Một số dạng chuẩn trong cơ sở dữ liệu mờ theo cách tiếp cận đại số gia tử

Fuzzy Normal Forms an Approach Hedge Algbebras

Nguyễn Công Hào

Abstract: *In this paper, we introduced some results about key in fuzzy databases model an approach hedge algebras [5]. Fuzzy normal forms such as fuzzy first normal form, fuzzy second normal form, fuzzy third normal form is discussed. Finally, dependency-preserving decomposition into k-F3NF and lossless-join decomposition into k-FBCNF is proposed.*

I. GIỚI THIỆU

Trong những năm gần đây, CSDL mờ và các vấn đề liên quan đã được nhiều tác giả trong và ngoài nước quan tâm nghiên cứu và đã có những kết quả đáng kể [7][8][9][10][14][16]. Có nhiều cách tiếp cận khác nhau như cách tiếp cận theo lý thuyết tập mờ [7][8][14], theo lý thuyết khả năng, tương tự [9][10][16] ...Tất cả các cách tiếp cận trên nhằm mục đích nắm bắt và xử lý một cách thỏa đáng trên một luận điểm nào đó các thông tin không chính xác (*unexact*), không chắc chắn (*uncertainty*) hay những thông tin không đầy đủ (*incomplete*). Do sự đa dạng của những loại thông tin này nên chúng ta gặp rất khó khăn trong biểu thị ngữ nghĩa và thao tác với chúng. Chẳng hạn như theo cách tiếp cận lý thuyết tập mờ muốn tìm một giá trị ngôn ngữ xấp xỉ một tập mờ thì có rất nhiều yếu tố ảnh hưởng đến việc xấp xỉ ngôn ngữ, gây nên sự phức tạp. Mặc khác do có nhiều phương pháp khử mờ mà mỗi phương pháp cho kết quả khác nhau nên dẫn đến khả năng sai số lớn.

Vì vậy, các nhà nghiên cứu tìm kiếm các phương pháp tiếp cận đến việc thao tác dữ liệu dựa trên logic mờ theo nghĩa của Zadeh mà không sử dụng phương pháp truyền thống. N. Cat Ho và W.Wechler [3] đã khởi xướng phương pháp tiếp cận đại số đến cấu trúc tự nhiên của miền giá trị của biến ngôn ngữ. Theo cách tiếp cận này, mỗi giá trị ngôn ngữ của biến ngôn ngữ nằm trong một cấu trúc đại số gọi là đại số gia tử. Dựa trên những tính chất ngữ nghĩa của ngôn ngữ được phát hiện, bằng phương pháp tiên đề hóa nhiều tác giả đã tập trung phát triển lý thuyết đại số gia tử và đã có nhiều đóng góp trong lập luận ngôn ngữ [1][2][4] mà các tác giả chưa quan tâm đến vấn đề xây dựng mô hình cơ sở dữ liệu theo cách tiếp cận này. Để góp phần nghiên cứu cơ sở dữ liệu mờ, theo hướng tiếp cận này, chúng tôi xem miền trị của thuộc tính là một đại số gia tử (*thuộc tính này phải được xem như là biến ngôn ngữ*). Để so sánh hai giá trị trong miền trị của thuộc tính chúng tôi đã xây dựng quan hệ "*xấp xỉ theo mức*" [5] dựa trên tính mờ của các giá trị ngôn ngữ, theo cách tiếp cận này có thể thấy một số ưu điểm như sau:

– Đảm bảo tính thuần nhất về kiểu dữ liệu.

– Việc tổ chức lưu trữ và thao tác dữ liệu trở nên đơn giản, trực quan hơn.

– Tập trung nổ lực vào việc lựa chọn độ đo tính mờ của các gia tử và chúng trở thành hệ tham số của cách tiếp cận. Vì vậy, cách quản lý ngữ nghĩa dữ liệu rõ ràng.

– Không cần khử mờ (trong hầu hết cách tiếp cận với mô hình trước đây đều khử mờ, mỗi phương pháp khử mờ sẽ cho kết quả khác nhau).

Dựa trên mô hình đã được [5] đề xuất, bài báo tập trung giới thiệu một số kết quả như phụ thuộc hàm mờ, khoá và một số vấn đề liên quan đến các dạng chuẩn mờ.

II. MỘT SỐ KHÁI NIỆM CƠ SỞ

Trong phần này, chúng tôi nhắc lại một số định nghĩa, bổ đề, tính chất và định lý đã được đề xuất ở [5] làm cơ sở đi xây dựng định nghĩa một khái niệm mới và nghiên cứu trong các phần tiếp theo.

Định nghĩa 2.1 [5]. Gọi *fm* là độ đo tính mờ trên đại số gia tử X. Với mỗi $x \in X$, ta ký hiệu $I(x) \subseteq [0,1]$ và $|I(x)|$ là độ dài của I(x).

Một họ các $J = \{I(x): x \in X\}$ 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 $\{I(h_i x)$: i=1..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, tức là: hoặc $I(h_1x) < I(h_2x) < I(h_3x) <\ldots\ldots< I(h_{p+q}x)$

hoặc $I(h_1x) > I(h_2x) > I(h_3x) >\ldots\ldots> I(h_{p+q}x)$.

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 2.2 [5]. Xét $P^k = \{I(x): x \in X^k\}$ với $X^k = \{x \in X: |x| = k\}$ là một phân hoạch, ta nói rằng u *xấp xỉ* 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 là $\forall u,v \in X$, $u \approx_k v \Leftrightarrow \exists \Delta^k \in P^k : I(u) \subseteq \Delta^k$ và $I(v) \subseteq \Delta^k$.

Bổ đề 2.1 [5]. Quan hệ $\approx_k$ là một quan hệ tương đương trên P^k.

Bổ đề 2.2 [5]. Cho $u = h_n \ldots h_1 x$ và $v = h'_m \ldots h'_1 x$ là biểu diễn chính tắc của u và v đối với x.

(1) : Nếu u = v thì $u \approx_k v$ với mọi k.

(2) : Nếu $h_1 \neq h'_1$ thì $u \approx_{|x|} v$

Định lý 2.1 [5]. Xét $P^k = \{I(x): x \in X^k\}$ với $X^k = \{x \in X: |x| = k\}$ là một phân hoạch, $u = h_n \ldots h_1 x$ và $v = h'_m \ldots h'_1 x$ là biểu diễn chính tắc của u và v đối với x.

(1) : Nếu $u \approx_k v$ thì $u \approx_{k'} v$, $\forall\ 0 < k' \leq k$.

(2) : Nếu tồn tại một chỉ số $j \leq \min(m,n)$ lớn nhất sao cho với mọi s =1..j ta có $h_s = h'_s$ thì $u \approx_{j+|x|} v$.

Định nghĩa 2.3. Cho U là lược đồ quan hệ, $X \subseteq U$ là tập các thuộc tính, quan hệ r xác định trên U. Giá trị của hai bộ dữ liệu t_1 và t_2 trên tập thuộc tính X xấp xỉ mức k trong P^k ký hiệu $t_1[X] \approx_k t_2[X]$ khi và chỉ khi giá trị của hai bộ dữ liệu t_1 và t_2 trên mọi thuộc tính của X xấp xỉ mức k trong P^k. Có nghĩa là $t_1[X] \approx_k t_2[X] \Leftrightarrow \forall t_1, t_2 \in r$, $\forall A_i \in X$ ta có $t_1[A_i] \approx_k t_2[A_i]$.

Mệnh đề 2.1. Cho U là một lược đồ quan hệ, r là một quan hệ xác định trên U. $X, Y \subseteq U$ là tập các thuộc tính. Nếu $t_1[X] \approx_k t_2[X]$ thì $t_1[Y] \approx_k t_2[Y]$, $\forall t_1, t_2 \in r$ và $Y \subseteq X$.

Chứng minh: Nếu X=Y, theo định nghĩa hiển nhiên ta có nếu $t_1[X] \approx_k t_2[X]$ thì $t_1[X] \approx_k t_2[X]$, $\forall t_1, t_2 \in r$. Nếu $Y \subset X$, ta đặt $X = \{A_1, A_2, \ldots A_p\}$, $Y = \{A_1, A_2, \ldots A_q\}$, với p>q. Ta có $\forall t_1, t_2 \in r$, nếu $t_1[X] \approx_k t_2[X]$ thì $t_1[A_i] \approx_k t_2[A_i]$, $\forall A_i \in X$ hay $t_1[A_i] \approx_k t_2[A_i]$, $\forall A_i \in Y$. Vậy $t_1[Y] \approx_k t_2[Y]$.

Mệnh đề 2.2. Nếu $t_1[X] \approx_k t_2[X]$ và $t_2[X] \approx_k t_3[X]$ thì $t_1[X] \approx_k t_3[X]$, $\forall t_1, t_2, t_3 \in r$.

Chứng minh : Theo giả thiết $t_1[X] \approx_k t_2[X]$ nên theo định nghĩa ta có $\forall t_1, t_2 \in r$, $\forall A_i \in X$ ta có $t_1[A_i] \approx_k t_2[A_i]$ (*).Theo giả thiết $t_2[X] \approx_k t_3[X]$ nên theo định nghĩa ta có $\forall t_2, t_3 \in r$, $\forall A_i \in X$ ta có $t_2[A_i] \approx_k t_3[A_i]$ (**). Từ (*) và (**) ta suy ra $\forall t_1, t_3 \in r$ ta có $t_1[A_i] \approx_k t_3[A_i]$ (bổ đề 2.1), hay $\forall t_1, t_3 \in r$ ta có $t_1[X] \approx_k t_3[X]$.

Mệnh đề 2.3. Ta có $t_1[X] \approx_k t_1[X]$, $\forall t_1, t_2 \in r$, $\forall k$.

Chứng minh: Ta chứng minh bằng phương pháp quy nạp.

Nếu $|t_1[X]|=1$ thì $t_1[X]=c^+$ hoặc $t_1[X]=c^-$.

Ta có $\exists \Delta^1 = I(c^+) \in P^1 : I(c^+) = I(t_1[X]) \subseteq \Delta^1$ hoặc $\exists \Delta^1 = I(c^-) \in P^1 : I(c^-) = I(t_1[X]) \subseteq \Delta^1$. Vậy $t_1[X] \approx_1 t_1[X]$.

Giả sử $|t_1[X]| = n$ đúng, có nghĩa $t_1[X]\approx_n t_1[X]$ đúng, ta cần chứng minh $t_1[X]\approx_{n+1}t_1[X]$ đúng.

Đặt $t_1[X]=h_1x'$, với $|x'|=n$. Vì $t_1[X]\approx_n t_1[X]$ nên theo định nghĩa ta có $\exists\Delta^n\in P^n$: $I(t_1[X])\subseteq\Delta^n$.

Mặc khác ta có $P^{n+1} =\{I(h_1x'), I(h_2x'),\ldots\ldots\}$, với $h_1\neq h_2\neq\ldots$là một phân hoạch của $I(x')$. Do đó $\exists\Delta^{n+1}=I(h_1x')\in P^{n+1}$: $I(h_1x')=I(t_1[X])\subseteq\Delta^{n+1}$. Vậy $t_1[X]\approx_{n+1}t_1[X]$ đúng.

Định nghĩa 2.4. Cho U là lược đồ quan hệ, r là quan hệ xác định trên U. $X,Y\subseteq U$ là tập các thuộc tính, k là mức phân hoạch xác định trên U. Ta nói rằng X xác định phụ thuộc hàm mờ Y hay Y phụ thuộc hàm mờ vào X theo mức k trong P^k ký hiệu $X{\sim}{>}_kY$ đúng trong quan hệ r khi và chỉ khi $\forall t_1, t_2\in r$ nếu $t_1[X]\approx_k t_2[X]$ thì $t_1[Y]\approx_k t_2[Y]$. Có nghĩa là $X{\sim}{>}_kY \Leftrightarrow \forall t_1, t_2\in r$ nếu $t_1[X]\approx_k t_2[X]$ thì $t_1[Y]\approx_k t_2[Y]$.

Từ định nghĩa ta có hệ tiên đề cho trường hợp phụ thuộc hàm mờ theo cách tiếp cận đại số gia tử

$AFFD_1$: (phản xạ) Nếu $Y\subseteq X$ thì $X{\sim}{>}_kY$

$AFFD_2$: (tăng trưởng) Nếu $X{\sim}{>}_kY$ thì $XZ{\sim}{>}_kYZ$, với $Z\subseteq U$

$AFFD_3$: (bắt cầu) Nếu $X{\sim}{>}_kY$ và $Y{\sim}{>}_kZ$ thì $X{\sim}{>}_k Z$

$AFFD_4$: (bao hàm mức phân hoạch) Nếu $X{\sim}{>}_kY$ thì $X{\sim}{>}_{k'}Y,\forall\ 0 <k'\leq k$

III. MỘT SỐ DẠNG CHUẨN MỜ

Trước hết chúng ta xem xét một số khái niệm như *k-khoá*, *k-siêu khoá*, phụ thuộc hàm mờ, phụ thuộc hàm mờ đầy đủ, phụ thuộc hàm mờ bộ phận, phụ thuộc hàm mờ trực tiếp để làm cơ sở xây dựng các dạng chuẩn mờ.

Cho $X,Y\subseteq U$ là hai tập thuộc tính, k là mức phân hoạch trên U. Y được gọi là phụ thuộc hàm mờ đầy đủ (*fully*) vào X theo mức k, ký hiệu $X{\sim}{>}_{kF}Y$ nếu $X{\sim}{>}_kY$ và không tồn tại $\emptyset\neq X_1\subset X : X_1{\sim}{>}_kY$.

Nếu tồn tại $\emptyset \neq X_1\subset X : X_1{\sim}{>}_kY$ thì Y được gọi là phụ thuộc bộ phận (*partial*) vào X, ký hiệu $X{\sim}{>}_{kP}Y$.

Y được gọi là phụ thuộc hàm mờ bắc cầu (*transitivity*) vào X theo mức k, ký hiệu $X{\sim}{>}_{kT}Y$ nếu $\exists Z\subseteq U$: $Y-Z \neq\emptyset$, $X{\sim}{>}_kZ$, $Z\ !{\sim}{>}_kX$, $Z {\sim}{>}_kY$.

Nếu $X{\sim}{>}_kY$ và Y không phụ thuộc hàm mờ bắc cầu vào X thì ta nói Y phụ thuộc hàm trực tiếp (*direct*) vào X.

1. k-*khoá* và k-*siêu khoá* trong cơ sở dữ liệu mờ

Trong mô hình cơ sở dữ liệu mờ, *k-khoá* và *k-siêu khoá* cũng được xây dựng như là một sự mở rộng của khoá và siêu khoá trong mô hình quan hệ.

Gọi $AF{\sim}^+$ là tập các phụ thuộc hàm mờ được suy dẫn từ tập phụ thuộc hàm mờ $AF{\sim}$ dựa vào hệ tiền đề $AFFD_1$-$AFFD_4$, có nghĩa $AF{\sim}^+= \{X{\sim}{>}_kY : AF{\sim} \models X{\sim}{>}_kY\}$

Cho $K\subseteq U$ và $AF{\sim}$ là tập phụ thuộc hàm mờ trên lược đồ U, K được gọi là *k-khoá* của U nếu và chỉ nếu:

$$\begin{cases} K {\sim}{>}_k U \in AF\ {\sim}^+ \\ \quad K {\sim}{>}_{kF}\ U \end{cases}$$

Nếu $K{\sim}{>}_kU\in AF{\sim}^+$ thì K được gọi là *k-siêu khoá*.

2. Các dạng chuẩn mờ

Mục đích của các dạng chuẩn trong cơ sở dữ liệu là giải quyết các vấn đề dư thừa dữ liệu và dị thường trong khi thao tác dữ liệu.

Cho U là lược đồ quan hệ, $AF{\sim}$: tập các phụ thuộc hàm mờ, K là một *k-khoá*.

Dạng chuẩn một mờ (*k*-F1NF): Các lược đồ quan hệ U xây dựng theo cách tiếp cận đại số gia tử đều ở dạng chuẩn một mờ.

Dạng chuẩn hai mờ (*k*-F2NF): Lược đồ quan hệ U được gọi là dạng chuẩn hai mờ, ký hiệu *k*-F2NF, nếu U là ở dạng *k*-F1NF và $K {\sim}{>}_{kF} A\in AF{\sim}^+$, với $A\in R$ là thuộc tính không *k-khoá*.

Dạng chuẩn ba mờ (*k*-F3NF): Lược đồ quan hệ U được gọi là dạng chuẩn ba mờ, ký hiệu *k*-F3NF, nếu U là ở dạng *k*-F1NF và với mọi thuộc tính không *k-khoá* đều phụ thuộc hàm trực tiếp vào *k-khoá*.

Dạng chuẩn Boyce-Codd mờ (*k*-FBCNF): Lược đồ quan hệ U được gọi là dạng chuẩn Boyce-Codd mờ, ký hiệu *k*-FBCNF, nếu U là ở dạng *k*-F1NF và với mọi thuộc tính đều phụ thuộc hàm trực tiếp vào *k-khoá*. (Hay mọi $X \sim>_k A$, $A \notin X$, X là một *k-siêu khoá* của U).

Ví dụ 3.1. Cho lược đồ U = (A, B, C, D) và AF~ = $\{A \sim>_3 B, B \sim>_3 A, AC \sim>_2 D\}$

Từ $A \sim>_3 B \Rightarrow A \sim>_2 B$, $B \sim>_3 A \Rightarrow B \sim>_2 A$.

Vậy ta có $AC \sim>_2 U$ và $AC \sim>_{2F} U$ nên AC là *2-khoá* của U.

Tương tự BC cũng là *2-khoá* của U. Lược đồ U ở dạng chuẩn *2*-F3NF.

IV. TÁCH BẢO TOÀN PHỤ THUỘC HÀM VỀ DẠNG CHUẨN *K*-F3NF

Trước tiên, chúng tôi giới thiệu một số khái niệm như phủ và phủ tối tiểu của tập phụ thuộc hàm mờ. Tiếp theo đề xuất một thuật toán và định lý để xác định một phép tách lược đồ về dạng chuẩn *k*-F3NF bảo toàn phụ thuộc hàm.

Định nghĩa 4.1 (Sự tương đương của hai tập phụ thuộc hàm mờ): Cho lược đồ quan hệ U. Hai tập phụ thuộc hàm mờ AF~ và AG~ được gọi là tương đương (hoặc AF~ là phủ của AG~ và ngược lại) trên U khi và chỉ khi $AF\sim^+ = AG\sim^+$

Định nghĩa 4.2 (Phủ tối tiểu của tập phụ thuộc hàm mờ)

Tập phụ thuộc hàm mờ AF~ được gọi là tối tiểu nếu AF~ thoả mãn các yêu cầu sau:

(1): $\forall X \sim>_k A \in AF\sim$, $A \in U$ là một thuộc tính.

(2): $\forall X \sim>_k A \in AF\sim$ ta có $\{AF\sim -(X \sim>_k A)\}^+ \neq AF\sim^+$.

(3): $\forall X \sim>_k A \in AF\sim$, $\forall Z \subset X$ và $Z \neq X$, ta có $\{[AF\sim - (X \sim>_k A)] \cup (Z \sim>_k A)\}^+ \neq AF\sim^+$.

Thuật toán 4.1. Tách bảo toàn phụ thuộc hàm về dạng chuẩn *k*-F3NF.

Vào: Lược đồ quan hệ U, tập các phụ thuộc hàm mờ AF~. Không mất tính tổng quát giả sử AF~ là phủ tối tiểu.

Ra: ρ: Là một phép tách bảo toàn phụ thuộc hàm bao gồm tập các lược đồ con trong đó mỗi lược đồ con đều ở dạng chuẩn *k*-F3NF với các phụ thuộc hàm là hình chiếu của AF~ lên lược đồ đó.

Phương pháp:

(1) *If* $\exists B \in U$: $\forall X \sim>_k A \in AF\sim$, $B \notin X$ và $B \neq A$ *then*

(2) $U_1 = \bigcup B$ và $U = U - U_1$

(3) *If* $\exists X \sim>_k A \in AF\sim$: $\bigcup XA = U$ *then*

(4) $\rho = \{U\}$

(5) *Else if* Với mỗi $X \sim>_k A \in AF\sim$, ta có $\rho = \{U_i\} = \{\bigcup_i = XA\}$

Định lý 4.1. Cho lược đồ quan hệ U, AF~ là một phủ tối tiểu của tập phụ thuộc hàm mờ. Thuật toán 4.1 xác định một phép tách $\rho = \{U_1, U_2, \ldots U_m\}$ bảo toàn tập phụ thuộc hàm chuẩn hoá lược đồ quan hệ U về dạng chuẩn *k*-F3NF.

Chứng minh

Ta cần chứng minh ρ bảo toàn tập phụ thuộc hàm và mỗi U_i là dạng chuẩn *k*-F3NF

ρ bảo toàn tập phụ thuộc hàm

Vì $\forall X \sim>_k A \in AF\sim$ ta có $X \sim>_k A \in \Pi \bigcup_i (AF\sim)|_k \subseteq \Pi \bigcup_i (AF\sim) = \{V \sim>_k W | V \sim>_k W \in AF\sim^+, VW \subseteq \bigcup_i\}$, nên ρ bảo toàn tập phụ thuộc hàm.

Mỗi U_i là dạng chuẩn k-F3NF

Ta chứng minh bằng phản chứng. Giả sử rằng U_i không phải ở dạng chuẩn *k*-F3NF, khi đó $\exists Y \sim>_k B \in \Pi \bigcup_i (AF\sim)|_k$: B là thuộc tính không *k-khoá* không phụ thuộc trực tiếp vào *k-khoá*. Hay, thuộc tính không *k-khoá* $B \notin Y$, Y không *k-siêu khoá* của XA. Mặc khác ta lại có $YB \subseteq XA$ (Chú ý rằng A là thuộc tính đơn).

Nếu A=B thì $YA \subseteq XA$, vì $A \notin Y$ nên ta có $Y \subseteq X$. Từ Y không *k-siêu khoá* của XA và X là *k-siêu khoá* của XA, ta suy ra $Y \subset X$. Do đó $Y \sim>_k A$..

Từ $Y \sim>_k A \in \Pi\bigcup_i(AF\sim)|_k \subseteq AF\sim^+$ ta có thể thay $X \sim>_k A \in AF\sim$ bởi $Y \sim>_k A$..Vì $\{[AF\sim - (X \sim>_k A)] \cup (Y \sim>_k A)\}^+ = AF\sim^+$ (do $Y \sim>_k A \Rightarrow X \sim>_k A$) và $AF\sim^+ = \{[AF\sim - (X \sim>_k A)] \cup (Y \sim>_k A)\}^+$ (do $Y \sim>_k A \in AF\sim^+$). Vậy $AF\sim$ và $\{[AF\sim - (X \sim>_k A)] \cup (Y \sim>_k A)\}$ là tương đương nhau, điều này mâu thuẩn với giả thiết vì $AF\sim$ là phủ tối tiểu.

Tương tự nếu $A \neq B$ thì từ X là *k-siêu khoá* của XA sẽ $\exists Z \subseteq X$: Z là một *k-khoá* của XA. Do $B \subseteq XA$ và $A \neq B$ nên $\exists B_1 \in B$ sao cho $B_1 \in X$. Vì B là thuộc tính không *k-khoá* nên B_1 là thuộc tính không *k-khoá* hay $B_1 \notin Z$, vậy $Z \subset X$. Do đó, từ $Z \subset X$ và $Z \sim>_k A$ có thể thay thế $X \sim>_k A \in AF\sim$, điều này mâu thuẫn với giả thiết. Vậy, mỗi $U_i \in \rho$ là dạng chuẩn *k*-F3NF.

Ví dụ 4.1. Cho lược đồ quan hệ U={A, B, C, D} và tập các phụ thuộc hàm mờ $AF\sim = \{CD \sim>_3 A, CD \sim>_3 B, D \sim>_2 B\}$.

Ta có CD là một *2-khoá* của U. Từ $CD \sim>_3 B \Rightarrow CD \sim>_2 B$, do đó $CD \sim>_{2P} B$. Vậy U không ở dạng chuẩn *2*-F2NF. Áp dụng thuật toán 4.1 ta có ρ=(CDA, CDB, DB). Ta có ρ bảo toàn phụ thuộc hàm và các lược đồ con CDA ở dạng chuẩn *3*-F3NF với $FCDA\sim = \Pi_{CDA}(AF\sim)|_3$, CDB ở dạng chuẩn *3*-F3NF với $FCDB\sim = \Pi_{CDB}(AF\sim)|_3$ và DB ở dạng chuẩn *2*-F3NF với $FDB\sim = \Pi_{DB}(AF\sim)|_2$

Ví dụ 4.2. Cho lược đồ quan hệ U={A, B, C, D, E} và tập các phụ thuộc hàm mờ $AF\sim = \{CD \sim>_3 A, CD \sim>_3 B, AD \sim>_3 E, CD \sim>_3 E, A \sim>_2 B, B \sim>_1 E\}$.

Ta có CD là một *3-khoá* của U và U ở dạng chuẩn *3*-F2NF nhưng không ở dạng chuẩn *3*-F3NF. Do đó áp dụng thuật toán 4.1 ta có phép tách ρ = (CDA, CDB, ADE, CDE, AB, BE). Ta có ρ bảo toàn phụ thuộc hàm và các lược đồ con CDA ở dạng chuẩn *3*-F3NF với $FCDA\sim = \Pi_{CDA}(AF\sim)|_3$, CDB ở dạng chuẩn *3*-F3NF với $FCDB\sim = \Pi_{CDB}(AF\sim)|_3$, ADE ở dạng chuẩn *3*-F3NF với $FADE\sim = \Pi_{ADE}(AF\sim)|_3$, CDE ở dạng chuẩn *3*-F3NF với $FCDE\sim = \Pi_{CDE}(AF\sim)|_3$. Lược đồ AB ở dạng chuẩn *2*-F3NF với $FAB\sim = \Pi_{AB}(AF\sim)|_2$ và lược đồ BE ở dạng chuẩn *1*-F3NF với $FBE\sim = \Pi_{BE}(AF\sim)|_1$

V. TÁCH BẢO TOÀN THÔNG TIN VỀ DẠNG CHUẨN *K*-FBCNF

Khi cho một lược đồ quan hệ, vấn đề đặt ra ở đây là *có hay không tồn tại một phép tách về dạng chuẩn k-FBCNF và bảo toàn thông tin*? Để trả lời cho câu hỏi, trong phần này chúng tôi giới thiệu thuật toán tách lược đồ là một sự mở rộng của phép tách trong lược đồ quan hệ trong [17].

Thuật toán 5.1. Tách bảo toàn thông tin về dạng chuẩn *k*-FBCNF

Vào: Lược đồ quan hệ U, tập các phụ thuộc hàm mờ $AF\sim$, $k = Min\{\alpha \mid X \sim>_\alpha Y \in AF\sim\}$.

Ra: ρ: Là một phép tách bảo toàn thông tin bao gồm tập các lược đồ con trong đó mỗi lược đồ con đều ở dạng chuẩn *k*-FBCNF với các phụ thuộc hàm là hình chiếu của $AF\sim$ lên lược đồ đó.

Phương pháp:

(1) ρ = {U}

(2) *While* tồn tại lược đồ $Q \in \rho$ không ở dạng chuẩn *k*-FBCNF *do*

(3) *for* each $X \sim>_k A \in \Pi_Q(AF\sim) = \{V \sim>_\beta W \mid V \sim>_\beta W \in AF\sim^+, VW \subseteq Q\}$, $A \notin X$ và X không *k-siêu khoá* của Q, ta thực hiện *do*

(4) $Q_1 = XA$

(5) $Q_2 = Q - A$

(6) $\rho = (\rho - \{Q\}) \cup (\{Q_1\} \cup \{Q_2\})$

Định lý 5.1. Cho lược đồ quan hệ U và tập các phụ thuộc hàm mờ $AF\sim$ xác định trên U, $k = Min\{\alpha \mid X \sim>_\alpha Y \in AF\sim\}$. Thuật toán 5.1 xác định một phép tách $\rho = \{U_1, U_2, \ldots U_m\}$ bảo toàn thông tin và các lược đồ con U_i đều ở dạng chuẩn *k*-FBCNF với tập phụ thuộc hàm tương ứng $\Pi\bigcup_i(AF\sim)$.

Chứng minh

Ta cần chứng minh mỗi U_i là dạng chuẩn *k*-F3NF và ρ bảo toàn thông tin.

Mỗi U_i là dạng chuẩn k-FBCNF

Giả sử rằng tồn tại một lược đồ $Q \in \rho$, Q nhận được bằng cách áp dụng thuật toán 5.1 nhưng không ở dạng chuẩn *k*-FBCNF, khi đó ta chia Q như sau: $Q_1=XA$, $Q_2=Q-A$, khi đó $Q_1 \subset Q$. Vì nếu $Q_1=Q$ thì từ $A \notin X$ ta có X=Q-A.

Từ $X \sim>_\alpha A$ = $Q-A \sim>_\alpha A$ (*) và $Q-A \sim>_\beta Q-A$ ($\forall\beta$). Nếu chọn $\beta \geq \alpha$ thì từ $Q-A \sim>_\beta Q-A \Rightarrow Q-A \sim>_\alpha Q-A$ (**). Từ (*) và (**) ta có $X \sim>_\alpha Q$, vậy X là một *α-siêu khoá* của Q. Vì $\alpha \geq k$ nên $X \sim>_k Q$, hay X cũng là một *k-siêu khoá* của Q, mâu thuẩn. Vậy với mọi $U_i \in \rho$ đều là ở dạng chuẩn *k-FBCNF*.

ρ là phép tách bảo toàn thông tin: Có thể chứng minh tương tự như trong [17].

Ví dụ 5.1. Cho lược đồ quan hệ U={A, B, C, D, E, G, H} và tập phụ thuộc hàm mờ AF~ ={$ABC \sim>_3 D$, $ABC \sim>_4 E$, $ABC \sim>_4 G$, $DE \sim>_3 G$, $D \sim>_2 H$, $H \sim>_1 G$}.

Ta có $ABC \sim>_3 D \Rightarrow ABC \sim>_2 D$, $ABC \sim>_4 E \Rightarrow ABC \sim>_2 E$, $ABC \sim>_4 G \Rightarrow ABC \sim>_2 G$. Vậy ABC là một *2-khoá* của U nên ABC cũng là một *1-khoá* của U. {1=Min(3,4,4,3,2,1)}. Mặc khác ta có $DE \sim>_3 G \Rightarrow DE \sim>_1 G$, DE không là một *1-siêu khoá*. $D \sim>_2 H \Rightarrow D \sim>_1 H$, D không là một *1-siêu khoá*.

$H \sim>_1 G$, H không là một *1-siêu khoá*. Do đó U không ở dạng chuẩn *1*-FBCNF nên áp dụng thuật toán 5.1 để tách U. Từ $DE \sim>_1 G$, DE không là một *1-siêu khoá* nên ta có ρ=(ABCDEH, DEG). Lược đồ ABCDEH có ABC là *1-khoá*, $D \sim>_2 H$ vi phạm dạng chuẩn *1*-FCNBF nên tách lược đồ ABCDEH thành ABCDE và DH. Trong lược đồ DEG ta có DE là một *1-khoá*, $D \sim>_1 G \in \Pi_{DEG}(AF\sim)$ vi phạm dạng chuẩn *1*-FCNBF nên tách DEG thành DE, DG.

Vậy ρ = (ABCDE, DH, DE, DG) phép tách bảo toàn thông tin và các lược đồ đều ở dạng chuẩn *1*-FBCNF.

VI. KẾT LUẬN

Trong bài báo này chúng tôi đã xây dựng một số khái niệm trên mô hình cơ sở dữ liệu mờ theo cách tiếp cận đại số gia tử như *k-khoá*, *k-siêu khoá*…. Từ đó đề xuất các dạng chuẩn *k*-F1NF, *k*-F2NF, *k*-F3CNF *k*-FBCNF. Hai thuật toán tách lược đồ thành các lược đồ con bảo toàn phụ thuộc hàm về dạng chuẩn *k*-F3NF và bảo toàn thông tin về dạng chuẩn *k*-FBCNF cũng được xem xét. Hai thuật toán này là một sự mở rộng của các thuật toán trong [11][13] theo ngữ nghĩa của mô hình cơ sở dữ liệu mờ theo cách tiếp cận mới. Vấn đề xây dựng các ngôn ngữ để thao tác dữ liệu mờ trên mô hình này sẽ được chúng tôi giới thiệu trong bài báo tới.

TÀI LIỆU THAM KHẢO

[1] Nguyễn Cát Hồ, Trần Thái Sơn, *Về khoảng cách giữa các giá trị của biến ngôn ngữ trong đại số gia tử,* Tạp chí Tin học và điều khiển học, T11 S1 (1995), 10-20.

[2] Trần Đình Khang*, Xây dựng hàm đo trên đại số gia tử và ứng dụng trong lập luận ngôn ngữ,* Tạp chí Tin học và điều khiển học, T13 S1 (1997), 16-30.

[3] Nguyen Cat Ho, W.Wechler, *Extended hedge algebras ans their application to fuzzy logic,* Fuzzy Sest and Systems 52 (1992), 259-282.

[4] Nguyen Cat Ho, *Quantifying hedge algebras and interpolation methods in approximate reasoning,* Proceeding of the Inter Workshop on fuzzy information processing FIP2003, Beijing, 105-111.

[5] Nguyễn Công Hào, *Mô hình cơ sở dữ liệu mờ theo cách tiếp cận đại số gia tử.* Kỷ yếu hội thảo quốc gia về các vấn đề chọn lọc công nghệ thông tin và truyền thông, Hải Phòng 2005, 285-293.

[6] Buckles B.P, Petry F.E. (1982), *A Fuzzy representation of data for Relational Databases,* Fuzzy Sets and Systems 7 (3), 213-226.

[7] Chen G, Kerre E. E., vandenbulcke J. (1994), *A Computational algorithm for the FFD Transitive Closure and a Complete Axiomatization of Fuzzy Functional Dependence,* Internatinal Journal of Intelligent Systems 9 (5), 421-439.

[8] H.Thuan, T.T.Thanh (2002), *Fuzzy Functional Dependencies with Linguistic Quantifiers,* Tạp chí Tin học và điều khiển học, Tập 18 (2), 97-108.

[9] 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, T17 S3 (2001),41-47.
[10] Guoqing Chen, Etienne E.Kerre, Jacques Vandenbulcke, *Fuzzy Normal Forms and a Dependency-Preserving decomposition into θ-F3NF,* Proceeding of IEEE World Congress on Computation Intelligence, USA 1994, 156-161
[11] Guoqing Chen, Etienne E.Kerre, Jacques Vandenbulcke, *Introducing keys and integrity rules in a fuzzy relational data model,* 1996 IEEE, 332-335
[12] Guoqing Chen, Etienne E.Kerre, Jacques Vandenbulcke, *An extended Boyce-Codd normal form in fuzzy relational databases,* 1996 IEEE, 1546-1551
[13] Mustafa LLKer Sozat, Adnan Yazici*, A complete axiomatization for fuzzy functional and multivalued dependencies in fuzzy database relations,* Fuzzy Sets and Systems 117 (2001), 161-181.
[14] Guoqing Chen, Etienne E.Kerre, Jacques Vandenbulcke, *On the lossless-join decomposition in a fuzzy relational data model,* Proceedings of International symposium on Uncertainty Modeling & Analysis, Maryland, 1993, 440-446
[15] S.Shensoi, A Melton, *Proximity relations in the fuzzy relational databases*, Fuzzy Sets & Systems 21 (1987), 19-34.
[16] J.D.Ullman, *Principles of Databases Systems*, Computer Sciences Press Inc 1982.

Ngày nhận bài: 02/08/2006

SƠ LƯỢC TÁC GIẢ

NGUYỄN CÔNG HÀO

Sinh ngày 25/07/1976 tại Huế

Tốt nghiệp Đại học Sư Phạm Huế năm 1997, nhận bằng Thạc sỹ năm 2002 tại Đại học Bách khoa Hà Nội, hiện đang là Nghiên cứu sinh tại Viện Công nghệ thông tin - Viện khoa học và Công nghệ Việt Nam.

Nơi công tác hiện nay: Khoa Công nghệ thông tin- Trường Đại học khoa học Huế

Hướng nghiên cứu: Cơ sở dữ liệu mờ, Đại số gia tử.

Email: nchao_hueit@yahoo.com