Tạp chí Tin học và Điều khiển học, T.22, S.1 (2006), 25–36

VỀ MỘT CƠ SỞ DỮ LIỆU MỜ VÀ ỨNG DỤNG TRONG QUẢN LÝ TỘI PHẠM HÌNH SỰ

PHƯƠNG MINH NAM[1], TRẦN THÁI SƠN[2]

[1] *Cục Công nghệ tin học, Bộ Công an*
[2] *Viện Công nghệ thông tin, Viện KH&CN Việt Nam*

Abstract. In this paper, a new approach to fuzzy data meaning in fuzzy DB based on theory of hedges algebra will be presented. With this approach, the meaning of fuzzy values in DB will be presented through ordered relations, that will be easier and more rational for comparative operators of relational algebra in management systems of relational DB and therefore, fuzzy information processing will be suggestive simpler than in existed approaches.

Tóm tắt. Trong bài báo này, chúng tôi trình bày về một cách tiếp cận mới trong việc xử lý thông tin mờ trong cơ sở dữ liệu mờ trên cơ sở lý thuyết đại số gia tử. Với cách tiếp cận này, ngữ nghĩa của các giá trị mờ trong cơ sở dữ liệu được biểu diễn qua quan hệ thứ tự, thuận tiện và hợp lý cho các phép toán so sánh của đại số quan hệ trong các hệ quản trị Cơ sở dữ liệu mô hình quan hệ và vì thế việc xử lý thông tin mờ sẽ thuận tiện và hợp lý hơn so với một số cách tiếp cận trước đây.

1. MỞ ĐẦU

Trong thực tiễn quản lý các tội phạm phạm hình sự, bên cạnh những dữ liệu kinh điển với kiểu dữ liệu số nguyên, số thực hay ký tự còn thường xuyên xuất hiện các dữ liệu với ngữ nghĩa mờ, không chắc chắn. Ví dụ, người cung cấp tin về tội phạm thường mô tả chiều cao, độ tuổi, màu tóc, nước da,... bằng các cụm từ ngôn ngữ với ngữ nghĩa không chắc chắn, không chính xác, ở đây được gọi là dữ liệu hay thông tin kiểu mờ. Tuy nhiên nó vẫn mang một lượng thông tin có giá trị. Những thông tin kiểu dạng như vậy được lưu trong hồ sơ, trong tri thức kinh nghiệm của các cán bộ cảnh sát hình sự và được sử dụng trong trao đổi, hội họp nghiệp vụ hình sự. Điều này ít nhất có hai ý nghĩa: (i) Mặc dù các thông tin kiểu mờ là không chính xác, nhưng chúng vẫn có giá trị thực tiễn và do đó vẫn có thể và cần được sử dụng trong tìm kiếm thông tin trong cơ sở dữ liệu (CSDL) bằng một phương pháp phù hợp nào đó. (ii) CSDL trong lĩnh vực quản lý tội phạm hình sự chỉ có ý nghĩa thực tiễn khi nó cho phép lưu trữ và xử lý được dữ liệu kinh điển và dữ liệu mờ một cách đồng thời.

Vấn đề đặt ra là cần có cách tiếp cận xử lý ngữ nghĩa các thông tin kiểu mờ một cách hiệu quả và phù hợp.

Có nhiều cách tiếp cận xử lý thông tin mờ. Chẳng hạn, cách tiếp cận truyền thống là biểu diễn ngữ nghĩa ngôn ngữ bằng tập mờ và xử lý dữ liệu dựa trên lý thuyết tập mờ. Theo cách này, trong cài đặt người ta phải khai báo kiểu dữ liệu mờ và phải có phương pháp xử

lý riêng khi đối sánh các kiểu dữ liệu khác nhau ([25, 26, 27]). Hiệu quả của phương pháp, đầu tiên, phụ thuộc vào việc biểu diễn ngữ nghĩa ngôn ngữ bằng hàm thuộc và đây là một bài toán phức tạp, mang tính chủ quan. Thứ hai, nó phụ thuộc vào phương pháp xử lý đối sánh dữ liệu, chẳng hạn phụ thuộc vào việc chọn ngưỡng lát cắt λ của tập mờ và cách thức đánh giá độ tương tự giữa hai hàm thuộc hay giữa hàm thuộc và một giá trị thực. Một cách tiếp cận khác là biểu diễn ngữ nghĩa ngôn ngữ dựa trên lý thuyết khả năng. Trong cách tiếp cận này, các từ ngôn ngữ, về hình thức, cũng được biểu diễn bằng hàm như trong cách tiếp cận tập mờ, nhưng ngữ nghĩa dựa trên lý thuyết khả năng (Possibility).

Mặc dù những cách tiếp cận như vậy đối với CSDL với thông tin không chắc chắn đều thu hút sự quan tâm lớn của giới nghiên cứu, cả hai cách tiếp cận trên đều có những bất tiện như việc biểu diễn ngữ nghĩa trừu tượng, phức tạp và việc thực hiện đối sánh dữ liệu cũng gặp nhiều khó khăn. Điều này nhìn chung không thuận tiện đối với CSDL quản lý tội phạm hình sự với đặc điểm là tỷ lệ khá lớn thông tin hình sự là không chắc chắn.

Trong bài báo này, dựa theo cách tiếp cận đại số đến ngữ nghĩa ngôn ngữ được nghiên cứu lần đầu tiên bởi Nguyễn Cát Hồ và Wolfgang Wechler ([16, 17]), chúng tôi đề cập một giải pháp khác trong đó các thông tin mờ biểu thị bằng ngôn ngữ được hiểu như là những phần tử trong đại số gia tử. Trong cấu trúc đại số như vậy, ngữ nghĩa của các từ được biểu diễn bằng quan hệ thứ tự và do vậy nó tương thích với việc so sánh lớn hơn hay nhỏ hơn trong CSDL. Sử dụng ánh xạ định lượng ngữ nghĩa của đại số gia tử chuyển các giá trị ngôn ngữ thành các giá trị thực bảo toàn thứ tự ngữ nghĩa có thể cho phép thao tác dữ liệu trên miền thực của các thuộc tính trong CSDL trong mối liên hệ với ngữ nghĩa của ngôn ngữ. Giải pháp như vậy cho phép thao tác dữ liệu cùng kiểu kinh điển và xử lý ngữ nghĩa ngôn ngữ đơn giản hơn.

Bố cục bài báo như sau. Trong Mục 2, các khái niệm cơ bản liên quan đến định lượng ngữ nghĩa của một thuộc tính mờ được định nghĩa, chỉ ra các tính chất; từ đó, một kiểu Cơ sở dữ liệu mờ được xác định trong Mục 3, cho phép thao tác thuận tiện, hợp lý với các giá trị thuộc tính mờ. Mục 4 là một ví dụ cụ thể về cơ sở dữ liệu mờ ứng dụng trong lĩnh vực hình sự.

2. NGỮ NGHĨA ĐỊNH LƯỢNG CỦA MIỀN GIÁ TRỊ NGÔN NGỮ CỦA THUỘC TÍNH TRONG CSDL

Giá trị của nhiều thuộc tính trong CSDL mô tả đối tượng có thể nhận các giá trị không chắc chắn như thuộc tính TUỔI, MÀU TÓC, NƯỚC DA,... và chúng được gọi là thuộc tính mờ. Giả sử A là một thuộc tính mờ và ký hiệu $LDom(A)$ là miền các giá trị ngôn ngữ của A. Có thể xem A là một biến ngôn ngữ và $LDom(A)$ là tập cơ sở của đại số gia tử (ĐSGT) được gán với A, ký hiệu là $\mathcal{DS}(A)$.

Để hiểu cách tiếp cận ngữ nghĩa ngôn ngữ dựa trên ĐSGT trong việc xây dựng CSDL với thông tin mờ, chúng ta nhắc lại một số tính chất cấu trúc cơ bản của ĐSGT. (Do phạm vi ứng dụng, từ đây trở về sau, chúng ta chỉ nhắc đến ĐSGT tuyến tính và đối xứng, tức ĐSGT với tập phần tử sinh chỉ bao gồm 2 phần tử và các gia tử là sắp thứ tự tuyến tính).

ĐSGT $\mathcal{DS}(A)$ được viết dưới dạng $\mathcal{DS}(A) = (X, G, H, \leqslant)$, trong đó $X = LDom(A)$, $G = \{c^-, c^+\}$ là tập các phần tử sinh, H là tập các gia tử được xem là các phép toán một ngôi và $\leqslant$ là quan hệ thứ tự ngữ nghĩa trên X. Gọi I là toán tử đồng nhất trên X và một cách tổng

quát cũng xem nó là gia tử nhân tạo.

Trong bài bài báo này chúng ta luôn giả thiết các tập G và tập các gia tử âm $H^- = \{h_{-1}, ..., h_{-q}\}$ và gia tử dương $H^+ = \{h_1, ..., h_p\}$ đều là tuyến tính với thư tự như sau: $h_1 < \cdots < h_p$ và $h_{-1} < \cdots < h_{-q}$, trong đó $p, q > 1$ và $H = H^- \cup H^+$.

Ký hiệu $H(x)$ là tập các phần tử của X được sinh từ x trong ĐSGT $ĐS(A)$.

Mệnh đề 2.1. *Đối với mọi ĐSGT $ĐS(A) = (X, G, H, \leqslant)$ chúng ta đều có các tính chất sau:*

p1) $H(hx) \subseteq H(x)$, *đối với mọi gia tử* h *và với mọi* $x \in X$.

p2) $\forall x \in X, \forall h, k \in H,\ h \neq k$ *chúng ta có* $H(hx) \cap H(kx) = \emptyset$.

p3) $\forall x \in X$, *ta có* $H(x) = \cup\{H(hx) : h \in H \cup I\}$.

p4) $\forall x \in X, \forall h, k \in H$, *nếu* $hx \leqslant kx$ *thì* $H(hx) \leqslant H(kx)$.

p5) $\forall x \in X$, *ta có hoặc* $h_p x < \cdots < h_1 x < x < h_{-1} x < \cdots < h_{-q} x$
hoặc $h_{-q} x < \cdots < h_{-1} x < x < h_1 x < \cdots < h_p x$.

Trong thực tế rất nhiều ứng dụng đòi hỏi ngữ nghĩa định lượng của giá trị ngôn ngữ. Vì vậy, trong [15] các tác giả đã đưa ra và nghiên cứu khái niệm ánh xạ định lượng ngữ nghĩa của một ĐSGT.

Định nghĩa 2.1. Một cách tổng quát, ánh xạ định lượng ngữ nghĩa là một ánh xạ $f : X \to [0, 1]$ thỏa mãn các tính chất sau:

Q1) f là một song ánh, tức là ánh xạ 1 - 1.

Q2) f là ánh xạ bảo toàn quan hệ thứ tự ngữ nghĩa của X, tức là $x < y \Rightarrow f(x) < f(y)$, và ta luôn có $f(0) = 0, f(1) = 1$.

Như vậy, nhìn chung một ĐSGT có thể có nhiều ánh xạ định lượng ngữ nghĩa. Tuy nhiên có thể thấy một ánh xạ tổng quát như vậy không có mối liên hệ trực tiếp đến ngữ nghĩa của các từ ngôn ngữ, chẳng hạn như chúng không mô tả định lượng được sự gần nhau giữa ngữ nghĩa của các từ. Để thiết lập mối liên hệ giữa tính chất của ánh xạ định lượng ngữ nghĩa với mức độ gần nhau giữa các từ, bài báo [15] đã đưa ra khái niệm độ đo tính mờ của các từ ngôn ngữ. Nghĩa là mỗi từ ngôn ngữ sẽ được gắn một giá trị trong $[0, 1]$ để chỉ mức độ mờ của từ đó và do đó [15] đưa ra định nghĩa sau:

Định nghĩa 2.2. Một ánh xạ $fm : X \to [0, 1]$ được gọi là một độ đo tính mờ của X nếu:

(fm1) $fm(c^-) + fm(c^+) = 1$ và $\forall u \in X,\ \sum\limits_{h \in H} fm(hu) = fm(u)$;

(fm2) nếu $H(x) = \{x\}$ thì $fm(x) = 0$, đặc biệt ta có $fm(0) = fm(W) = fm(1) = 0$;

(fm3) với mọi $x, y \in X,\ \forall h \in H,\ \dfrac{fm(hx)}{fm(x)} = \dfrac{fm(hy)}{fm(y)}$, nghĩa là tỷ lệ thức này không phụ thuộc vào các từ ngôn ngữ x và y hay nó chỉ phụ thuộc vào gia tử h, do vậy giá trị này được gọi là độ đo tính mờ của gia tử h và được ký hiệu là $\mu(h)$.

Ta có thể giải thích ý nghĩa của định nghĩa trên như sau. Điều kiện (fm1) nói rằng ngữ nghĩa của các giá trị ngôn ngữ đủ phủ miền giá trị thực mà nó mô tả với lưu ý rằng ta sử dụng miền khoảng [0,1] để chuẩn hoá. Trong (fm2), đẳng thức $H(x) = \{x\}$ nói rằng x không thể biến đổi ngữ nghĩa bằng các gia tử và do đó nó là giá trị kinh điển. Điều kiện (fm2) thể hiện rằng một giá trị kinh điển luôn luôn có độ đo tính mờ bằng không. Còn (fm3) nói rằng mức độ tác động của mỗi gia tử, một cách tương đối, không thay đổi từ giá trị ngôn ngữ này

sang giá trị ngôn ngữ khác.

Như vậy có thể thấy định nghĩa trên thể hiện khách quan ý nghĩa của độ đo tính mờ của các khái niệm mờ. Độ đo tính mờ của giá trị ngôn ngữ có một số tính chất căn bản sau:

Mệnh đề 2.2. *Giả sử độ đo tính mờ của giá trị ngôn ngữ fm và của gia tử μ được định nghĩa như trong Định nghĩa 2.2. Khi đó ta có:*

1) $fm(hx) = \mu(h)fm(x),\ \forall x \in X^*$.

2) $fm(c^-) + fm(c^+) = 1$.

3) $\sum\limits_{-q\leqslant i\leqslant p, i\neq 0} fm(h_i c) = fm(c)$, *trong đó* $c \in \{c^-, c^+\}$.

4) $\sum\limits_{-q\leqslant i\leqslant p, i\neq 0} fm(h_i x) = fm(x),\ x \in X$.

5) $\sum\{\mu(h_i) | -q \leqslant i \leqslant -1\} = \alpha$ và $\sum\{\mu(h_i) | 1 \leqslant i \leqslant p\} = \beta$ *trong đó* $\alpha + \beta = 1$.

Có thể biểu diễn mỗi độ đo tính mờ $fm(x)$ như là một khoảng của đoạn $[0, 1]$, thứ tự giữa các khoảng tương thích với thứ tự giữa các từ ngôn ngữ x. Gọi $K(x)$ là khoảng tương ứng với x có độ dài $fm(x)$.

Một cách tự nhiên chúng ta mong muốn là giá trị ngữ nghĩa định lượng của x sẽ là một phần tử nằm trong khoảng $K(x)$. Dựa trên ý tưởng trực quan đó, tác giả trong [15] đưa ra định nghĩa sau về ánh xạ định lượng ngữ nghĩa dựa trên độ đo tính mờ fm.

Định nghĩa 2.3. Giả sử $ĐS(A) = (X, G, H, \leqslant)$ là một ĐSGT và $fm(c^-), fm(c^+), \mu(h)$ là những độ đo tính mờ tương ứng của các phần tử sinh âm và dương c^-, c^+ và của các gia tử h thỏa mãn các điều kiện 2) và 5) trong Mệnh đề 2.2. Giả sử ν là ánh xạ được xác định bằng các công thức giải tích sau:

1) $\nu(W) = \mathcal{K} = fm(c^-)$, $\nu(c^-) = \mathcal{K} - \alpha fm(c^-) = \beta fm(c^-)$, $\nu(c^+) = \mathcal{K} + \alpha fm(c^+)$.

2) $\nu(h_j x) = \nu(x) + \mathrm{Sgn}(h_j x)\{\sum\limits_{i=\mathrm{Sgn}(j)}^{j} \mu(h_i) fm(x) - \omega(h_j x)\mu(h_j) fm(x)\}$.

trong đó $\omega(h_j x) = \frac{1}{2}[1 + \mathrm{Sgn}(h_j x)\mathrm{Sgn}(h_p h_j x)(\beta - \alpha)] \in \{\alpha, \beta\}$ với mọi $j, -q \leqslant j \leqslant p$ và $j \neq 0$.

Ánh xạ ν được xác định như vậy được gọi là ánh xạ định lượng ngữ nghĩa được cho bởi độ đo tính mờ fm. Hơn nữa, trong trường hợp ĐSGT là đầy đủ, trong [7] các tác giả đã chứng minh rằng tập ảnh $\nu(H(x))$ là trù mật trong đoạn $K(x)$. Điều này chứng tỏ các phần tử trong khoảng $K(x)$ về mặt ngữ nghĩa định lượng gần phần tử $\nu(x)$. Điều này có ý nghĩa quan trọng trong việc đưa ra khái niệm lân cận của phần tử $\nu(x)$.

Như vậy mỗi thuộc tính mờ A có thể được gắn với một ánh xạ định lượng ngữ nghĩa $\nu_{fm,A}$ được cho bởi fm. Trong một số thí dụ về sau, để cho gọn, đôi khi chúng tôi sử dụng ký hiệu ν_A thay cho $\nu_{fm,A}$ với sự ngầm hiểu rằng có một độ đo tính mờ fm đã xác định ở đó.

Nếu ký hiệu X_k là tập con của X gồm các từ có độ dài k, thì các tính chất 2), 3) và 4) của Mệnh đề 2.2 bảo đảm rằng:

K1) Các khoảng $K(h_i x), i \in [-q^\wedge p] = \{j : -q \leqslant j \leqslant p \text{ và } j \neq 0\}$ là phân hoạch có bậc $k+1$ của khoảng $K(x)$ và $\nu_{fm,A}(x)$ luôn luôn là một điểm mút của phân hoạch một phía của nó là các khoảng $K(h_i x)$ với $-q \leqslant i < 0$ còn phía kia của điểm mút là các khoảng $K(h_i x)$

với $0 < i \leqslant p$.

K2) Các khoảng $K(x), x \in X_k$ là một phân hoạch bậc $k = |x|$ của [0,1], tức là chúng rời nhau và phủ đoạn $[0, 1]$.

3. CƠ SỞ DỮ LIỆU MỜ VỚI NGỮ NGHĨA DỰA TRÊN ĐSGT

Cho một tập thuộc tính $U = \{A_i : i = 1, ..., n\}$, trong đó A_i là các thuộc tính kinh điển hoặc thuộc tính mờ. Miền giá trị của thuộc tính kinh điển của A_i là D_i và ta viết $Dom(A_i) = D_i$, miền giá trị của thuộc tính mờ là $Dom(A_j) = D_j \cup LDom(A_j)$.

Với miền giá trị như vậy, khái niệm quan hệ $r[R]$ trên lược đồ $R \subseteq U$ được định nghĩa hình thức như trong trường hợp quan hệ kinh điển. Chúng ta sẽ sử dụng những ký pháp quen thuộc như trong các tài liệu chuẩn về CSDL.

Trong mục này chúng ta sẽ đưa ra giải pháp xử lý ngữ nghĩa của các từ ngôn ngữ xuất hiện trong CSDL.

Trong trường hợp ngữ nghĩa của các từ được biểu thị bằng tập mờ, kiểu dữ liệu của thuộc tính mờ không thuần nhất bao gồm kiểu số thực và kiểu hàm.

Trong trường hợp ngữ nghĩa được biểu thị như là một phần tử trong ĐSGT, mỗi thuộc tính mờ được gắn với một ánh xạ ĐLNN $\nu_{fm,A}$. Khi đó mỗi từ x của thuộc tính mờ A xuất hiện trong CSDL được xem như là ký hiệu của giá trị thực $\nu_{fm,A}(x)$. Như vậy kiểu của thuộc tính mờ A trở nên thuần nhất. Tuy nhiên nếu ta xử lý giá trị này như giá trị thực thì đặc trưng mờ của dữ liệu trở nên vô nghĩa. Vì vậy ta cần có cách tiếp cận cho việc xử lý dữ liệu này.

Trước hết ta định nghĩa hệ lân cận mờ của phần tử $\nu_{fm,A}(x) \in [0, 1]$. Vì như đã nhận xét ở trên, các khoảng $K(y)$ gồm các phần tử gần ngữ nghĩa với phần tử $\nu_{fm,A}(y)$, và vì sự tương tự ngữ nghĩa có tính bắc cầu nên dựa trên các khoảng này ta có thể xác định hệ lân cận của $\nu_{fm,A}(x)$.

Định nghĩa 3.1. Cho ĐSGT $\mathcal{DS}(A)$ của thuộc tính mờ A và ánh xạ định lượng ngữ nghĩa $\nu_{fm,A}$. Xét phần tử $x \in X$, và giả sử x có độ dài $l, 5 > l > 0$. Giả sử miền thực của A có độ dài d. Khi đó lân cận ngữ nghĩa của x được xác định như sau:

(N1) Khoảng $dK(x)$, trong đó d là hệ số nhân, là một *lân cận bậc l* của từ x.

(N2) *Lân cận bậc $l + 1$* của x: Nếu $l + 1 < 5$, xét phân hoạch bậc $l + 1$ của $dK(x)$. Các phân hoạch này tương ứng với các phần tử $h_j x,\ j \in [-q^\wedge p]$, với độ dài là $d \times fm(h_j x)$. Lân cận bậc $l + 1$ của x là hợp (tập hợp) các khoảng của phân hoạch bậc $l + 1$ kế tiếp nhau sao cho tập kết quả là khoảng con thật sự của $dK(x)$ theo nghĩa nó không chứa hai khoảng con của phân hoạch nằm ở hai đầu của khoảng $dK(x)$.

(N3) *Bước lặp (Lân cận bậc $l' = l + i$)*: Nếu $l' < 5$ ta lấy hai khoảng K và K' của phân hoạch bậc $l' - 1$ có điểm nút chung là $\nu_{fm,A}(x)$ (hai khoảng này tồn tại theo K1)), và xét phân hoạch bậc l' của khoảng tạo bởi tập $K \cup K'$. Nhớ rằng mỗi phân hoạch K hay K' có dạng như trong K1) do đó điểm mút chung $\nu_{fm,A}(x)$ sẽ kề với khoảng phân hoạch tương ứng với phần tử có gia tử h_{-q} hoặc h_p ở tiếp đầu ngữ. Khi đó:

(i) Lân cận bậc l' của x là hợp các khoảng của phân hoạch đang xét kế tiếp nhau từ khoảng phân hoạch kề với điểm $\nu_{fm,A}(x)$ đến khoảng phân hoạch đầu tiên tương ứng với phần tử có chứa gia tử h_{-1} hay h_1 ở tiếp đầu ngữ.

(ii) Lân cận bậc l' chặt của x là hợp 2 khoảng của phân hoạch đang kề với điểm $\nu_{fm,A}(x)$, tức là hợp 2 khoảng phân hoạch tương ứng với hai phần tử có gia tử h_{-q} hoặc h_p ở tiếp đầu ngữ.

Lưu ý rằng lân cận bậc l *chặt* được xác định ở bước (N1) và (N2) là trùng nhau, chúng chỉ khác nhau ở bước lặp (N3).

Điều kiện $l + i < 5$ được áp đặt vì trong thực tiễn thường người ta chỉ sử dụng tối đa 4 gia tử tác động liên tiếp vào từ nguyên thủy (hay phần tử sinh).

Ví dụ 3.1. Cho ĐSGT của biến ngôn ngữ LỨA TUỔI là $\mathcal{AX} = (X, G, C, H, \leqslant)$, trong đó $G = \{young,\ old\}$, $H^- = \{R, L\}$ và $H^+ = \{M, V\}$, với R, L, M và V tương ứng là chữ viết tắt cho các gia tử *Rather, Little, More* và *Very*. Giả sử miền tham chiếu $D_A = [0, 120]$, $fm(old) = 0{,}55$, $fm(young) = 0{,}45$, $\mu(R) = 0{,}32$, $\mu(L) = 0{,}20$, $\mu(M) = 0{,}30$ và $\mu(V) = 0{,}18$. Như vậy ta có $\alpha = 0{,}52$ và $\beta = 0{,}48$.

Xây dựng các lân cận của *young*: Theo định nghĩa của ánh xạ định lượng ngữ nghĩa ta có $\nu_A(young)$ = (0,45 − 0,45×0,52)×120 =25,92, với lưu ý rằng ta cần nhân với 120 để giá trị của ánh xạ vào miền giá trị thực của lứa tuổi.

Trước hết ta tìm các lân cận (ngữ nghĩa) bậc 1 và 2 của *young*.

Khoảng $K(young)$ = [0, 0,45× 120] = [0, 54] là lân cận ngữ nghĩa bậc 1 của *young*. Phân hoạch bậc 2 của khoảng $K(young)$ bao gồm các khoảng con [0, 0,18× 54] = [0, 9,72]; (9,72, 9,72 + 0,30×54] = (9,72, 25,92]; (25,92, 25,92 + 0,32×4] = (25,92, 43,20] và (43,20, 54]. Do vậy hệ các lân cận ngữ nghĩa bậc 2 của *young* trong ví dụ này chỉ bao gồm một khoảng (9,72, 43,20].

Để xây dựng hệ lân cận bậc 3 của *young* ta lấy 2 khoảng phân hoạch bậc 2 kề với điểm 25,92, tức là 2 khoảng (9,72, 25,92] và (25,92, 43,20] và tính phân hoạch bậc 3 của chúng như sau:

Khoảng (9,72, 25,92] tương ứng với phần tử *Myoung* còn khoảng (25,92, 43,20] tương ứng với phần tử *Ryoung*. Khi đó phân hoạch bậc 3 tương ứng với các phần tử được sắp thứ tự như sau: $VMyoung < MMyoung < RMyoung < LMyoung < young < VRyoung < MRyoung < RRyoung < LRyoung$. Dựa vào các phần tử này ta tính các khoảng phân hoạch và thu được:

(9,72, 12,636], (12,636, 17,496], (17,496, 22,68], (22,68, 25,92]; (25,92, 29,0304], (29,0304, 34,2144], (34,2144, 39,744], (39,744, 43,20].

Trong ví dụ này $h_{-1} = R$ và khoảng bậc 3 tương ứng là K_{-1} = (17,496, 22,68] còn $h_1 = M$ và khoảng bậc 3 tương ứng là K_1= (29,0304, 34,2144]. Như vậy lân cận bậc 3 của *young* là hợp các khoảng kề với giá trị 25,92 đến các khoảng K_{-1} và K_1 và là khoảng (17.496, 34.2144]. Lân cận bậc 3 chặt của young là khoảng (22,68, 25,92] ((25,92, 29,0304] = (22,68, 29,0304].

Dễ dàng kiểm tra là $\nu_{fm,A}(Ryoung) = 34{,}2144$ và lân cận bậc 2 của *Ryoung* là *K(Ryoung)* = (25,92, 43,20], lân cận bậc 3 của *Ryoung* là (29,0304, 39,744].

Bây giờ chúng ta tính lân cận bậc 4 của *young*. Tương tự như trên, chúng ta tính phân hoạch bậc 4 của 2 khoảng (22,68, 25,92] tương ứng với phần tử *LMyoung* và (25,92, 29,0304] tương ứng với phần tử *VRyoung*. Trước hết ta xác định thứ tự của các phần tử tương ứng với các lớp (tức các khoảng) của phân hoạch như sau: $LLMyoung < RLMyoung < MLMyoung < VLMyoung < young < VVRyoung < MVRyoung < RVRyoung < LVRyoung$. Tuy nhiên,

theo (N3) chúng ta quan tâm tính các khoảng phân hoạch ở gần giá trị 25,92 tương ứng phần tử *young* và thu được các khoảng sau:

(24,3648, 25,3368], (25,3368, 25,92], (25,92, 26,479872], (26,479872, 27,412992].

Khi đó lân cận bậc 4 của *young* là (24,3648, 27,412992] và lân cận bậc 4 *chặt* của *young* là (25,3368, 26,479872].

Trên cơ sở khái niệm lân cận, chúng ta có thể đưa ra các định nghĩa về các quan hệ đối sánh giữa các phần tử trong miền giá trị *Dom(A)* của thuộc tính mờ A.

Định nghĩa 3.2. Đẳng thức bậc k (*chặt*): Giả sử A là thuộc tính mờ và với bất kỳ $a, b \in Dom(A) = D \cup LDom(A)$ ta nói chúng bằng nhau bậc k và ta viết $a =_k b (a =_{k,c} b)$ nếu:
(i) Với $a, b \in D$ thì $a = b$ hoặc tồn tại một lân cận $V_k(x)$ bậc k (*chặt*) sao cho $a, b \in V_k(x)$.
(ii) Với $a \in D$ và $b \in LDom(A)$ thì phải tồn tại một lân cận $V_k(b)$ bậc k (chặt) của b (tức là của $\nu_{fm,A}(b) \in [0, 1]$) sao cho $a \in V_k(b)$.
(iii) Với $a, b \in LDom(A)$ phải tồn tại một lân cận $V_k(b)$ bậc k (chặt) của b sao cho $\nu_{fm,A}(a) \in V_k(b)$ hoặc phải tồn tại một lân cận $V_k(a)$ bậc k (chặt) của a sao cho $\nu_{fm,A}(b) \in V_k(a)$.

Ví dụ 3.2. Giả sử trong CSDL mờ, cột thuộc tính A thuộc có các giá trị 28, *Ryoung*, 33, 35. Chúng ta xây dựng hệ lân cận như trong Ví dụ 3.1. Dễ kiểm tra thấy là $\nu_{fm,A}(Ryoung) = 34,2144$. Theo định nghĩa trên ta có $young =_2 28 =_2 Ryoung =_2 33 =_2 35$, tức chúng bằng nhau bậc 2; $young =_3 28 =_3 Ryoung =_3 33$ và chúng $\neq_3 35$, nhưng chỉ có $young =_{3,c} 28$ và $\{young, 28\} \neq_{3,c} \{Ryoung, 33, 35\}$; $young \neq_4 \{28, Ryoung, 33, 35\}$!.

Định nghĩa 3.3. Bất đẳng thức bậc k (chặt): Giả sử A là thuộc tính mờ và với bất kỳ $a, b \in Dom(A) = D \cup LDom(A)$ ta viết $a \geqslant_k b (a \geqslant_{k,c} b)$, nếu:
(i) Với $a, b \in D$ thì $a \geqslant b$.
(ii) Với $a \in D$ và $b \in LDom(A)$ thì phải tồn tại một lân cận $V_k(b)$ bậc k (chặt) của b (tức là của $\nu_{fm,A}(b) \in [0, 1]$) sao cho $a \geqslant V_k(b)$.
(iii) Với $a, b \in LDom(A)$ phải tồn tại một lân cận $V_k(b)$ bậc k (chặt) của b và một lân cận $V_k(a)$ bậc k (chặt) của a sao cho $V_k(a) \geqslant V_k(b)$.

Ví dụ 3.3. Ta tiếp tục sử dụng các giả thiết và các kết quả trong các Ví dụ 3.1 và 3.2. Như vậy lân cận bậc 4 của *young* là (24,3648, 27,412992] và lân cận bậc 3 của *Ryoung* là (29,0304, 39,744] và do đó lân cận bậc 4 của *Ryoung V(Ryoung)* phải bị chứa trong (29,0304, 39,744]. Vậy theo Định nghĩa 3.2 và 3.3 ta có $young =_3 Ryoung$ nhưng $Ryoung \geqslant_4 young$.

4. MỘT ỨNG DỤNG TRONG XÂY DỰNG MÔ HÌNH CSDL HÌNH SỰ

Trước hết cũng xin lưu ý rằng các đẳng thức và bất đẳng thức bậc k được định nghĩa như trên nhằm để khai thác thông tin trong CSDL với thông tin mờ mà không sử dụng trong thao tác về mặt syntax để "quản lý về mặt ký hiệu ".

Giả sử U là tập vũ trụ các thuộc tính và R là một lược đồ quan hệ trên U, tức là $R \subseteq U$. Mỗi thuộc tính A_i của U được gán một miền giá trị ký hiệu là $Dom(A_i)$. Nếu A_i chỉ nhận giá trị kinh điển, nó được gọi là thuộc tính kinh điển và miền giá trị của A_i là D_i, tức là $Dom(A_i) = D_i$. Nếu thuộc tính A_j có thể vừa nhận giá trị trong một miền kinh điển vừa nhận giá trị mờ được gọi là thuộc tính mờ và khi đó miền giá trị của nó là

$Dom(A_j) = D_j \cup LDom(A_j)$.

Mỗi một thuộc tính mờ A_j của U được gán với một hàm định lượng ngữ nghĩa $\nu_{fm,A_j} : LDom(D_j) \to D_j$, được xác định bởi biểu thức giải tính được cho trong Định nghĩa 2.3 với các tham số là độ đo tính mờ fm của từ sinh nguyên thủy và các gia tử. Với các tham số như vậy, ngữ nghĩa của giá trị thuộc tính có thể hiệu chỉnh cho phù hợp với thực tế khai thác, sử dụng CSDL.

Để thuận tiện trong trình bày, giả sử các phần tử của D_{Ai} được ký hiệu bằng các chữ cái thường đầu bảng chữ cái như $a, b, c, ...$ nếu cần với chỉ số và các phần tử của tập $LDom(A_i)$ sẽ được ký hiệu bằng các chữ cái thường cuối bảng chữ cái như $x, y, z, u, v, ...$ nếu cần với chỉ số.

Như thông thường, một bộ t trên R là một ánh xạ $t : R \to D(A_1) \cup \cdots \cup D(A_n)$ sao cho $t(A_i) \in D(A_i)$, với $1 \leqslant i \leqslant n$. Các bộ sẽ được ký hiệu là t, s nếu cần sử dụng thêm chỉ số. Ký hiệu $t[A_i]$ được hiểu là giá trị của bộ t tại thuộc tính A_i. Nếu X là tập con của $U, t[X]$ được hiểu là ánh xạ t được hạn chế trên tập X và được gọi là một bộ trên X.

Cho lược đồ quan hệ R trên U. Một thể hiện của R là một tập các bộ trên R và được gọi là một quan hệ trên R. Các quan hệ trên R được ký hiệu là $r[R], s[R], ...$ Nếu R được ngầm hiểu và không gây hiểu lầm, ký hiệu R có thể bỏ qua trong các ký pháp này.

Đối với một thuộc tính mờ A, một bộ t có thể nhận giá trị kinh điển hay giá trị ngôn ngữ tại A. Để có thể thao tác được với các giá trị ngôn ngữ như vậy, trong bài báo này chúng ta thao tác ngữ nghĩa ngôn ngữ dựa trên ĐSGT và ánh xạ định lượng ngữ nghĩa.

Như đã trình bày trong Mục 3, việc thao tác dữ liệu trong mô hình CSDL nghiên cứu trong bài này sẽ được thực hiện bằng các kỹ thuật kinh điển, trong đó các quan hệ đối sánh đối với dữ liệu kinh điển vẫn là $=, \neq, \leqslant, \geqslant, <$ và $>$, còn đối với dữ liệu ngôn ngữ chúng sẽ là các quan hệ đối sánh mờ $=_{k1}, \neq_{k2}, \leqslant_{k3}, \geqslant_{k4}, <_{k5}$ và $>_{k6}$.

Giả sử r là một quan hệ trên lược đồ R. Một câu truy vấn đối với quan hệ r có thể phát biểu dưới dạng $(A_{i1}\theta_{k1}a_{i1}, ..., A_{im}\theta_{km}a_{im})$, trong đó $A_{ij} \in R$, $a_{ij} \in D(A_{ij})$, còn θ_{kj} là một trong các quan hệ đối sánh kinh điển hay mờ nói trên tuỳ theo A_{ij}, $j = 1, ..., m$ là thuộc tính kinh điển hay thuộc tính mờ.

Thực tiễn trong các hoạt động hình sự, các thuộc tính thường hay nhận các giá trị mờ, không chính xác như lứa tuổi, chiều cao, màu tóc, màu da, giọng nói, dáng người ... Như vậy CSDL mờ sẽ rất phù hợp cho việc xây dựng CSDL hình sự.

CSDL hình sự sử dụng mô hình CSDL mờ trình bày ở trên có nhiều ưu điểm:

- Thứ nhất, như đã trình bày trong Mục 3, không giống như đối với các mô hình CSDL mờ khác, kiểu dữ liệu của mỗi thuộc tính kinh điển lẫn thuộc tính mờ đều thuần nhất.

- Thứ hai, các giá trị mờ được ánh xạ thành phần tử kinh điển làm "đại diện". Việc xử lý ngữ nghĩa thông qua các đẳng thức và bất đẳng thức bậc k có thể thực hiện bằng các kỹ thuật thao tác dữ liệu kinh điển.

- Để thuận tiện và nhanh chóng, có thể thiết lập các bảng lưu trữ các giá trị đại diện tính sẵn đối với từng thuộc tính cùng với hệ lân cận của chúng.

Với những ưu việt như vậy, việc xây dựng CSDL hình sự, bên cạnh các công việc xây dựng như đối với các CSDL thông thường, chỉ cần tiến hành thêm các công việc sau (với giả thiết hạn chế trong bài báo này là chúng chỉ gồm các thuộc tính mờ nhận các giá trị kinh điển hay giá trị ngôn ngữ của một ĐSGT với 1 hoặc 2 phần tử sinh nguyên thuỷ):

1) Xác định các thuộc tính mờ cùng các giá trị mờ của chúng. Xây dựng và cài đặt một thủ tục cho phép khai báo các thuộc tính và các tham số cho mỗi thuộc tính mờ như khai báo tên thuộc tính, các tham số xác định một ĐSGT mà người thiết kế gán cho thuộc tính và khai báo các tham số để tính hàm định lượng ngữ nghĩa dựa theo Định nghĩa 2.3.

Xây dựng một thủ tục sinh các giá trị mờ (giá trị ngôn ngữ). Các giá trị này sẽ được sử dụng để mô tả đối tượng và khi đó chúng sẽ xuất hiện trong cột tương ứng của quan hệ.

2) Vì các giá trị mờ của mỗi thuộc tính là hữu hạn, do mỗi từ ngôn ngữ chỉ có tối đa 4 gia tử, ta có thể xây dựng và cài đặt một thủ tục chung cho các thuộc tính cho phép tính sẵn các giá trị đại diện cũng như hệ lân cận của chúng để khai thác ngữ nghĩa dữ liệu. Lưu ý rằng, như đã nói ở trên, với các thuộc tính mờ mà giá trị của nó có thể là các giá trị ngôn ngữ thì các giá trị ngôn ngữ này chứ không phải là giá trị của ánh xạ định lượng ngữ nghĩa sẽ được lưu trong cột tương ứng của quan hệ, mặc dù khi xử lý, ta sẽ sử dụng giá trị của ánh xạ định lượng ngữ nghĩa đã tính sẵn. Lý do là việc lưu giá trị ngôn ngữ cho ta cái nhìn cảm tính rõ hơn và mang nhiều thông tin hơn.

3) Như đã chứng tỏ trong Mục 3, việc thao tác và xử lý dữ liệu ở mức ngữ nghĩa đối với các câu truy vấn được chuyển về việc thao tác xử lý các giá trị thực và các khoảng lân cận, còn việc thao tác và xử lý dữ liệu ở mức cú pháp (mức ký hiệu) hoàn toàn trùng với trường hợp kinh điển. Điều này cho phép chúng ta phát triển CSDL mờ theo cách tiếp cận này dựa trên CSDL kinh điển.

Chúng ta sẽ đưa ra một ví dụ để làm sáng tỏ điều trên. Để đơn giản việc trình bày chúng ta giả sử có hai thuộc tính mờ là LỨA TUỔI và DÁNG NGƯỜI. Giả sử các phần tử sinh nguyên thủy của các thuộc tính này tương ứng là *già* và *trẻ, cao* và *thấp* và, đối với thuộc tính DÁNG NGƯỜI để đơn giản ta giả sử chỉ có 2 phần tử sinh *béo* và *gầy.*

- Các tham số của LỨA TUỔI được cho như trong Ví dụ 3.1 với các lân cận của giá trị *young* là $K_1(young) = [0,\ 54]$, $K_2(young) = (9{,}72,\ 43{,}20]$, $K_3(young) = (17{,}496,\ 34{,}2144]$ và $K_{3,c}(young) = (22{,}68,\ 29{,}0304]$, $K_4(young) = (24{,}3648,\ 27{,}412992]$ và $K_{4,c}(young) = (25{,}3368,\ 26{,}479872]$, trong đó chỉ số chữ số chỉ bậc của lân cận và chỉ số c nói rằng nó là lân cận bậc k chặt.

Ngoài ra, ta có $\nu_{fm,A}(Ryoung) = 34,2144$, lân cận bậc 2 của $Ryoung$ là $K(Ryoung) = K_2(Ryoung) = (25,92,43,20]$ và lân cận bậc 3 là $K_3(Ryoung) = (29,0304,39,744]$, trong đó R là gia tử *Rather.*

- Bây giờ xác định các tham số của thuộc tính DÁNG NGƯỜI. Đây là một thuộc tính phức tạp, để đơn giản ta giới hạn các giá trị mô tả chỉ bao gồm các khái niệm sinh ra từ hai khái niệm nguyên thủy *béo* và *gầy.* Các khái niệm khác có thể làm tương đồng với một khái niệm sinh ra được từ hai khái niệm nguyên thủy này. Các gia tử sử dụng để sinh các khái niệm bao gồm *Rất (V), Hơn (M), Khá (R), Ít (L),* tương ứng với các gia tử tiếng Anh là $Very$, $More$, $Rather$ và $Little$.

Theo kinh nghiệm của các cán bộ y tế, mức độ béo, gầy có thể được xác định định lượng qua mối quan hệ giữa giá trị phần dư C của chiều cao của một người theo số đo centimet sau khi trừ đi 1 mét và trọng lượng W tính theo kg. Để phản ảnh định lượng các khái niệm mờ về DÁNG NGƯỜI, theo kinh nghiệm của các điều tra viên hình sự, chúng phải được định nghĩa không phụ thuộc vào chiều cao của con người. Do vậy, ta sử dụng tỷ lệ W/C được xác định trong khoảng [0,6, 1,4] với độ dài $d = 0,8$ (lưu ý là tỷ lệ C/W không phù hợp) để

mã hoá khái niệm dáng người. Chẳng hạn nếu giá trị này càng gần 1,4 thì người đó càng có dáng người béo. Chẳng hạn một người cao 160 cm và nặng 80 kg thì giá trị định lượng dáng người sẽ là 80/60 = 1,33, và chắc chắn người đó rất béo. Ngược lại, nếu giá trị này càng gần 0.6 thì người đó thiên về càng gầy.

Giả sử ta chọn các tham số của hàm định lượng ngữ nghĩa như sau

$fm(béo) = fm(gầy) = 0{,}5$, $\mu(V) = \mu(L) = 0{,}2$ và $\mu(M) = \mu(R) = 0{,}3$. Như vậy $\alpha = \beta = 0{,}5$.

Giá trị ngữ nghĩa định lượng của béo là $\nu_{fm,A}(béo) = 0{,}6 + (0{,}5 + 0{,}5/2) = 1{,}35$.

Xét giá trị *Khá béo (Rbéo)* của thuộc tính $\nu_{fm,A}(Rbéo) = 0{,}6 + (0{,}5 + 0{,}2\times\ 0{,}5 + \alpha\times 0{,}3 \times\ 0{,}5)\times$d = 1,14.

Các lân cận của giá trị *Rbéo* được xác định như sau:

Vì độ dài của *Rbéo* là 2 nên nó không có lân cận bậc 1. Lân cận bậc 2 của *Rbéo* là $(0{,}6 + (0{,}5 + 0{,}2\times\ 0{,}5)\times d,\ 0{,}6 + (0{,}5 + 0{,}2\times 0{,}5)\times d + 0{,}3\times 0{,}5\times d] = (1{,}08,\ 1{,}20]$.

Lân cận bậc 3 của *Rbéo* là $(1{,}08 + 0{,}2\times 0{,}3\times 0{,}5\times d,\ 1{,}08 + 0{,}2\times 0{,}3\times\ 0{,}5\times d + 0{,}3\times 0{,}3\times 0{,}5\times d + 0{,}3\times\ 0{,}3\times\ 0{,}5\times d] = (1{,}104,\ 1{,}176]$.

Đối với giá trị *MRbéo* (hiểu là béo hơn là *Khá béo*) ta có giá trị ngữ nghĩa định lương là $\nu_{fm,A}(MRbéo) = \nu_{fm,A}(Rbéo) + 0{,}3\times 0{,}3\times 0{,}5/2 = 1{,}14 + 0{,}0225 = 1{,}1625$.

Trong thực tế cài đặt, hệ thống sẽ tính tất cả các giá trị ngữ nghĩa định lượng của các khái niệm mờ và hệ lân cận của chúng. Ở đây chúng ta chỉ tính một số giá trị để sử dụng trong ví dụ để làm sáng tỏ phương pháp tiếp cận đề xuất trong bài báo này.

Giả sử trong CSDL mờ chúng ta có những thông tin sau:

SỐ TT	TÊN	TUỔI	DÁNG NGƯỜI	SỐ LẦN PHẠM TỘI
1	Tửu	18	béo	4
2	Trúc	trẻ	VRbéo	6
3	Ngưu	24	gầy	3
4	Mục	31	MRbéo	8

Giả sử hệ CSDL nhận được câu truy vấn sau:

(TUỔI $=_2$ *Ryoung*, DÁNG NGƯỜI $=_2$ *Rbéo*).

Theo tính toán trên, lân cận bậc 2 của*Ryoung* là $(25.92, 43.20]$ và lân cận bậc 2 của *Rbéo* là $(1.08, 1.20]$. Như vậy đối tượng ở hai hàng 2 và 4 của bảng trên thỏa câu truy vấn trên.

Nếu ta cần độ chính xác cao hơn, ta đưa ra câu truy vấn có dạng (TUỔI $=_3$ *Ryoung*, DÁNGNGƯỜI $=_3$ *Rbéo*). Khi đó lân cận bậc 3 của *Ryoung* là $(29,0304, 39,744]$ và lân cận bậc 3 của *Rbéo* là $(1,104, 1,176]$. Như vậy chỉ có đối tượng ở hàng 4 của bảng thỏa câu truy vấn vừa cho.

5. KẾT LUẬN

Trong bài báo này, chúng tôi đã đề xuất một phương pháp mới xây dựng cơ sở dữ liệu mờ dựa trên các kết quả mới nhất trong nghiên cứu về đại số gia tử. Phương pháp này, theo ý kiến của chúng tôi, cho phép xử lý các thông tin mờ trong cơ sở dữ liệu một cách nhất quán và hiệu quả, đồng thời không đòi hỏi những tính toán quá phức tạp. Do đó, hy vọng là có

thể xây dựng được những cơ sở dữ liệu thực tế đáp ứng những yêu cầu quản lý trong ngành Công an nói riêng và các bài toán quản lý có liên quan đến các thông tin mờ nói chung.

TÀI LIỆU THAM KHẢO

[1] B.P. Buckles, F.E. Petry, A fuzzy representation of data for relational databases, *Fuzzy Sets and Systems* **7** (3) (1982) 213–226.

[2] G. Chen, E.E. Kerre, J. Vandenbulcke, A computational algorithm for the FFD transitive closure and a complete axiomatization of fuzzy functional dependence, *International Journal of Intelligent Systems* **9** (5) (1994) 421–439.

[3] Bùi Thị Thúy Hiền, N.C. Hồ, Functional dependencies with context dependent values in relational databases, *MIF'99: The International Symposium on Medical Informatics and Fuzzy Technology*, 271–279.

[4] N.C. Ho, Fuzziness in structure of linguistic truth values: a foundation for development of fuzzy reasoning, *Proc. of Int. Symp. on Multiple-Valued Logic*, May 26-28, 1987, Boston University, Boston, Massachusetts, IEEE Computer Society Press, 1987, 325–335.

[5] 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).

[6] N.C. Ho, A topological completion of refined hedge algebras and a model of fuzziness of linguistic terms, *submitted to Fuzzy Sets and Systems.*

[7] N.C. Ho, N.V. Long, Fuzziness measure on complete hedge algebras and quantitative semantics of terms in linear hedge algebras, *submitted to Fuzzy Sets and Systems.*

[8] N.C. Ho, N. H. Chau, Quantitative semantics in hedge algebras and interpolation methods, *Proc. of ICT,* Hanoi, 2003.

[9] N.C. Ho, H.V. Nam, Ordered structure-based semantics of linguistic terms of linguistic variables and approximate reasoning, *AIP Conf. Proceed. on Computing Anticipatory Systems, CASYS'99, 3th Inter. Conf.,* 98 – 116.

[10] N.C. Ho, H.V. Nam, Lattice character of the refinement structure of hedge algebras, *Journal of Informatics and Cybernetics* **12** (1) (1996) 07–20.

[11] N.C. Ho, H.V. Nam, A theory of refinement structure of hedge algebras and its application to linguistic-valued fuzzy logic. In D. Niwinski & M. Zawadowski (Eds), *Logic, Algebra and Computer Science, Banach Center Publications* Vol. **46** (PWN - Polish Scientific Publishers 1999).

[12] N.C. Ho, H.V. Nam, Towards an algebraic foundation for a zadeh fuzzy logic, *Fuzzy Set and System* **129** (2002) 229–254.

[13] 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** (4) (1999) 347–361.

[14] N.C. Ho and T.T. Son, On distance between values of linguistic variable based on the structure of hedge algebras, *Journal of Informatics and Cybernetics* **11** (1) (1995) (Vietnamese).

[15] N.C. Ho, T.T. Son, T.D. Khang, and L.X. Viet, Fuzzyness measure, quantified semantic mapping and interpolative method of approximate reasoning in medical expert systems, *Journal of Informatics and Cybernetics* **18** (3) (2002) 237–252.

[16] N.C. Ho and W. Wechler, Hedge algebras: An algebraic approach to structures of sets of linguistic domains of linguistic truth variable, *Fuzzy Sets and Systems* **35** (3) (1990) 281–293.

[17] N. Cat Ho and W. Wechler, Extended hedge algebras and their application to fuzzy logic, *Fuzzy Sets and Systems* **52** (1992) 259–281.

[18] L. Di Lascio, A. Gisolfi, and V. Loia, A new model for linguistic modifiers, *International Journal of Approximate Reasoning* **15** (1996) 25–47.

[19] L. Di Lascio and A. Gisolfi, Averaging linguistic truth values in fuzzy approximate reasoning, *International Journal of Intelligent Systems* **13** (1998) 301–318.

[20] L.T. Koczy and K. Hirota, Interpolative reasoning with insufficient evidence in sparse fuzzy rules bases, *Inform. Sci.* **71** (1993) 169–201.

[21] L. Di Lascio, A. Gisolfi, and V. Loia, A new model for linguistic modifiers, *International Journal of Approximate Reasoning* **15** (1996) 25–47.

[22] L. Di Lascio and A. Gisolfi, Averaging linguistic truth values in fuzzy approximate reasoning, *International Journal of Intelligent Systems* **13** (1998) 301–318.

[23] Mustafa LLKer Sozat, Adnan Yazici, A complete axiomatization for fuzzy functional and multivalued dependencies in fuzzy database relations, *Fuzzy Set and Systems* **117** (2001) 161–181.

[24] E. Petry and P. Bosc, *Fuzzy Databases Principles and Applications*, Kluwer Academic Publishers, 1996.

[25] Hồ Thuần, Hồ Cẩm Hà, An approach to extending the relational database model for handing incomplete information and data dependencies, *Tạp chí Tin học và Điều khiển học* **1** (3) (2001) 41–47.

[26] Le Tien Vuong, Ho Thuan, A relational database extended by application of fuzzy set theory and linguistic variables, *Computer and Artificial Intelligence* **8** (2) (1989) 153–168.

Nhận bài ngày 10 - 2 - 2006
Nhận lại sau sửa ngày 24 - 5 - 2006