 130]
iD3	15 (14.625, 22.125]	25 (14, 30]		iD3	15 (14.625, 22.125]	25 (14, 30]	(14, 52.125]
iD4	22 (14.625, 22.125]	30 (14, 30]		iD4	22 (14.625, 22.125]	30 (14, 30]	(14, 52.125]

**Định lý:** Thuật toán **NQAO** và **NQMO** luôn dừng và đúng đắn.

*Chứng minh:*

1. *Chứng minh tính dừng:* Tập các thuộc tính, phương thức của đối tượng là hữu hạn ( $n$ ,  $p$ ,  $m$  là hữu hạn) nên thuật toán sẽ dừng sau khi duyệt xong tất cả các đối tượng.
2. *Chứng minh tính đúng đắn:* Thuật toán **NQAO** và **NQMO** sử dụng Định nghĩa 6 xác định mờ thuộc tính và Định nghĩa 7 xác định mờ phương thức. Khái niệm xác định mờ thuộc tính và xác định mờ phương thức là sự mở rộng của khái niệm phụ thuộc hàm mờ trong CSDL quan hệ mờ. Sự suy diễn logic này dựa vào sự suy diễn của phụ thuộc mờ trong CSDL quan hệ mờ. Tính đúng đắn của phép suy diễn bởi phụ thuộc hàm mờ trong CSDL quan hệ mờ đã được chứng minh. ■

Độ phức tạp thuật toán: Thuật toán **NQAO** và **NQMO** có độ phức tạp tính toán là  $O(n*m)$ , với  $n$  là số đối tượng và  $m$  tập thuộc tính.

## 5. KẾT LUẬN

Bài báo đã đề xuất một cách tiếp cận để trả lời các truy vấn Null sử dụng lập luận tương tự trong mô hình CSDL HDT mờ dựa trên cách tiếp cận của ĐSGT. Dựa trên lân cận ngữ nghĩa của ĐSGT, bài báo đã đưa ra các định nghĩa xác định mờ thuộc tính và xác định mờ phương thức. Việc xây các bước thực hiện trong hai thuật toán **NQAO** và **NQMO** nhằm tìm ra một câu trả lời gần đúng cho các truy vấn Null đối với thuộc tính lẫn phương thức mà có các giá trị dữ liệu bị thiếu từ một CSDL không đầy đủ. Dựa trên cơ sở các xác định mờ thuộc tính và xác định mờ phương thức, các dạng chuẩn lớp mờ và phương pháp tách lớp về các dạng chuẩn trong lớp đối tượng mờ sẽ được trình bày trong các nghiên cứu tiếp theo.

## TÀI LIỆU THAM KHẢO

- [1] V. Biazzo, R. Giugno, T. Lukasiewicz, V.S.Subrahmanian, Temporal probabilistic object bases, *IEEE Transaction on Knowledge and Engineering*, 2002, (921-939).

- [2] D.V.Thang, D.V.Ban, Query data with fuzzy information in object-oriented databases an approach the semantic neighborhood of hedge algebras, *International Journal of Computer Science and Information Security (IJCSIS)*, **9** (5) (2011) 37–42.
- [3] D.V.Thang, D.V.Ban, Query data with fuzzy information in object-oriented databases an approach interval values, *International Journal of Computer Science and Information Security (IJCSIS)*, **9** (2) (2011) 1–6.
- [4] Đ.V.Ban, H.C.Hà, V.Đ.Quảng, Phương pháp phát hiện các lớp bộ phận trong lớp đối tượng mờ, *Tạp chí Tin học và điều khiển học* **26** (4) (2010) 321–331.
- [5] S. Dutta, Approximate reasoning by analogy to answer null queries, *Int. J. of Appr. Reasoning* (5) (1991) 373–398.
- [6] 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 (105-112).
- [7] N.C.Ho, *Lý thuyết tập mờ và công nghệ tính toán mềm, Hệ mờ, mạng nơron và ứng dụng*, NXB Khoa học & Kỹ thuật, 2001 (37–74).
- [8] 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* **24** (4) (2008) 281–294.
- [9] Nguyễn Công Hào, Một số dạng chuẩn trong cơ sở dữ liệu mờ theo cách tiếp cận đại số gia tử, *Tạp chí Tin học và Điều khiển học* **17** (2007) 101–107.
- [10] Phương Minh Nam, Trần Thái Sơn, Về một cơ sở dữ liệu mờ và ứng dụng trong quản lý tội phạm hình sự, *Tạp chí Tin học và Điều khiển học* **22** (1) (2006) 67–73.
- [11] S.L. Wang, and T.J. Huang, Analogical reasoning to answer null queries in fuzzy object-oriented data model, *Proc. of the 6th IEEE Inter. Conf. on Fuzzy Systems*, Barcelona, Spain, July 1-5, 1997 (31-36).

*Nhận bài ngày 16 - 6 - 2011*

*Nhận lại sau sửa ngày 20 - 10 - 2011*