Tạp chí Khoa học và Công nghệ ***51*** *(2) (2013) 137-152*

PHỤ THUỘC HÀM MỜ TRONG CƠ SỞ DỮ LIỆU QUAN HỆ VỚI DỮ LIỆU NGÔN NGỮ

Nguyễn Cát Hồ[1,*], Vũ Minh Lộc[2], Hoàng Tùng[3], Nguyễn Tân Ân[4]

[1]*Viện Công nghệ thông tin, Viện Hàn lâm KHCNVN, 18 Hoàng Quốc Việt, Cầu Giấy, Hà Nội*
[2]*Trường Đại học CNTT Gia Định*
[3]*Trường Đại học Đồng Nai*
[4]*Khoa Công nghệ thông tin, Trường Đại học Sư phạm Hà Nội*

[*]Email: *ncatho@gmail.com*

Đến Tòa soạn: 15/5/2012; Chấp nhận đăng: 2/6/2013

TÓM TẮT

Cơ sở dữ liệu (CSDL) quan hệ với dữ liệu ngôn ngữ dựa trên đại số gia tử (ĐSGT) đã được giới thiệu và nghiên cứu trong [1, 2], trong đó mỗi dữ liệu ngôn ngữ x được biểu diễn thông qua hai thành phần ngữ nghĩa, thành phần thứ nhất là một giá trị ngữ nghĩa thuộc miền thực $\boldsymbol{D}_A$ của x, thành phần thứ hai là một tập các lân cận dựa trên các khoảng mờ của x. Các phép toán so sánh được định nghĩa dựa trên khái niệm đẳng thức mức k, trong đó k là độ dài biểu diễn chính tắc của các phần tử trong ĐSGT. Với cách nhìn ngữ nghĩa dữ liệu ngôn ngữ như vậy, một kiểu phụ thuộc hàm mờ mới trong CSDL quan hệ với dữ liệu ngôn ngữ sẽ được giới thiệu và nghiên cứu. Chúng có thể được xem như là những ràng buộc ngữ nghĩa trên CSDL trong môi trường mờ của thế giới thực. Bài báo cũng đưa ra một tập các quy tắc suy dẫn cho các phụ thuộc hàm mờ này và chứng minh tính đúng đắn và đầy đủ của chúng.

Từ khóa: CSDL mờ, ngữ nghĩa của dữ liệu ngôn ngữ, đại số gia tử, khoảng mờ, bằng mờ, phụ thuộc hàm mờ.

1. GIỚI THIỆU

Như chúng ta đã biết, phụ thuộc hàm đóng vai trò rất quan trọng trong việc duy trì ràng buộc toàn vẹn dữ liệu đối với các CSDL nói chung và CSDL quan hệ mờ nói riêng. Vì vậy, đã có nhiều những kết quả nghiên cứu về phụ thuộc hàm trong các CSDL mờ với nhiều cách tiếp cận khác nhau dựa trên các cách hình thức hóa khác nhau để biểu diễn thông tin mờ. Bởi vậy cách xây dựng phụ thuộc hàm trong CSDL mờ cũng phụ thuộc vào cách mà ngữ nghĩa của dữ liệu mờ được biểu diễn trong CSDL cũng như vào cách thực hiện so sánh giữa những kiểu dữ liệu khác nhau bao gồm cả dữ liệu mờ. Trong [2] đã giới thiệu tổng quan một số phương pháp định nghĩa những khái niệm so sánh dữ liệu mờ khác nhau dựa trên tập mờ. Những khái niệm này dẫn đến những kiểu đẳng thức không chắc chắn khác nhau trên các miền thuộc tính tương ứng của chúng. Rõ ràng, những kiểu đẳng thức không chắc chắn khác nhau lại kéo theo những kiểu phụ thuộc dữ liệu mờ khác nhau về bản chất cho phép mô tả các ràng buộc ngữ nghĩa khác nhau đối với dữ liệu được cho phép lưu trữ trong một CSDL. Do vậy chúng ta có những cách khác nhau để thao tác với các ràng buộc toàn vẹn dữ liệu.

Có hai kiểu đẳng thức mờ (không chắc chắn) trên miền trị của một thuộc tính A được định nghĩa dựa trên hai loại quan hệ mờ khác nhau. Thứ nhất là quan hệ tương đương mờ hay còn được gọi là quan hệ tương tự, quan hệ giống nhau. Đó là những quan hệ mờ có các tính chất phản xạ, đối xứng và bắc cầu. Thứ hai là quan hệ xấp xỉ được đặc trưng bởi các tính chất phản xạ và đối xứng.

Cho một quan hệ mờ S là quan hệ tương đương hoặc quan hệ xấp xỉ mờ trên miền trị $\boldsymbol{D}_A$ của thuộc tính A. Hàm thành viên của S là μ_S: $\boldsymbol{D}_A \times \boldsymbol{D}_A \to [0,1]$. Quan hệ mờ S xác định một phép so sánh bằng mờ, kí hiệu là $=_S$, trên miền trị $\boldsymbol{D}_A$, theo đó với bất kì hai phần tử a, $b \in \boldsymbol{D}_A$, biểu thức $a =_S b$ nói rằng hai phần tử này bằng nhau với mức độ chân lí là $\mu_S(a, b)$.

Giả sử mỗi thuộc tính A của CSDL mờ được gắn với một quan hệ S_A – bằng nhau. Phép so sánh bằng mờ xác định bởi S_A được kí hiệu là $=_S$ hoặc =, khi không gây nhầm lẫn. Cho $\boldsymbol{X} \subseteq \boldsymbol{U}$, hai bộ t và s có quan hệ S – bằng nhau trên $\boldsymbol{X}$ với mức chân lí được xác định bởi (1-1) [3 - 6]:

$$\mu_S(s[\boldsymbol{X}], t[\boldsymbol{X}]) = \min\{\mu_S(s(A), t(A)): A \in \boldsymbol{X}\} \tag{1-1}$$

Quan hệ S – bằng nhau trên $\boldsymbol{X}$ giữa hai bộ s và t có thể được viết dưới dạng $s[\boldsymbol{X}] =_S t[\boldsymbol{X}]$. Mức độ bằng nhau giữa chúng được tính theo (1-1). Với khái niệm bằng mờ như vậy, phụ thuộc hàm mờ có thể được định nghĩa như sau:

Cho hai tập thuộc tính $\boldsymbol{X}$ và $\boldsymbol{Y}$ của lược đồ quan hệ R, mọi biểu thức có dạng $f = \boldsymbol{X} \sim\!\!\to \boldsymbol{Y}$ được gọi là một phụ thuộc hàm mờ (PTHM) với ngữ nghĩa được định nghĩa như sau: f là đúng trong một quan hệ r, hoặc thỏa mãn quan hệ r, nếu mọi cặp hai bộ s và t trong r, mức bằng nhau của $s[\boldsymbol{X}] =_S t[\boldsymbol{X}]$ phải không lớn hơn mức bằng nhau của $s[\boldsymbol{Y}] =_S t[\boldsymbol{Y}]$. Một cách hình thức điều kiện này có thể được viết như sau:

$$s[\boldsymbol{X}] =_S t[\boldsymbol{X}] \ \leq \ s[\boldsymbol{Y}] =_S t[\boldsymbol{Y}] \tag{1-2}$$

Mặt khác, theo [7 - 9] ngữ nghĩa của PTHM với điều kiện (1-2) có thể được thay thế bởi:

$$s[\boldsymbol{X}] =_{S,\lambda} t[\boldsymbol{X}] \ \Rightarrow \ s[\boldsymbol{Y}] =_{S,\lambda} t[\boldsymbol{Y}] \tag{1-3}$$

trong đó $\lambda \in [0,1]$ là một số cho trước và cố định đối với mỗi CSDL mờ và $s[\boldsymbol{X}] =_{S,\lambda} t[\boldsymbol{X}]$ có nghĩa là hai bộ s và t có quan hệ S – bằng nhau với mức độ chân lí không nhỏ hơn λ. PTHM như vậy được kí hiệu là $\boldsymbol{X} \sim\!\!\to_\lambda \boldsymbol{Y}$. Tất nhiên điều kiện (1-3) yếu hơn điều kiện (1-2) và bởi vậy PTHM theo định nghĩa (1-2) cũng là PTHM theo định nghĩa (1-3).

Trong [4, 6, 8], ngữ nghĩa đẳng thức mờ dựa trên quan hệ xấp xỉ và một số tính chất của PTHM cũng như phụ thuộc đa trị mờ đã được nghiên cứu. Hơn nữa trong [4, 6] một hệ tiên đề (còn được gọi là những quy tắc suy diễn) cho PTHM và/hoặc FMD đã được chứng minh là đúng đắn và đầy đủ. PTHM và PTHM đa trị trên cơ sở của quan hệ tương tự mờ đã được nghiên cứu. Chẳng hạn trong [1] một hệ tiên đề đúng và đầy đủ cho những phụ thuộc mờ này đã được nghiên cứu phát triển.

Cách tiếp cận tập mờ đối với CSDL cho phép người ta có thể linh hoạt đưa ra ngữ nghĩa của các hạng từ (giá trị của biến ngôn ngữ hay giá trị ngôn ngữ) nhờ các tập mờ hay các phân bố khả năng. Nó lí giải vì sao lại có nhiều mô hình CSDL mờ cũng như nhiều kiểu PTHM trong các công trình nghiên cứu, vì chúng phụ thuộc vào quan điểm tiếp cận đối với ngữ nghĩa và cách biểu diễn toán học của các thông tin mờ. Dẫn chứng thêm, trong [5] những hạng từ được biểu diễn bởi các khoảng, nghĩa là các khoảng của miền trị của một thuộc tính và tất nhiên nó sẽ kéo theo một kiểu đẳng thức mờ hay không chắc chắn mới cùng với kiểu phụ thuộc dữ liệu mờ mới. Trong [5] ta thấy việc gán các khoảng cho các hạng từ là khá tự do và không có một hạn chế nào.

Tiếp cận ĐSGT tới ngữ nghĩa của các hạng từ là hoàn toàn khác biệt. Độ đo tính mờ của các hạng từ mờ giữ một vai trò chủ chốt trong việc biểu diễn ý nghĩa của các hạng từ trong CSDL. Điều này phù hợp với bản chất của thông tin mờ. Chúng là những thông tin có thể được xác định theo kinh nghiệm của người sử dụng trong mỗi ứng dụng cụ thể và là cái gốc để xác định ngữ nghĩa của dữ liệu mờ. Bởi vậy, trong [2] mô hình CSDL ngôn ngữ mờ đã được đưa ra và nghiên cứu dựa trên ngữ nghĩa dữ liệu ĐSGT. Trong bài báo này, chúng tôi sẽ sử dụng khái niệm đẳng thức mờ và những kết quả nghiên cứu trong [2] để định nghĩa một kiểu PTHM mới trong CSDL ngôn ngữ mờ. Chúng tôi áp dụng cách biểu diễn ngữ nghĩa của dữ liệu ngôn ngữ dựa trên cơ sở của ĐSGT đã được nghiên cứu trong [10 - 15] để định nghĩa khái niệm đẳng thức mức k hay k-đẳng thức, trong đó k biểu thị mức độ mờ của nó, với k chỉ độ dài của những hạng từ ngôn ngữ được xem là những phần tử của ĐSGT. Khái niệm đẳng thức này, một cách hình thức, là quan hệ tương đương trên miền giá trị của thuộc tính tương ứng. Bởi vậy, khái niệm đẳng thức mờ này rất giống và có cùng tính chất hình thức hóa với khái niệm đẳng thức nhị phân (đẳng thức thông thường) trên cùng miền trị đang xét. Điều đó cho ta một nền tảng tốt để nghiên cứu về một tập PTHM mới. Một hệ tiên đề cho các PTHM mới này cũng sẽ được đưa ra và nghiên cứu.

Bài báo được bố cục như sau: Mục 2 sẽ nhắc lại một số khái niệm và những kí hiệu được dùng trong CSDL quan hệ với dữ liệu ngôn ngữ và biểu diễn ngữ nghĩa dữ liệu của chúng dựa trên ĐSGT đã được nghiên cứu trong [2]. Trong Mục 3, khái niệm PTHM mới và ngữ nghĩa của chúng sẽ được đưa ra và nghiên cứu. Mục 4 sẽ được dành cho việc nghiên cứu một số tính chất của các PTHM đồng thời sẽ đưa ra một hệ tiên đề đúng đắn và đầy đủ cho chúng. Mục 5 trình bày một số kết luận.

2. CSDL QUAN HỆ VỚI DỮ LIỆU NGÔN NGỮ VÀ NGỮ NGHĨA CỦA CHÚNG

Trong phần này chúng ta sẽ nhắc lại một số khái niệm và kí pháp trong [2]. Một mô hình CSDL quan hệ ngôn ngữ được kí hiệu là $\boldsymbol{DB}$, và được định nghĩa bởi

$$\boldsymbol{DB} = \{\boldsymbol{U}, R_1, R_2, \ldots, R_m; Const\}$$

trong đó:

- $\boldsymbol{U} = \{A_1, A_2, \ldots, A_n\}$ là tập vũ trụ của các thuộc tính, một số trong chúng là thuộc tính ngôn ngữ, tức là những thuộc tính có thể nhận thêm các giá trị ngôn ngữ. Giá trị ngôn ngữ là giá trị trong miền trị của biến ngôn ngữ, được gọi chung là các hạng từ. Miền giá trị của biến ngôn ngữ và các hạng từ được mô hình hóa bởi ĐSGT.

- R_i, $i = 1, \ldots, m,$ là các lược đồ quan hệ, mỗi lược đồ quan hệ là một tập con của $\boldsymbol{U}$,

- Const (viết tắt của constraints) là tập các ràng buộc dữ liệu trên $\boldsymbol{DB}$.

Mỗi thuộc tính A_i được gán một miền trị: $\mathbf{D}(A_i) = \boldsymbol{D}_{Ai} \cup LDom(A_i)$, trong đó $\boldsymbol{D}_{Ai}$ là một miền thực thông thường, $LDom(A_i)$ là miền trị ngôn ngữ. $LDom(A_i)$ là tập rỗng nếu A_i không phải là thuộc tính ngôn ngữ. Các phần tử của $\boldsymbol{D}_{Ai}$ sẽ được kí hiệu là $a, b, c, \ldots$ và các phần tử của $LDom(A_i)$ sẽ được kí hiệu là $x, y, z, u, v\ldots$

Thông thường, một bộ t trên U là một ánh xạ t: $\boldsymbol{U} \rightarrow \mathbf{D}(A_1) \cup \ldots \cup \mathbf{D}(A_n)$ sao cho $t(A_i) \in \mathbf{D}(A_i)$, với $1 \leq i \leq n$. Các bộ sẽ được kí hiệu là t, s với chỉ số nếu cần thiết. Cho một bộ t, kí hiệu $t[A_i]$ là giá trị của bộ t tại thuộc tính A_i. Với bất kì tập $\boldsymbol{X} \subseteq \boldsymbol{U}$, kí hiệu $t[\boldsymbol{X}]$ là hạn chế của ánh xạ t trên $\boldsymbol{X}$. Xét một lược đồ quan hệ R trên $\boldsymbol{U}$. Mỗi thể hiện của R được gọi là một quan hệ. Nó là một tập các bộ phân biệt trên R. Các quan hệ của R được kí hiệu là $r[R]$, $s[R]$, ... hay $r, s, \ldots$, khi việc không xuất hiện kí hiệu R không gây nhầm lẫn.

Đối với CSDL mờ, ngữ nghĩa thông tin mờ và phương pháp biểu diễn chúng có ý nghĩa quyết định đến chất lượng tổ chức và phương pháp khai thác dữ liệu. Vì thông tin mờ trong CSDL đều có nguồn gốc ngôn ngữ, nên ý tưởng của CSDL ngôn ngữ là cho phép lưu trữ dữ liệu ngôn ngữ với ngữ nghĩa vốn có của chúng nhờ việc áp dụng ĐSGT. Trong [2] đã đưa ra phương pháp biểu diễn ngữ nghĩa các giá trị ngôn ngữ bằng tập hữu hạn các khoảng lân cận ngữ nghĩa. Cách biểu diễn như vậy tạo nền tảng cho phép thao tác dữ liệu mờ thống nhất với dữ liệu kinh điển. Một điều quan trọng là các khoảng lân cận ngữ nghĩa của các từ ngôn ngữ có thể được sinh ra tự động khi cho các tham số tính mờ của các thuộc tính ngôn ngữ của CSDL, nghĩa là độ đo tính mờ của một trong hai phần tử sinh và của $(p + q - 1)$ gia tử. Thường $p = q = 2$, và do đó ta chỉ có 4 tham số. Như vậy, khác với CSDL mờ theo tiếp cận tập mờ, người thiết kế phải xây dựng các tập mờ cho *từng dữ liệu ngôn ngữ*, nghĩa là có bao nhiêu hạng từ thì phải xây dựng ngần ấy tập mờ, đối với CSDL ngôn ngữ nhà thiết kế chỉ phải lựa chọn giá trị của 4 tham số tính mờ.

Với mỗi thuộc tính ngôn ngữ A_j, ta giả thiết tập các giá trị ngôn ngữ là tập $X_{j,(kj)}$ các phần tử của một ĐSGT có độ dài $\le k_j$, nghĩa là số lần xuất hiện các gia tử trong các hạng từ ngôn ngữ không vượt quá $k_j - 1$. Kí hiệu $X_{j,k}$ là tập con của $X_{j,(kj)}$ gồm tất cả các hạng từ có độ dài đúng bằng k, $k \le k_j$. Một khi các giá trị tham số tính mờ của thuộc tính A_j đã xác định, các đại lượng sau đây hoàn toàn tính được bằng các thuật toán:

(i) Các giá trị ngữ nghĩa định lượng của các hạng từ của $X_{j,(kj)}$, $\{\upsilon(x) : x \in X_{j,(kj)}\}$, với υ là ánh xạ định lượng ngữ nghĩa xác định bởi các tham số tính mờ.

(ii) Tập các khoảng tính mờ của mỗi hạng từ trong $X_{j,k}$. Chúng lập thành một phân hoạch của miền thuộc tính.

(iii) Biểu diễn toán học của hạng từ: Trong [2], mỗi từ $x \in X_{j,(kj)}$ được biểu diễn bằng hệ các khoảng lân cận, kí hiệu là $\mathrm{IRp}(x)$, được định nghĩa như sau:

$$\mathrm{IRp}(x) = \{\mathrm{O}_{\min,k}(x) : 1 \le k \le k_j\} \tag{3.5}$$

trong đó $\mathrm{O}_{\min,k}(x)$ là hợp của hai khoảng tính mờ mức k' kề với giá trị $\upsilon(x)$ (chỉ mục "min" nói rằng nó là tập nhỏ nhất mức k có thể được sinh ra như vậy).

Biểu diễn (3.5) mang ý nghĩa quan trọng sau đây:

- Nó biểu diễn được tính mờ của hạng từ x vì nó được tính từ các tham số tính mờ. Các giá trị trong khoảng $\mathrm{O}_{\min,k}(x)$ được xem là tương tự hay phù hợp với ngữ nghĩa của x, hay với $\upsilon(x)$, ở mức k (giống như tập mức của tập mờ).

- Nó là cơ sở chuyển đổi một câu hỏi ngôn ngữ mờ về câu hỏi kinh điển. Chẳng hạn, [2] chỉ ra rằng, điều kiện câu hỏi *Tuổi*(a) = "*khá trẻ*" được thỏa mãn ở mức k được kiểm chứng bằng điều kiện $\upsilon(Tuổi(a)) \in \mathrm{O}_{\min,k}(khá\ trẻ)$, trong đó nếu *Tuổi*($a$) là kinh điển thì quy ước là $\upsilon(Tuổi(a)) = Tuổi(a)$.

(iv) Các khoảng tương tự $\mathrm{S}_k(x)$ của tất cả các từ trong $X_{j,(k)}$, với $k = 1, \ldots, k_j$: Hệ khoảng này được xây dựng từ các khoảng tính mờ mức $k' > k$ sao cho:

- Mỗi $\mathrm{S}_k(x)$ là hợp của một số khoảng tính mờ mức k' và chỉ chứa duy nhất một giá trị định lượng $\upsilon(x)$, $x \in X_{j,(k)}$;

- Tập $\{\mathrm{S}_k(x): x \in X_{j,(k)}\}$ là một phân hoạch của đoạn miền tham chiếu [0,1] của A_j.

Hệ khoảng này xác định một quan hệ tương đương nhị phân (crisp) được kí hiệu là $=_k$, với ngữ nghĩa về sự tương đương hay về *đẳng thức ở mức k*. Giá trị k càng lớn nói rằng các từ độ dài k càng có nghĩa cá biệt và sự xấp xỉ ngữ nghĩa càng gần nhau hơn. Cần nhớ rằng phân hoạch mức $k > 0$ mịn hơn phân hoạch mức $k - 1$, nghĩa là mỗi lớp tương đương của phân hoạch mức k

là tập con của một lớp tương đương của phân hoạch mức k– 1. Ngược lại, mỗi lớp tương đương của phân hoạch mức k – 1 *là hợp của các lớp tương đương* của phân hoạch mức k.

Quan hệ $=_k$ có thể hiểu là đẳng thức mờ mức k trong CSDL ngôn ngữ. k càng lớn thì miền giá trị của lớp tương đương càng gần nhau. Như vậy, khác với CSDL mờ thông thường, trong CSDL ngôn ngữ *có nhiều quan hệ đẳng thức mờ*.

(v) Ngữ nghĩa của đẳng thức mờ mức k được định nghĩa như sau [2]:

Hai giá trị $t[A]$, $s[A]$ thuộc $\underline{\mathbf{D}}(A)$ được gọi là bằng nhau mức k, kí hiệu $t[A] =_k s[A]$, nếu một trong các điều kiện sau đây thỏa mãn:

(a) $t[A]$ và $s[A]$ đồng nhất nhau về kí hiệu;

b) Tồn tại một lớp tương tự $S_k(u)$ của quan hệ tương tự S_k mức k sao cho

$$O_{\min,k}(t[A]) \subseteq S_k(u) \text{ và } O_{\min,k}(s[A]) \subseteq S_k(u).$$

Tuy nhiên, việc kiểm chứng xem $t[A] =_k s[A]$ theo điều kiện (b) ở trên có phần phức tạp. Trong [2] đã chỉ ra rằng, với ngữ nghĩa định lượng đã cho, và với bất kì hai giá trị $t[A]$ và $s[A]$ nào của thuộc tính ngôn ngữ A, tiêu chuẩn kiểm tra đẳng thức $t[A] =_k s[A]$ là các điều kiện sau:

Tồn tại một khoảng tương tự mức k, $S_k(u)$, của A sao cho

- Nếu $t[A]$ và $s[A]$ cùng là giá trị kinh điển thì ta phải có $t[A], s[A] \in S_k(u)$;
- Nếu chỉ một trong hai $t[A]$ và $s[A]$ là ngôn ngữ, ví dụ là $t[A]$, thì ta $\upsilon(t[A]), s[A] \in S_k(u)$;
- Nếu cả hai giá trị $t[A]$ và $s[A]$ cùng là ngôn ngữ, thì $\upsilon(t[A]), \upsilon(s[A]) \in S_k(u)$.

Để tính các khoảng tương tự mức $k > 0$ tùy ý, ta sử dụng các khoảng tính mờ mức $k' > k$ có đủ số lượng sao cho giá trị định lượng của mỗi từ trong tập $X_{j,(k)}$ có ít nhất 2 khoảng tính mờ mức k' của riêng nó. Trong [2] đã chứng tỏ rằng điều kiện đối với k' là với p, $q > 1$, thì $k' > k$. Trái lại thì $k' > k+1$.

Tóm lại, trong CSDL ngôn ngữ chúng ta có:

(i) Các hạng từ ngôn ngữ được biểu diễn bằng hệ lân cận các khoảng;

(ii) Trên mỗi thuộc tính ngôn ngữ, có nhiều mối quan hệ đẳng thức mức k, $0<k \leq k_j$.

3. PHỤ THUỘC HÀM MỜ TRONG CSDL NGÔN NGỮ

Như đã thảo luận trong Mục 1, ngữ nghĩa mới của quan hệ đối sánh mờ sẽ kéo theo khái niệm phụ thuộc dữ liệu mới. Trong mục này chúng ta sẽ đưa ra khái niệm PTHM trong CSDL ngôn ngữ và khảo sát ngữ nghĩa của chúng. Xét CSDL ngôn ngữ ***DB*** [2] với ngữ nghĩa dữ liệu được biểu diễn như trên. Đặt $\boldsymbol{U}_{\text{Ling}} = \{A \in \boldsymbol{U}$: A là một thuộc tính ngôn ngữ$\}$ và $\boldsymbol{U}_{\text{CL}} = \boldsymbol{U} \setminus \boldsymbol{U}_{\text{Ling}}$, là tập các thuộc tính thông thường. Mỗi CSDL ngôn ngữ mờ là mô hình lưu giữ thông tin không chắc chắn của một thế giới thực. Trong một thế giới thực như vậy, có thể tồn tại một số phụ thuộc dữ liệu liên quan tới thông tin không chắc chắn, gọi là PTHM. Cũng như đối với CSDL mờ thông thường ta có thể đưa ra một số ví dụ sau để minh họa ràng buộc mờ:

Ví dụ 3.1: Trước khi tổ chức thành các bảng đấu, ban tổ chức một giải bóng đá nào đó thường phải phân nhóm trình độ các đội thành nhóm hạt giống, nhóm trình độ trung bình, yếu, … Lược đồ quan hệ Xeploaitrinhdo của CSDL Giải đấu X, gồm các thuộc tính MaD (Mã đội), TenD (Tên đội), Thanhtichdoikhang (Thành tích đối kháng), Xeploai (Xếp loại). Trong lược đồ quan hệ này, một cách tự nhiên, dễ thấy có PTHM "những đội có thành tích đối kháng trong lịch sử (tốt, rất tốt, khá, kém, …) tương tự nhau thì sẽ được xếp vào cùng nhóm".

Ví dụ 3.2: Lược đồ quan hệ Nhomkhaithac (Nhóm khai thác) của một CSDL quản lí rừng, bao gồm các thuộc tính MaCay (Mã cây), TenCay (Tên cây), Muctuoi (Mức tuổi của cây), Kichthuoc (Kích thước), Nhom (Nhóm). Trong lược đồ quan hệ này cũng dễ thấy có PTHM "những cây có cùng tên, tương tự nhau về tuổi (tuổi cây có thể là: già, rất già, trẻ, khá trẻ, ...) và tương tự nhau về kích thước (kích thước có thể là: lớn, khá lớn, bé, tương đối bé,…) sẽ được xếp vào cùng nhóm khai thác".

Hai phụ thuộc dữ liệu mà chúng ta thấy trong ví dụ 3.1 và 3.2 là những PTHM. Chúng khác với phụ thuộc hàm rõ hay còn gọi là phụ thuộc hàm kinh điển (PTH) ở chỗ: hai bộ t, s được xem là thỏa mãn PTH $X \rightarrow Y$ nếu t và s *"bằng nhau"* trên X thì sẽ *"bằng nhau"* trên Y. Còn hai bộ như vậy được xem thỏa mãn một PTHM với vế trái là X và vế phải là Y nếu chúng *"tương tự nhau"* trên X thì sẽ *"tương tự nhau"* trên Y. Sự khác biệt ở đây thuộc về ngữ nghĩa của khái niệm "tương tự nhau". "bằng nhau" đối với PTH có nghĩa rằng các phần tử là đồng nhất. Còn "tương tự nhau" đối với PTHM có nghĩa các phần tử thỏa một quan hệ tương tự ở một mức nào đấy.

Trong CSDL mờ thông thường có nhiều cách tiếp cận để định nghĩa quan hệ "tương tự", nhưng hầu hết chúng không thỏa mãn quan hệ bắc cầu. May mắn là trong CSDL ngôn ngữ các quan hệ bằng nhau mức k đều là các quan hệ phản xạ, đối xứng và bắc cầu. Do vậy, chúng giống như quan hệ đẳng thức thông thường, ngoại trừ *chúng không phải là đồng nhất thức*. Bởi vậy, mọi thuộc tính ngôn ngữ A của CSDL trong truy vấn sẽ được kết hợp với một tập các quan hệ tương tự mức k, gọi là k–đẳng thức hay bằng nhau mức k, được kí hiệu là "$=_{A,k}$", với $0 < k \leq k_A$, và k_A là độ dài lớn nhất của các hạng từ trong $LDom(A)$. Để thuận tiện, mỗi thuộc tính thông thường sẽ được kết hợp với một đẳng thức mức 0, $=_0$, nó là một quan hệ bằng nhau thông thường, vì vậy, chỉ số 0 có thể bỏ qua và ta có kí hiệu $=$.

Vì đối với thuộc tính ngôn ngữ chúng ta có nhiều đẳng thức mờ, nên để biết rằng ta sử dụng một đẳng thức nào trong những đẳng thức ấy, ta cần đưa ra một số kí pháp. Chúng ta giả thiết $\boldsymbol{\kappa}$ là một hàm, $\boldsymbol{\kappa}: \boldsymbol{U} \rightarrow \mathrm{N}$ có tính chất nếu A là thuộc tính kinh điển thì $\boldsymbol{\kappa}(A) = 0$, nghĩa là quan hệ tương tự của A là đẳng thức thông thường. Nếu ngược lại thì $k_A \geq \boldsymbol{\kappa}(A) > 0$, nghĩa là quan hệ tương tự của A là $=_{\boldsymbol{\kappa}(A)}$. Như vậy hàm $\boldsymbol{\kappa}$ cho biết các mức tương tự của các đẳng thức sử dụng đối với các thuộc tính của $\boldsymbol{U}$. Với bất kì $\boldsymbol{X} \subseteq \boldsymbol{U}$, ta kí hiệu $\boldsymbol{\kappa}_X$ hay $\boldsymbol{\kappa}(\boldsymbol{X})$ là hạn chế của $\boldsymbol{\kappa}$ trên $\boldsymbol{X}$.

Cho $\boldsymbol{\kappa}$ và $\boldsymbol{\kappa}'$, ta viết $\boldsymbol{\kappa}_X \geq \boldsymbol{\kappa}'_X$ nếu $\boldsymbol{\kappa}(A) \geq \boldsymbol{\kappa}'(A)$, với mọi $A \in \boldsymbol{X}$.

Cho $\boldsymbol{X} \subseteq \boldsymbol{U}$, chúng ta nói rằng hai bộ t, s trên $\boldsymbol{U}$ là bằng nhau với mức tương tự đã cho $\boldsymbol{\kappa}$, và viết $t[\boldsymbol{X}] =_{\kappa} s[\boldsymbol{X}]$, nếu chúng ta có $t[A] =_{\boldsymbol{\kappa}(A)} s[A]$, với mọi $A \in \boldsymbol{X}$. Trong trường hợp này chúng ta cũng nói rằng t và s là *bằng nhau mức* $\boldsymbol{\kappa}$ hoặc *tương tự mức* $\boldsymbol{\kappa}$ trên $\boldsymbol{X}$. Trong trường hợp ngược lại, chúng ta viết $t[\boldsymbol{X}] \neq_{\kappa} s[\boldsymbol{X}]$. Với cách kí hiệu như vậy, chúng ta đưa ra định nghĩa sau về phụ thuộc hàm mờ.

Định nghĩa 3.1. Cho $\boldsymbol{DB}$ là một CSDL ngôn ngữ và R là một lược đồ quan hệ của $\boldsymbol{DB}$. Mọi biểu thức có dạng $f = \boldsymbol{X} \rightarrow_{\boldsymbol{\kappa}} \boldsymbol{Y}$ được gọi là phụ thuộc hàm mờ với mức $\boldsymbol{\kappa}$ (hoặc viết tắt là $\boldsymbol{\kappa}$-PTHM) và nó xác định một ràng buộc đối với lược đồ R như sau: với mỗi quan hệ r trên R, r phải thỏa f, nghĩa là

$$(\forall t, s \in \mathrm{r})(t[\boldsymbol{X}] =_{\kappa} s[\boldsymbol{X}] \ \Rightarrow \ t[\boldsymbol{Y}] =_{\kappa} s[\boldsymbol{Y}]).$$

Như vậy, chúng ta thấy ngoại trừ tính mờ, tính không chắc chắn của các đẳng thức của các thuộc tính ngôn ngữ, hình thức hóa hay cú pháp của khái niệm PTHM rất giống với PTH kinh điển. Như chúng ta sẽ thấy, các PTHM như vậy rất gần gũi với các PTH thông thường.

Trong trường hợp không cần thiết phải chỉ rõ κ, một κ-PTHM f được gọi đơn giản là PTHM f.

Ví dụ 3.3: Xét CSDL "phân luồng học sinh năm cuối THCS" trong đó có lược đồ quan hệ phân nhóm học sinh, $R1$ = {#MAHS, TEN, TOAN, VAN, XEPHANG, SUCKHOE, #NHOM}, ở đây #MAHS (Mã học sinh), #NHOM (Nhóm) là những thuộc tính thông thường, TOAN (Toán), VAN (Văn), XEPHANG (Xếp hạng), SUCKHOE (Sức khỏe) là những thuộc tính ngôn ngữ. Xét quan hệ r của $R1$ được cho trong bảng 1.

Bảng 1. Một thể hiện r của lược đồ quan hệ R1.

#MAHS	TEN	TOAN	VAN	XEPHANG	SUCKHOE	#NHOM
1	Chương	*high*	65	82	*high*	G2
2	Trí	75	*Lhigh*	71	*Rhigh*	G2
3	Phúc	90	95	*Ehigh*	*Lhigh*	G1
4	Hoa	*Rhigh*	80	78	*Elow*	G4
5	Quân	48	70	65	*Llow*	G3
6	Tuệ	45	*high*	*Lhigh*	*Llow*	G3
7	Tiếu	90	*Ehigh*	93	*Lhigh*	G1
8	Lập	35	*Low*	31	*Vhigh*	G5

Trong ngữ cảnh của thế giới thực đang xét, một cách tự nhiên, ta có thể giả thiết r tuân theo quy tắc "nếu các giá trị điểm trên thuộc tính TOAN và VAN của hai học sinh tương tự nhau, thì xếp hạng của chúng cũng sẽ tương tự nhau". Thêm nữa, nếu chúng ta giả thiết việc xếp nhóm học sinh theo quy tắc "nếu các học sinh tương tự nhau về XEPHANG và SUCKHOE, thì chúng sẽ được xếp vào cùng nhóm". Theo đó, chúng ta có thể giả thiết lược đồ quan hệ $R1$ tuân theo hai PTHM, f = {TOAN, VAN} $\rightarrow_{\kappa}$ XEPHANG và g = {XEPHANG, SUCKHOE} $\rightarrow_{\kappa'}$ #NHOM, giả sử với κ = {κ(TOAN) = 2, κ(VAN) =2, κ(XEPHANG) = 2} và κ' = {κ'(XEPHANG, SUCKHOE) = 2, κ'(#NHOM) = 0}.

Ta sẽ thấy quan hệ r trên $R1$ thỏa mãn các PTHM f và g theo ngữ nghĩa ngôn ngữ đã trình bày ở trên với một ngữ nghĩa định lượng phù hợp. Thật vậy, để đơn giản việc trình bày chúng ta giả thiết các thuộc tính ngôn ngữ này có cùng miền tham chiếu thực là đoạn [0, 100], và có cùng miền các hạng từ, tức là được nhúng vào cùng một ĐSGT tuyến tính $\mathbf{AX} = (X, G, C, H, \leq)$, trong đó $G = \{low, high\}$, $C = \{\mathbf{0}, \mathbf{W}, \mathbf{1}\}$, $H = \{E, V, R, L\}$, trong đó E, V, R và L là các kí hiêu viết tắt theo thứ tự của *Extremely*, *Very*, *Rather* và *Little*, và $\leq$ là một phép toán thứ tự dựa trên ngữ nghĩa. Ngữ nghĩa định lượng ĐSGT được xác định với việc lựa chọn các giá trị tham số tính mờ như sau: $fm(high) = fm(c^+) = 0.45$, $fm(low) = fm(c^-) = 0,55$, $\mu(R) = 0,25$, $\mu(L) = 0,25$, $\mu(V) = 0,30$ và $\mu(E) = 0,20$. Như vậy chúng ta có $\alpha = \beta = 0,5$.

Đầu tiên, chúng ta tính các giá trị ngữ nghĩa của những hạng từ sau cho *tất cả* các thuộc tính ngôn ngữ (theo giả thiết ở trên) với lưu ý rằng do $\alpha = \beta = 0,5$, $\upsilon_{A,r}(x)$ là điểm trong của khoảng $\mathfrak{I}_r(x)$, ở đây r là chỉ số dưới chỉ rõ các giá trị thuộc miền thực $[a, b]$ của thuộc tính A trong truy vấn, thay vì lấy giá trị trong [0,1], như quy ước ở [2].

$\upsilon_{A,r}(high) = [fm(low) + 0,5\times fm(high)]\times 100 = [0,55 + 0,5\times 0,45]\times 100 = 77,5$

$\upsilon_{A,r}(low) = 0.5 \times fm(low) \times 100 = 0,5 \times 0,55 \times 100 = 27,5$

$\upsilon_{A,r}(Llow) = \upsilon_{A,r}(low) + (\mu(R) \times fm(low) + \alpha \times \mu(L) fm(low)) \times 100$

$= 27,5 + (0,25 \times 0,55 + 0,5 \times 0,25 \times 0,55) \times 100 = 48,13$

$\upsilon_{A,r}(Lhigh) = [fm(low) + 0.5 \times fm(Lhigh)] \times 100 = [0,55 + 0,5 \times 0,25 \times 0,45] \times 100 = 60,625$

$\upsilon_{A,r}(Rhigh) = [fm(low) + fm(Lhigh) + 0,5 \times fm(Rhigh)] \times 100$

$= [0,55 + 0,25 \times 0,45 + 0,5 \times 0,25 \times 0,45] \times 100 = 71,875$

$\upsilon_{A,r}(Vhigh) = \upsilon_{A,r}(high) + 0,5 \times fm(Vhigh) \times 100 = 77,5 + 0,5 \times 0,30 \times 0,45 \times 100 = 84,25$

$\upsilon_{A,r}(Ehigh) = 100 - 0,5 \times fm(Ehigh) \times 100 = 100 - 0,5 \times 0,20 \times 0,45 \times 100 = 95,5$

Bây giờ chúng ta sẽ tính các khoảng tương tự mức 2 của miền [0, 100] như sau:

$S_{2,r}(high) = \theta_{2,r}(high) = \mathfrak{I}_r(Rhigh) \cup \mathfrak{I}_r(Vhigh)$

$= (77,5 - 0,25 \times 0,45 \times 100, 77,5 + 0,30 \times 0,45 \times 100] = (66,25, 91,00].$

$S_{2,r}(Ehigh) = \mathfrak{I}_r(Ehigh) = (100 - 0,20 \times 0,45 \times 100, 100] = (91, 100].$

$S_{2,r}(Lhigh) = \mathfrak{I}_r(Lhigh) = (0,55 \times 100, 0,55 \times 100 + 0,25 \times 0,45 \times 100] = (55, 66,25].$

$S_{2,r}(Llow) = \mathfrak{I}_r(RLlow) \cup \mathfrak{I}_r(VLhigh) = ((\upsilon_{A,r}(Llow) - \mu(R) \times \mu(L) \times fm(low)) \times 100, (\upsilon_{A,r}(Llow) + \mu(V) \times \mu(L) \times fm(low)) \times 100]$

$= (48,13 - (0,25 \times 0,25 \times 0,55) \times 100, (48,13 + 0,30 \times 0,20 \times 0,55) \times 100] = (44,7, 51,43].$

Bây giờ chúng ta có thể kiểm tra những phụ thuộc hàm được đề cập ở trên có thỏa mãn bởi r với κ và κ' đã cho hay không dựa vào (iv) Mục 2. Ví dụ, chúng ta thấy rằng $high =_{\kappa} 75$ (vì 75, $\upsilon_{A,r}(high) \in S_{2,r}(high) = (66.25, 91.00]$) và $65 =_{\kappa} Lhigh$ (vì 65, $\upsilon_{A,r}(Lhigh) \in S_{2,r}(Lhigh) = (55, 66,25]$) và bởi vậy t_1[TOAN, VAN] $=_{\kappa}$ t_2[TOAN, VAN]. Tương tự, với thuộc tính XEPHANG chúng ta có $82 =_{\kappa'} 71$ và điều này chứng tỏ t_1 và t_2 thỏa mãn f. Với thuộc tính SUCKHOE ta có $high =_{\kappa'} Rhigh$ và vì vậy t_1 và t_2 cũng thỏa g. Hoàn toàn tương tự như vậy, chúng ta có thể kiểm tra thấy r thỏa các PTHM f và g.

Chúng ta cũng có thể thấy rằng PTHM {TOAN, VAN} $\rightarrow_{\kappa}$ XEPHANG biểu thị đầy đủ ngữ nghĩa mong muốn và nó khác biệt với ngữ nghĩa của PTH có cùng vế trái và vế phải, {TOAN, VAN} $\rightarrow$ XEPHANG, bởi vì rõ ràng là hai bộ t và s mà bằng nhau kinh điển, tức là bằng nhau về mặt kí hiệu trên A, dù là kí hiệu giá trị ngôn ngữ, thì *chúng phải được xếp cùng hạng*. Nghĩa là phụ thuộc hàm kinh điển {TOAN, VAN} $\rightarrow$ XEPHANG cũng thỏa trong r.

Một cách hình thức, theo định nghĩa của PTHM ở trên, dễ thấy PTHM kiểu này khác với kiểu PTHM thông thường. Bây giờ chúng ta sẽ cùng bàn luận về một số đặc trưng ngữ nghĩa của kiểu PTHM mới này. Khác với ngưỡng λ, $0 < \lambda \leq 1$, trong trường hợp của PTHM thông thường, nó là một ngưỡng chung cho tất cả các PTHM thông thường trong truy vấn, trong nghiên cứu này, mỗi PTHM được kết hợp với một mức κ (một vector mức) và những mức này diễn đạt những đặc điểm ngữ nghĩa rõ ràng của kiểu PTHM mới. Nếu $\kappa(\boldsymbol{XY}) \equiv 0$, có nghĩa là $\boldsymbol{X}$ và $\boldsymbol{Y}$ không bao gồm các thuộc tính ngôn ngữ, thì $\boldsymbol{X} \rightarrow_{\kappa} \boldsymbol{Y}$ là một phụ thuộc hàm kinh điển (PTH). Tuy nhiên, ngay cả khi những tập này bao gồm những thuộc tính ngôn ngữ thì vẫn có thể tồn tại phụ thuộc hàm kinh điển PTH $\boldsymbol{X} \rightarrow \boldsymbol{Y}$ xác định trên cùng các tập $\boldsymbol{X}$ và $\boldsymbol{Y}$, như đã đề cập ở trên. Như vậy. với bất kì thuộc tính ngôn ngữ A nào đều có những đẳng thức mờ, $=_{\kappa(A)}$, $\kappa_A > 0$, đồng thời vẫn có một đẳng thức kinh điển (bằng nhau trên kí hiệu) = trên $\mathbf{D}(A) = \boldsymbol{D}_{Ai} \cup LDom(A_i)$, được xem như một đồng nhất về cú pháp.

Từ cách nhìn này, chúng ta có thể xem các PTH kinh điển như ràng buộc ở mức cú pháp hay *ràng buộc cú pháp*, bởi chúng kéo theo phép bằng của dữ liệu trong truy vấn, và chúng biểu thị một kiểu ràng buộc ngữ nghĩa của thế giới thức. Các PTHM mới là những *ràng buộc ở mức ngữ nghĩa* ngôn ngữ khi người sử dụng biểu thị các ràng buộc của thế giới thực bằng ngôn ngữ.

Ví dụ 3.4: Xét CSDL "Hướng nghiệp", lưu trữ thông tin hướng nghiệp cho học sinh phổ thông. Lược đồ quan hệ phân nhóm nghề nghiệp, *R2*, bao gồm các thuộc tính MAHS (Mã học sinh), TENHS (Tên học sinh), DMTN (Điểm các môn tự nhiên), DMXH (Điểm các môn xã hội), DNK (Điểm ngoại khóa), NHOM (Nhóm nghề nghiệp), trong đó các thuộc tính DMTN, DMXH, DMNK là các thuộc tính ngôn ngữ.

Bảng 2. Quan hệ *r* trên *R2*.

MAHS	TENHS	DMTN	DMXH	DNK	NHOM
H001	Hùng	4.5	8.0	3.5	Nhóm 5
H002	Hoan	7.0	5.5	7.0	Nhóm 3
H003	Tước	*Rather high*	*Rather low*	Cao	Nhóm 3
H004	Chuân	*Little low*	7.5	7.0	Nhóm 7
H005	Phương	*Very high*	*Little low*	*Rather low*	Nhóm 6
H006	Bàng	4.37	*Very high*	3.9	Nhóm 5
H007	Sùng	4.95	*High*	*Very high*	Nhóm 7

Xét quan hệ r của *R2* được cho trong bảng 2. Theo tính chất và mục đích của CSDL, ta có thể thấy trong *R2* tồn tại phụ thuộc dữ liệu mờ "những học sinh có điểm tổng kết các môn ở DMTN, DMXH và DNK tương tự nhau thì sẽ được định hướng vào chung một nhóm nghề nghiệp". Nghĩa là ta phải có các PTHM *h*={DMTN, DMXH, DNK} $\rightarrow_{\kappa}$ NHOM. Giả sử với κ[DMTN] = 2; κ[DMXH] = 2; κ[DNK] = 2; κ[MAHS] = κ[TENHS] = 0; κ[NHOM] = 0. Với r trong Bảng 2, ta sẽ thấy nó thỏa mãn PTHM *h*, thật vậy, để kiểm chứng, chúng ta giả thiết các thuộc tính ngôn ngữ này có cùng miền thực là đoạn [0, 10], và có cùng miền các hạng từ, tức là được nhúng vào cùng một ĐSGT tuyến tính $\mathbf{AX} = (X, G, C, H, \leq)$, ở đây $G = \{low, high\}$, $C = \{\mathbf{0}, \mathbf{W}, \mathbf{1}\}$, $H = \{L, R, V, E\}$ trong đó, *L, R, V* và *E* là các kí hiệu viết tắt theo thứ tự của *Little, Rather, Very, Extreamly*. Để định lượng ĐSGT này, chúng ta giả thiết $fm(high) = fm(c^+) = 0{,}6$, $fm(low) = fm(c^-) = 0{,}4$, $\mu(L) = 0{,}15$, $\mu(R) = 0{,}25$, $\mu(h) = 0{,}20$ và $\mu(Extreamly) = 0{,}40$. Như vậy chúng ta có $\alpha = 0{,}4$ và $\beta = 0{,}6$.

Ta cũng quy ước $\upsilon_{A,r}(x)$ lấy giá trị thuộc miền thực [*a*, *b*] thay vì lấy giá trị trong [0,1] như quy ước trong ví dụ 3.3.

$\upsilon_{DMTN,r}(low) = \beta\times fm(low) = 0{,}6\times0{,}4\times10 = 2{,}4$;

$S_{2,DMTN,r}(Llow) = \theta_{2,DMTN,r}(Llow) = \Im_r(VLlow) \cup \Im_r(LLlow) = ((fm(low) + \beta^* fm(Llow) - fm(VLlow), (fm(low) + \beta\times fm(Llow) + fm(LLlow)] = ((0{,}4 + 0{,}15\times0{,}6) - 0{,}2\times0{,}15\times0{,}6, (0{,}4 + 0{,}15\times0{,}6) + 0{,}15\times0{,}15\times0{,}6] \times (10 - 0) = (4{,}72, 5{,}035]$

$\upsilon_{DMXH,r}(high) = fm(low) + fm(high) \times\alpha = (0{,}4 + 0{,}4\times0{,}6) \times 10 = 6{,}4$;

$S_{2,DMXH,r}(high) = \theta_{2,DMXH,r}(high) = \Im_r(Lhigh) \cup \Im_r(Vhigh) = (\upsilon_{DMXH,r}(high) - \mu(L) \times fm(high)), (\upsilon_{DMXH,r}(high) + \mu(V) \times fm(high)] = ((0,64 - 0,15\times0,6,\ 0,64 + 0,2\times0,6]) \times10 = (5,5,\ 7,6]$

$S_{2,DMXH,r}(Vhigh) = \theta_{2,DMXH,r}(Vhigh) = \Im_r(LVhigh) \cup \Im_r(VVhigh) = (\upsilon_{DMXH,r}(high) + fm(RVhigh), (\upsilon_{DMXH,r}(high) + fm(Vhigh)) - fm(EVhigh)] = ((0,64 + 0,25\times0,2\times0,6,\ 0,64 + 0,2\times0,6 - 0,4\times0,2\times0,6]) \times10 = (6,7,\ 7,12]$.

Qua các kết quả tính toán ở trên ta thấy bộ thứ 4, t_4 và bộ thứ 7, t_7 có: t_4[DMTN] $=_2$ t_7[DMTN]; t_4[DMXH] $=_2$ t_7[DMXH]; t_4[DNK] $=_2$ t_7[DNK] suy ra t_4[DMTN, DMXH, DNK] $=_2$ t_7[DMTN, DMXH, DNK] đồng thời ta lại có t_4[NHOM] $=_0$ t_7[NHOM], điều đó có nghĩa là t_4 và t_7 thỏa PTHM *h*. Tính toán tương tự như vậy ta có thể chỉ ra với hai cặp bộ bất kì của r chúng đều thỏa PTHM *h*. Vậy, r được cho như trong bảng 2 thỏa PTHM *h*.

Lưu ý rằng, mặc dù $t[X] = s[X]$ suy ra $t[X] =_\kappa s[X]$, với mọi κ, nhưng có thể dễ dàng chỉ ra bằng ví dụ rằng nếu ta có PTH $f = X \rightarrow Y$ thì không thể suy ra κ-PTHM $f' = X \rightarrow_\kappa Y$, với một mức κ nào đó. Nghĩa là, tồn tại một quan hệ r saocho "r thỏa mãn PTH *f* nhưng không thỏa mãn κ-PTHM *f*'. Ngược lại, nếu có κ-PTHM $f = X \rightarrow_\kappa Y$ cũng không suy ra được PTH $g = X \rightarrow Y$.

Tuy nhiên, như một hệ quả trực tiếp từ định nghĩa PTHM ta có mệnh đề sau:

Mệnh đề 3.1. *Nếu một quan hệ* r *thỏa mãn* PTHM $f = X \rightarrow_\kappa Y$, *thì* r *thỏa mãn* PTHM $f' = X \rightarrow_{\kappa'} Y$, *với bất kì* κ' *sao cho* $\kappa'(X) \geq \kappa(X)$ *và* $\kappa'(Y) \leq \kappa(Y)$.

Giả thiết về κ' trong mệnh đề trên nói rằng quan hệ bằng nhau mức κ' mạnh hơn trên X nhưng yếu hơn trên Y. Vì vậy nếu hai bộ bằng nhau với mức κ' trên X thì chúng cũng bằng nhau mức κ trên X. Lập luận tương tự, ta suy ra chúng cũng phải bằng nhau mức κ' trên Y. Trường hợp ngược lại, chẳng hạn, κ' thỏa $\kappa'(Y) > \kappa(Y)$, ta có thể chỉ ra ví dụ mà κ-PTHM $f = X \rightarrow_\kappa Y$ thỏa quan hệ r nhưng κ'-PTHM $g = X \rightarrow_{\kappa'} Y$ lại không thỏa quan hệ r đó. Nghĩa là, với κ' thỏa $\kappa'(Y) > \kappa(Y)$, κ-PTHM *f* không suy dẫn ra κ'-PTHM *g* hay, nói khác đi, κ'-PTHM *g* có ngữ nghĩa ràng buộc dữ liệu của riêng nó. Như vậy, khác với trường hợp PTH kinh điển, giữa hai tập thuộc tính X và Y chỉ tồn tại tối đa một phụ thuộc hàm, ở đây giữa chúng có thể có 2 PTHM khác nhau, κ-PTHM *f* và κ'-PTHM *g*.

Như vậy, mỗi κ-PTHM, đóng vai trò như một ràng buộc và sở hữu một ngữ nghĩa xác định của riêng nó. Điều này rất quan trọng vì nó sẽ là đặc trưng phải được tận dụng để xây dựng tập các quy tắc suy dẫn cho các PTHM. Như vậy, kỹ thuật chứng minh sử dụng quan hệ gồm 2 bộ chỉ bao gồm các phần tử **0** và **1** được sử dụng trong [5], ở đây **0** và **1** kí hiệu phần tử nhỏ nhất và lớn nhất trong miền thuộc tính tương ứng, không đủ để chứng minh tính đầy đủ của các quy tắc suy dẫn cho các PTHM, vì quan hệ gồm 2 bộ như vậy cho thấy không mô tả được sự khác nhau về ngữ nghĩa giữa $X \rightarrow Y$, $X \rightarrow_\kappa Y$ và $X \rightarrow_{\kappa'} Y$ như đã thảo luận ở trên.

4. HỆ CÁC QUY TẮC SUY DẪN ĐÚNG VÀ ĐẦY ĐỦ CHO CÁC PTHM

Hãy xét một tập các PTHM trên tập vũ trụ các thuộc tính U, $\mathbf{F} = \{X_i \rightarrow_{\kappa i} Y_i: i = 1, 2, \ldots, n\}$, trong đó $X_i, Y_i \subseteq U$. Lưu ý rằng theo định nghĩa PTHM, với mỗi κ_i-PTHM $X_i \rightarrow_{\kappa i} Y_i$, κ_i được ngầm định là xác định trên $X_iY_i = X_i \cup Y_i$. Tiếp theo chúng ta sẽ định nghĩa các khái niệm về hệ quả ngữ nghĩa và hệ quả cú pháp của tập **F**. Đầu tiên, chúng ta định nghĩa một số kí hiệu.

Đặt $\kappa = \{k(A): A \in X\}$ và $\kappa' = \{k'(B): B \in Y\}$ là hai mức tương tự cho trước. Nếu $\kappa(\mathbf{Z}) = \kappa'(\mathbf{Z})$, với $\mathbf{Z} \subseteq \mathbf{X} \cap \mathbf{Y}$, thì κ và κ' được gọi là tương thích trên $\mathbf{Z}$. Với hai mức tương tự bất kì κ và κ', thì $\kappa \vee \kappa'$ là mức tương tự $\{k(A): A \in X \setminus Y\} \cup \{k'(B): B \in Y \setminus X\} \cup \{k(A) \vee k'(A): A \in X \cap Y\}$, ở đây $k(A) \vee k'(A) = \max\{ k(A), k'(A) \}$. Nếu κ và κ' là tương thích trên $\mathbf{X} \cap \mathbf{Y}$ thì $\kappa \vee \kappa'$ là một mở rộng của κ và κ' trên tập $\mathbf{XY}$ và, trong trường hợp này, chúng ta có thể viết $\kappa \cup \kappa' = \kappa \vee \kappa'$.

Một PTHM $f = \mathbf{X} \rightarrow_{\kappa} \mathbf{Y}$ được gọi là *hệ quả ngữ nghĩa* của tập **F** các PTHM nếu với mọi r trên *R*, nếu r thỏa tất cả PTHM trong **F**, thì r cũng thỏa *f*. Khi đó, chúng ta viết $\mathbf{F} \models f$.

Kí hiệu $\mathbf{F}^* = \{f = \mathbf{X} \rightarrow_{\kappa} \mathbf{Y}: \mathbf{F} \models f\}$, nghĩa là $\mathbf{F}^*$ là họ tất cả các PTHM với những mức tương tự κ khác nhau trên $\mathbf{U}$ mà là hệ quả ngữ nghiã của **F**. Để định nghĩa khái niệm *hệ quả cú pháp* của **F**, đầu tiên chúng ta đưa ra các quy tắc suy diễn cú pháp sau:

(κ1) *Phản xạ*: $\mathbf{X} \rightarrow_{\kappa} \mathbf{X}'$, với $\mathbf{X}' \subseteq \mathbf{X}$;

(κ2) *Làm chặt*[*]: $\mathbf{X} \rightarrow_{\kappa} \mathbf{Y} \Rightarrow \mathbf{X} \rightarrow_{\kappa^*} \mathbf{Y}$, với mọi κ^* thỏa $\kappa^*_X \geq \kappa_X$ và $\kappa^*_Y \leq \kappa_Y$;

(κ3) *Tăng trưởng*: $\mathbf{X} \rightarrow_{\kappa} \mathbf{Y} \Rightarrow \mathbf{XZ} \rightarrow_{\kappa \vee \kappa^*(Z)} \mathbf{YZ}$, với bất kì $\mathbf{Z} \subseteq \mathbf{U}$ và bởi bất kì mức tương tự κ^* trên Z sao cho $\kappa^*_{Y \cap Z} \leq \kappa_{Y \cap Z}$, ở đây $\mathbf{XZ} = \mathbf{X} \cup \mathbf{Z}$ và $\mathbf{YZ} = \mathbf{Y} \cup \mathbf{Z}$;

(κ4) *Bắc cầu*: $\mathbf{X} \rightarrow_{\kappa} \mathbf{Y}, \mathbf{Y} \rightarrow_{\kappa^*} \mathbf{Z} \Rightarrow \mathbf{X} \rightarrow_{\kappa \vee \kappa^*} \mathbf{Z}$, ở đây $\kappa^*_Y \leq \kappa_Y$.

Như chúng ta đã biết, một dẫn suất (cú pháp) từ **F** là một dãy các PTHM $f_1, \ldots, f_m$ sao cho với mỗi i, $1 \leq i \leq m$, f_i hoặc là một PTHM của **F** hoặc là một PTHM được suy diễn từ một hoặc hai PTHM trong dãy $f_1, \ldots, f_{i-1}$, bằng việc áp dụng một trong số các quy tắc suy diễn (κ1) – (κ4). Một PTHM f được gọi là *suy dẫn được* từ **F** nếu tồn tại một dẫn suất $f_1, \ldots, f_m$ từ **F** sao cho $f = f_m$. Khi đó, ta viết $\mathbf{F} \vdash f$ và f được gọi là một hệ quả cú pháp của **F**.

Một quy tắc suy diễn được gọi là *đúng* nếu nó bảo toàn được tính đúng của các PTHM có mặt trong tiền đề của quy tắc, nghĩa là với bất kì quan hệ r, nếu các PTHM ở tiền đề là đúng trong r thì PTHM kết luận cũng sẽ đúng trong r. Định lí sau khẳng định tính đúng đắn của các quy tắc suy diễn:

Định lí 4.1. *Các quy tắc suy diễn* (κ1) – (κ4) *cho các* PTHM *là đúng đắn.*

Chứng minh. Chứng minh (κ1) và (κ3) rất đơn giản. (κ2) là mệnh đề 3.1. Để chứng minh (κ4), giả sử rằng $\mathbf{X} \rightarrow_{\kappa} \mathbf{Y}, \mathbf{Y} \rightarrow_{\kappa'} \mathbf{Z}$, với $\kappa'_Y \leq \kappa_Y$, là đúng trong r và $s, t \in$ r là những bộ bất kì sao cho $s[\mathbf{X}] =_{\kappa \vee \kappa'} t[\mathbf{X}]$. Vì $s[A] =_{\kappa(A) \vee \kappa'(A)} t[A]$ suy ra $s[A] =_{\kappa(A)} t[A]$, bởi định nghĩa của $\kappa \vee \kappa'$, $\forall A \in \mathbf{X}$, tức là ta có $s[\mathbf{X}] =_{\kappa} t[\mathbf{X}]$. Vậy, $s[\mathbf{Y}] =_{\kappa} t[\mathbf{Y}]$. Điều này dẫn đến $s[\mathbf{Y}] =_{\kappa'} t[\mathbf{Y}]$, do $\kappa'_Y \leq \kappa_Y$. Vì $\mathbf{Y} \rightarrow_{\kappa'} \mathbf{Z}$ đúng trong r, nên ta lại thu được $s[\mathbf{Z}] =_{\kappa'} t[\mathbf{Z}]$. Kết hợp các đẳng thức mờ đã thu được, ta thấy $s[\mathbf{Y} \cap \mathbf{Z}] =_{\kappa} t[\mathbf{Y} \cap \mathbf{Z}]$ và $s[\mathbf{Z} \setminus \mathbf{XY}] =_{\kappa'} t[\mathbf{Z} \setminus \mathbf{XY}]$. Vì $\kappa'_Y \leq \kappa_Y$, cuối cùng ta có $\mathbf{X} \rightarrow_{\kappa \vee \kappa'} \mathbf{Z}$. ■

Với một tập PTHM **F** cho trước. Kí hiệu $\mathbf{F}^+$ là họ nhỏ nhất các PTHM chứa **F** và đóng đối với các quy tắc suy dẫn (κ1) – (κ4). Hoặc hiểu theo cách khác, $\mathbf{F}^+$ là tập nhỏ nhất các PTHM có thể được suy dẫn từ **F**. Theo định lí 4.1, chúng ta có $\mathbf{F}^+ \subseteq \mathbf{F}^*$.

Tương tự như trong trường hợp PTH, chúng ta cũng có khái niệm phủ: cho một tập PTHM **F**, tập PTHM **G** được gọi là một phủ của **F** nếu $\mathbf{G}^+ \supseteq \mathbf{F}^+$. Rõ ràng, **F** là một phủ của chính nó. **F** và **G** được gọi là tương đương nếu $\mathbf{G}^+ = \mathbf{F}^+$.

---

[*] Thuật ngữ này lấy từ ý nghĩa của điều kiện $\kappa^*_X \geq \kappa_X$, nghĩa là điều kiện bằng nhau mức κ^*_X chặt hơn điều kiện bằng nhau mức κ_X.

Định lí 4.2. *Cho tập* **F** *các PTHM, tập* $\mathbf{F}^+$ *có những tính chất sau*:

(i) $\boldsymbol{X} \rightarrow_{\kappa} \boldsymbol{Y} \in \mathbf{F}^+ \Rightarrow \boldsymbol{X} \rightarrow_{\kappa} \boldsymbol{Y'} \in \mathbf{F}^+$, *với mọi* $\boldsymbol{Y'} \subseteq \boldsymbol{Y}$;

(ii) $\boldsymbol{X} \rightarrow_{\kappa} \boldsymbol{Y} \in \mathbf{F}^+$ & $\boldsymbol{X} \rightarrow_{\kappa'} \boldsymbol{Z} \in \mathbf{F}^+ \Rightarrow \boldsymbol{X} \rightarrow_{\kappa\vee\kappa'} \boldsymbol{YZ} \in \mathbf{F}^+$, *ở đây* $\boldsymbol{\kappa}$ *và* $\boldsymbol{\kappa'}$ *là tương thích trên* $\boldsymbol{X}$;

(iii) $\boldsymbol{X} \rightarrow_{\kappa} \boldsymbol{Y} \in \mathbf{F}^+$ & $\boldsymbol{V} \rightarrow_{\kappa'} \boldsymbol{W} \in \mathbf{F}^+ \Rightarrow \boldsymbol{XV} \rightarrow_{\kappa\vee\kappa'} \boldsymbol{YW} \in \mathbf{F}^+$, *với* $\boldsymbol{\kappa'}_{Y\cap V} \leq \boldsymbol{\kappa}_{Y\cap V}$ và $\boldsymbol{\kappa'}_{Y\cap W} \geq \boldsymbol{\kappa}_{Y\cap W}$;

(iv) *Tập* $\mathbf{G} = \{\boldsymbol{X} \rightarrow_{\kappa} A : \boldsymbol{X} \rightarrow_{\kappa} \boldsymbol{Y} \in \mathbf{F}\ \&\ A \in \boldsymbol{Y}\}$, *thì*, $\mathbf{G}^+ = \mathbf{F}^+$.

Chứng minh. Dễ dàng chứng mình rằng (i) là đúng nhờ áp dụng (κ1) và (κ4). Còn để chứng minh (ii), đặt $\boldsymbol{Y}_0 = \{B \in \boldsymbol{Y} \cap \boldsymbol{Z} : \boldsymbol{\kappa'}_B \leq \boldsymbol{\kappa}_B\}$ và $\boldsymbol{Y'} = (\boldsymbol{Y}\backslash\boldsymbol{Z})\boldsymbol{Y}_0$. Do (i) chúng ta có $\boldsymbol{X} \rightarrow_{\kappa} \boldsymbol{Y'}$. Áp dụng (κ3) với $\boldsymbol{\kappa}^*_X = \boldsymbol{\kappa}_X$ ta thu được PTHM $\boldsymbol{X} \rightarrow_{\kappa} \boldsymbol{Y'X} \in \mathbf{F}^+$, vì khi đó $\boldsymbol{\kappa'}_{Y'\cap Z} \leq \boldsymbol{\kappa}_{Y'\cap Z}$. Áp dụng (κ3) cho PTHM thứ hai của (ii), chúng ta có $\boldsymbol{Y'X} \rightarrow_{\underline{\kappa'}} \boldsymbol{Y'Z} \in \mathbf{F}^+$, trong đó $\underline{\boldsymbol{\kappa'}} = \boldsymbol{\kappa'} \vee \boldsymbol{\kappa}_{Y'}$ hoặc nó được xác định bởi $\underline{\boldsymbol{\kappa'}}_{XZ} = \boldsymbol{\kappa'}$ và $\underline{\boldsymbol{\kappa'}}_{Y'\backslash XZ} = \boldsymbol{\kappa}_{Y'\backslash XZ}$, do $\boldsymbol{\kappa'}_X = \boldsymbol{\kappa}_X$ theo giả thiết trong (ii). Rõ ràng, $\underline{\boldsymbol{\kappa'}}_{Y'X} \leq \boldsymbol{\kappa}_{Y'X}$, nên theo (κ4) với $\boldsymbol{\kappa} \vee \underline{\boldsymbol{\kappa'}} = \boldsymbol{\kappa} \vee \boldsymbol{\kappa'}$, ta suy ra $\boldsymbol{X} \rightarrow_{\kappa\vee\kappa'} \boldsymbol{YZ} \in \mathbf{F}^+$.

Bây giờ chúng ta sẽ chứng minh (iii). Áp dụng (κ3) cho PTHM thứ nhất của (iii), với điều kiện đã cho là $\boldsymbol{\kappa'}_{Y\cap V} \leq \boldsymbol{\kappa}_{Y\cap V}$ và $\boldsymbol{\kappa'}_{Y\cap W} \geq \boldsymbol{\kappa}_{Y\cap W}$, chúng ta thu được $\boldsymbol{XV} \rightarrow_{\kappa\vee\kappa'(V)} \boldsymbol{YV} \in \mathbf{F}^+$ và $\boldsymbol{YV} \rightarrow_{\kappa(Y)\vee\kappa'} \boldsymbol{YW} \in \mathbf{F}^+$. Vì $\boldsymbol{\kappa'}_{Y\cap V} \leq \boldsymbol{\kappa}_{Y\cap V}$ và $\boldsymbol{\kappa'}_{Y\cap W} \geq \boldsymbol{\kappa}_{Y\cap W}$, nên ta có $(\boldsymbol{\kappa}_Y \vee \boldsymbol{\kappa'})_{YV} = (\boldsymbol{\kappa} \vee \boldsymbol{\kappa'}_V)_{YV}$ là đúng. Vì vậy áp dụng (κ4) ta suy ra $\boldsymbol{XV} \rightarrow_{\boldsymbol{\kappa}\vee\boldsymbol{\kappa'}} \boldsymbol{YW} \in \mathbf{F}^+$, do $(\boldsymbol{\kappa} \vee \boldsymbol{\kappa'}_V) \vee (\boldsymbol{\kappa}_Y \vee \boldsymbol{\kappa'}) = \boldsymbol{\kappa} \vee \boldsymbol{\kappa'}$.

Rõ ràng mệnh đề (iv) là hệ quả của (i) và (ii). Từ (i) ta có $\mathbf{G} \subseteq \mathbf{F}^+$ và như vậy $\mathbf{G}^+ \subseteq \mathbf{F}^+$. Ngược lại, với mỗi $\boldsymbol{X} \rightarrow_{\kappa} \boldsymbol{Y} \in \mathbf{F}$, từ $\boldsymbol{X} \rightarrow_{\kappa} A \in \mathbf{G}$, với mọi $A \in \boldsymbol{Y}$, và (ii) kéo theo $\boldsymbol{X} \rightarrow_{\kappa} \boldsymbol{Y} \in \mathbf{G}^+$, nghĩa là chúng ta có $\mathbf{G}^+ \supseteq \mathbf{F}$. Suy ra $\mathbf{G}^+ \supseteq \mathbf{F}^+$. ■

Định lí 4.3. *Tập các quy tắc suy dẫn* (κ1) – (κ4) *trong định lí 3.1 là đầy đủ, nghĩa* $\mathbf{F}^+ \supseteq \mathbf{F}^*$, *và vì vậy chúng ta có* $\mathbf{F}^+ = \mathbf{F}^*$.

Chứng minh. Đầu tiên chúng ta có nhận xét: Dựa vào (κ2) nên dễ dàng thấy rằng với bất kì PTHM $\boldsymbol{Z} \rightarrow_{\boldsymbol{\kappa}} A \notin \mathbf{F}^+$ chúng ta có $\boldsymbol{Z} \rightarrow_{\boldsymbol{\kappa'}} A \notin \mathbf{F}^+$ với mọi $\boldsymbol{\kappa'}$ thỏa mãn $\boldsymbol{\kappa'}_Z = \boldsymbol{\kappa}_Z$ và $\boldsymbol{\kappa'}_A > \boldsymbol{\kappa}_A$. Thật vậy, vì ngược lại nếu ta có $\boldsymbol{Z} \rightarrow_{\boldsymbol{\kappa'}} A \in \mathbf{F}^+$ thì theo (κ2) chúng ta có $\boldsymbol{Z} \rightarrow_{\boldsymbol{\kappa}} A \in \mathbf{F}^+$ với mọi $\boldsymbol{\kappa}$ thỏa $\boldsymbol{\kappa'}_Z = \boldsymbol{\kappa}_Z$ và $\boldsymbol{\kappa'}_A \geq \boldsymbol{\kappa}_A$.

Theo định lí 4.1 ta đã có $\mathbf{F}^+ \subseteq \mathbf{F}^*$, nên ta chỉ cần chỉ ra bao hàm thức ngược lại. Dựa vào khẳng đinh (ii) của Định lí 4.2, ta chỉ cần chứng minh rằng với mỗi PTHM $f = \boldsymbol{X} \rightarrow_{\kappa} A \notin \mathbf{F}^+$ (với sự bằng nhau mức $\boldsymbol{\kappa}$ trên $\boldsymbol{X} \cup \{A\}$), thì:

$\exists$r thỏa mãn tất cả các PTHM trong $\mathbf{F}^+$ nhưng không thỏa mãn $f = \boldsymbol{X} \rightarrow_{\kappa} A$ (*)

Đặt $\boldsymbol{X}_{\boldsymbol{\kappa}}^+ = \{B \in \boldsymbol{U}: \exists g = \boldsymbol{X} \rightarrow_{\boldsymbol{\kappa'}} B \in \mathbf{F}^+$, với mọi $\boldsymbol{\kappa'}$ thỏa $\boldsymbol{\kappa'}_X = \boldsymbol{\kappa}_X\}$. Với mỗi $B \in \boldsymbol{X}_{\boldsymbol{\kappa}}^+$, đặt $\boldsymbol{k}^*_B = \max\{\boldsymbol{\kappa'}_B: \exists\boldsymbol{\kappa'}$ sao cho $\boldsymbol{X} \rightarrow_{\boldsymbol{\kappa'}} B \in \mathbf{F}^+$ và $\boldsymbol{\kappa'}_X = \boldsymbol{\kappa}_X\}$. Rõ ràng, $\boldsymbol{k}^*_B = \boldsymbol{\kappa}_B$, với mọi $B \in \boldsymbol{X}$. Với bất kì $\boldsymbol{Z} \subseteq \boldsymbol{X}_{\boldsymbol{\kappa}}^+$, kí hiệu $\boldsymbol{k}^*_Z$ là tập$\{\boldsymbol{k}^*_B: B \in \boldsymbol{Z}\}$.

Bảng 3. Quan hệ 2 bộ với $b_i =_{\boldsymbol{k}^*(B_i)} 1$ nhưng $b_i \neq_{\boldsymbol{k}^*(B_i)+1} 1$, $i = 1, \ldots, n\!-\!1$, và $a =_{\boldsymbol{k}^*(A)-1} 1$ nhưng $a \neq_{\boldsymbol{k}^*(A)} 1$. Rõ ràng là $0 \neq_{\kappa(A)} 1$, với $\forall\kappa$.

Bộ	$\boldsymbol{X}_{\kappa}^+$		$\boldsymbol{U} \setminus \boldsymbol{X}_{\kappa}^+$
	$\boldsymbol{X}$	$B_{m+1} \ldots B_{n-1}\ A$	
t_1	1 1… … 1	1 … … 1 1	1 … … 1
t_2	$b_1\ b_2 \ldots \ldots b_m$	$b_{m+1} \ldots \ldots b_{n-1}\ a$	0 … … 0

Bây giờ quan hệ r thỏa mãn đòi hỏi trên sẽ được xây dựng theo từng trường hợp như sau:

Với trường hợp $A \notin \boldsymbol{X}_{\boldsymbol{\kappa}}^{+}$, quan hệ r gồm 2 bộ với các phần tử 0 và 1 sẽ được xây dựng tương tự như trong trường hợp PTH kinh điển sao cho nó thỏa mãn điều kiện (*) ở trên với các biểu thức trong $\mathbf{F}^{+}$ và f được xem là các PTH thông thường. Nhưng vì chỉ có hai giá trị kinh điển 0 và 1 trong quan hệ như vậy nên với hai giá trị bất kì a và b, nếu $a = b$ ta suy ra $a =_{\boldsymbol{\kappa}} b$ và nếu $a \neq b$ ta cũng suy ra được $a \neq_{\boldsymbol{\kappa}} b$, với mọi $\boldsymbol{\kappa}$. Ở đây, 0 và 1 kí hiệu tương ứng là phần tử nhỏ nhất và lớn nhất trong miền trị của thuộc tính tương ứng, nên ta luôn luôn có $0 \neq_{\boldsymbol{\kappa}} 1$, với mọi $\boldsymbol{\kappa}$. Do vậy ta suy ra các PTHM trong $\mathbf{F}^{+}$ cũng thỏa trong r nhưng f thì không, nghĩa là điều kiện (*) được thỏa mãn.

Với trường hợp $A \in \boldsymbol{X}_{\boldsymbol{\kappa}}^{+}$, quan hệ r sẽ được xây dựng như sau:

Giả thiết rằng các thuộc tính trong $\boldsymbol{X}_{\boldsymbol{\kappa}}^{+}$ được sắp xếp sao cho $\boldsymbol{X}_{\boldsymbol{\kappa}}^{+} = \{B_i : i = 1, 2, \ldots, n\}$, $B_n = A$ và $\boldsymbol{X} = \{B_i : i = 1, 2, \ldots, m\}$, $m < n$. Với mỗi thuộc tính B_i, $i = 1, 2, \ldots, n\text{-}1$, nhận một giá trị $b_i \in \boldsymbol{D}_{Bi} = [0,1]$ sao cho $b_i =_{\boldsymbol{k}^*(B_i)} 1$ nhưng $b_i \neq_{\boldsymbol{k}^*(B_i)+1} 1$ và tồn tại một giá trị $a \in \boldsymbol{D}_A = [0,1]$ sao cho $a =_{\boldsymbol{k}^*(A)} 1$ nhưng $a \neq_{\boldsymbol{k}^*(A)} 1$. Những giá trị như vậy tồn tại với mỗi $k > 1$, do mỗi lớp tương đương của quan hệ bằng nhau mức $k - 1$, $=_{k-1}$, luôn luôn bao gồm ít nhất hai lớp tương tương mịn hơn của quan hệ bằng nhau $=_k$. Vì vậy, ta luôn luôn có thể xây dựng được quan hệ r như trong bảng 2.

Chúng ta sẽ chỉ ra rằng mọi PTHM $g = \boldsymbol{Z} \to_{\kappa'} B \in \mathbf{F}^{+}$ luôn thỏa mãn trong r. Thật vậy, do trường hợp $B \in \boldsymbol{Z} \subseteq \boldsymbol{X}_{\boldsymbol{\kappa}}^{+}$, PTHM g là tầm thường, nghĩa là nó luôn luôn thỏa mãn trong mọi quan hệ, nên ta xét trường hợp $B \notin \boldsymbol{Z}$. Với $\boldsymbol{Z} \not\subseteq \boldsymbol{X}_{\kappa}^{+}$, rõ ràng g được thỏa mãn bởi r, bởi $t_1[Z] \neq_{\kappa'} t_2[Z]$. Với $\boldsymbol{Z} \subseteq \boldsymbol{X}_{\kappa}^{+}$, theo định nghĩa của X_{κ}^{+} ta suy ra $B \in X_{\kappa}^{+}$. Xét trường hợp $\boldsymbol{\kappa}'_Z \not\leq \boldsymbol{k}^*_Z$. Khi đó tồn tại $B_j \in \boldsymbol{Z}$ sao cho $\boldsymbol{\kappa}'_{Bj} > \boldsymbol{k}^*_{Bj}$. Dựa theo cách xây dựng quan hệ r, ta có $b_j \neq_{\boldsymbol{k}^*(B_j)+1} 1$, nghĩa là ta có $t_1[B_j] \neq_{\kappa'(Bj)} t_2[B_j]$, nên g cũng thỏa r.

Bây giờ ta giả thiết $\boldsymbol{\kappa}'_Z \leq \boldsymbol{k}^*_Z$. Nếu $B \neq A$ thì rõ ràng g được thỏa mãn bởi r, bởi chúng ta luôn luôn có $t_1[B] =_{\kappa'(B)} t_2[B]$, do cách xây dựng r. Xét trường hợp $B = A$, thức là $g = \boldsymbol{Z} \to_{\kappa'} A$. Do $\boldsymbol{Z} \subseteq \boldsymbol{X}_{\kappa}^{+}$, từ định nghĩa của $\boldsymbol{X}_{\kappa}^{+}$ và $\boldsymbol{k}^*$ và áp dụng (ii), Định lí 4.2, từ điều kiện $\boldsymbol{X} \to_{\boldsymbol{\kappa}'} B' \in \mathbf{F}^{+}$, $B' \in Z$, chúng ta thu được $\boldsymbol{X} \to_{\kappa''} \boldsymbol{Z} \in \mathbf{F}^{+}$, với mức tương tự $\boldsymbol{\kappa}''$ được định nghĩa trên $\boldsymbol{XZ}$ thỏa mãn điều kiện $\boldsymbol{\kappa}''_X = \boldsymbol{\kappa}_X$ và $\boldsymbol{\kappa}''_{B'} = \boldsymbol{k}^*_{B'}$, $\forall B' \in \boldsymbol{Z}$. Lại theo định nghĩa của $\boldsymbol{k}^*$, từ điều kiện $\boldsymbol{Z} \subseteq \boldsymbol{X}_{\kappa}^{+}$ suy ra $\boldsymbol{\kappa}'_Z \leq \boldsymbol{\kappa}''_Z$. Sử dụng quy tắc (κ4), từ $\boldsymbol{X} \to_{\kappa''} \boldsymbol{Z} \in \mathbf{F}^{+}$ và $g = \boldsymbol{Z} \to_{\kappa'} A \in \mathbf{F}^{+}$ chúng ta suy ra $\boldsymbol{X} \to_{\kappa'' \vee \kappa'} A \in \mathbf{F}^{+}$. Do vậy, theo nhận xét ở phần đầu của chứng minh, từ giả thiết đã cho là $f = \boldsymbol{X} \to_{\kappa} A \notin \mathbf{F}^{+}$ ta suy ra $\boldsymbol{\kappa}_A$ không thể thỏa mãn $\boldsymbol{\kappa}_A \leq \boldsymbol{\kappa}''_A \vee \boldsymbol{\kappa}'_A$ và, vì vậy, ta có $\boldsymbol{\kappa}_A > \boldsymbol{\kappa}''_A \vee \boldsymbol{\kappa}'_A = \boldsymbol{\kappa}'_A$, nghĩa là đẳng thức mức $\boldsymbol{\kappa}_A$ mịn hơn mức $\boldsymbol{\kappa}'_A$. Do cách xây dựng r, chúng ta có $t_1[A] =_{\kappa'(A)} t_2[A]$, nhưng $t_1[A] \neq_{\kappa(A)} t_2[A]$. Điều này chứng tỏ g thỏa mãn quan hệ r, nhưng f lại không thỏa r. ■

5. KẾT LUẬN

Đối với CSDL với thông tin không chắc chắn [3, 16, 4 – 6, 8 - 9], dường như vấn đề cốt lõi là cần phải xây dựng được mô hình toán học hợp lí biểu diễn được ngữ nghĩa của những khái niệm mờ đúng bản chất tự nhiên vốn có của nó, đồng thời cách biểu diễn những khái niệm này trong CSDL cần phải đơn giản và hiệu quả. Nghiên cứu ý nghĩa của các hạng từ ngôn ngữ trên miền trị của các biến ngôn ngữ hoặc thuộc tính ngôn ngữ, ĐSGT có thể được xem như mô hình toán học biểu diễn được đầy đủ cấu trúc ngữ nghĩa của các miền hạng từ ngôn ngữ. Công cụ định lượng trong ĐSGT cho phép có thể biểu diễn và thao tác với các khái niệm mờ trong CSDL ngôn ngữ một cách đơn giản và thuận lợi [2]. Các phép toán so sánh được định nghĩa dựa trên khái niệm k-bằng, ở đây k là độ dài của biểu diễn chính tắc của những hạng từ, nó chỉ rõ mức tương tự trong quan hệ bằng này. Tính không chắc chắn hoặc tính mờ của đẳng thức không chắc chắn dựa trên các khoảng mờ mức k của các hạng từ mờ được sử dụng để định nghĩa những đẳng thức này. Tuy nhiên, chúng là những quan hệ bằng nhau rõ trên miền tham chiếu của các thuộc tính được xem xét và bởi vậy chúng rất giống với quan hệ bằng thông thường về mặt hình thức.

Với ngữ nghĩa của dữ liệu ngôn ngữ như vậy, bài báo đã giới thiệu và đã trình bày một kiểu phụ thuộc hàm mờ mới. Mỗi PTHM được kết hợp với một mức tương tự κ xác định trên những thuộc tính liên quan, chúng chỉ rõ mức không chắc chắn của PTHM. PTH và PTHM với cùng vế trái và vế phải đã được thảo luận và chỉ ra rằng PTH có thể được xem nhưng những ràng buộc cú pháp của CSDL ngôn ngữ, trong khi đó PTHM được xem như những ràng buộc ngữ nghĩa. Với một họ các PTHM như vậy, một tập các quy tắc suy dẫn đã được thiết lập và đã được chứng minh là đúng và đầy đủ.

Lời cảm ơn. Nhóm tác giả cám ơn Quỹ Phát triển Khoa học và Công nghệ Quốc gia (Nofosted) thông qua đề tài mã số 102.01-2011.06 đã hỗ trợ chúng tôi thực hiện bài báo này.

TÀI LIỆU THAM KHẢO

1. Ho N. C. - A model of relational databases with linguistic data of hedge algebras - based semantics, Proceed. of ICT.rda’2006, 20-21/05/2006.
2. Nguyễn Cát Hồ, Lê Xuân Vinh, Nguyễn Công Hào - Thống nhất dữ liệu và xây dựng quan hệ tương tự trong CSDL ngôn ngữ bằng ĐSGT, Tạp chí Tin học và Điều khiển học **25** (4) (2009) 314-332.
3. Bhattacharjee T. K. and Mazumdar A. K. - Axiomatisation of fuzzy multivalued dependencies in fuzzy relational data model, Fuzzy Sets and Systems **96** (1998) 343-352.
4. Jyothi S., Syam Babu M. - Multi-dependencies in fuzzy relational databases and lossless join decomposition, Fuzzy sets and systems **88** (1997) 315-332.
5. Liu W. Y. - Fuzzy data dependencies and implication of fuzzy data dependencies, Fuzzy sets and systems **92** (1997) 341-348.
6. Raju K.V. S. V. N, Majumdar Ar. K. - Fuzzy Functional Dependencies and Lossless Join Decomposition of Fuzzy Relational Database Systems, ACM Transactions on Database Systems **13** (2) (1988) 129-166.
7. Dukíc N., Avdagíc Z. - Formalization of Provenes Fuzzy Functional Dependency in Fuzzy Databases, Mathware & Soft Computing **11** (2004) 31-44.
8. Sadeq Al-Hamouz and Ranjit Biswas - Fuzzy Functional Dependencies in Relational Databases, International Journal of Computational Cognition (http://www.ijcc.us) **4** (1) (2006) 39-43.
9. Shenoi S., Melton A. - Functional dependencies and normal forms in the fuzzy relational databases model, Information Science **60** (1992) 1-28.
10. Ho N.C, 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.

11. Ho N. C, Nam H. V. - 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.
12. Ho N. C, Nam H. V. - 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).
13. Ho N. C, Nam H. V. - Towards an Algebraic Foundation for a Zadeh Fuzzy Logic, Fuzzy Set and System **129** (2002) 229-254.
14. Ho N. C. and Wechler W. - Hedge algebras: An algebraic approach to structures of sets of linguistic domains of linguistic truth variable, Fuzzy Sets and Systems **35** (1990) 281-293.
15. Ho N. C. and Wechler W. - Extended hedge algebras and their application to Fuzzy logic. Fuzzy Sets and Systems **52** (1992) 259-281.
16. Chiang D. A., Chow L. R., and Hsien N. C. - Fuzzy information in extended fuzzy relational databases, Fuzzy Sets and Systems **92** (1997) 1-20.
17. Ho N. C., Long N. V. - Complete and linear hedge algebras, fuzziness measure of vague concepts and linguistic hedges and application, (Best paper Award of the Conference), AIP Conf. Proceed. on Computing Anticipatory Systems, CASYS'05, Liege, Belgium 8-13 August 2005. ed. Daniel M. Dubois, pp. 331-339.
18. Ho N.C., Long N. V. - Fuzziness Measure on Complete Hedge Algebras and Quantitative Semantics of Terms in Linear Hedge Algebras, Fuzzy Sets and Systems **158** (2007) 452-471.
19. Ho N. C., Nam H. V., Khang T. D., and Chau L. H. - Hedge Algebras, Linguistic- valued Logic and their Application to Fuzzy Reasoning, Inter.J. of Uncertainty, Fuzziness and Knowledge-Based System **7** (1999) 347-361.
20. 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), p.105-112.

ABSTRACT

FUZZY FUNCTIONAL DEPENDENCIES IN FUZZY RELATIONAL DATABASES WITH LINGUISTIC DATA

Nguyen Cat Ho[1,*], Vu Minh Loc[2], Hoang Tung[3], Nguyen Tan An[4]

[1] *Institute of Information Technology, VAST, 18 Hoang Quoc Viet, Cau Giay, Hanoi, Vietnam*
[2] *Gia Dinh Information Technology University*
[3] *DongNai University*
[4] *Faculty of Information Technology, Ha Noi National University of Education*

[*] Email: *ncatho@gmail.com*

Relational databases with linguistic data based on hedge algebras – based semantics were introduced and investigated in [1, 2], in which each linguistic datum x are represented in two semantic components, a semantic value belonging to the real domain $\boldsymbol{D}_A$ of x and a set of fuzziness-intervals-based neighbourhoods of *x*. The data comparison operations are defined based on *k*-equalities, where *k* indicates the length of the representation of elements in hedge algebras. On this new viewpoint of linguistic data semantics, in this paper, a new type of fuzzy functional dependencies in these databases will be defined and examined. New fuzzy functional dependencies can be considered as semantic constraints on the databases in fuzzy environment. It is shown that a sound and complete set of inference rule for these fuzzy functional dependencies can be established.

Keywords: fuzzy database, linguistic data semantics, hedge algebras, fuzziness intervals, uncertain equality, fuzzy functional dependency.

Ý kiến Tổng biên tập:

1. Đề nghị tác giả rà soát lại phiên bản 'proof'.
2. Đề nghị kiểm tra việc trích dẫn: các tài liệu từ 17-20 chua thấy trích dẫn trong bài viết. Nếu đánh số lại thì cần chỉnh lại số tài liệu trích dẫn cho khớp.