*Tạp chí Khoa học và Công nghệ **53** (5) (2015) 568-582*

PHỤ THUỘC DỮ LIỆU TRONG MÔ HÌNH CƠ SỞ DỮ LIỆU HƯỚNG ĐỐI TƯỢNG MỜ SỬ DỤNG ĐẠI SỐ GIA TỬ

Nguyễn Công Hào[1, *], **Trương Thị Mỹ Lê**[2]

[1]*Trung tâm Công nghệ thông tin Đại học Huế, số 2 Lê Lợi, TP Huế*

[2]*Trường Đại học Quang Trung, số 130 Trần Hưng Đạo, TP Quy Nhơn*

[*]Email: *nchao@hueuni.edu.vn*

Đến Tòa soạn: 19/11/2014; Chấp nhận đăng: 11/7/2015

TÓM TẮT

Khái niệm quan trọng nhất trong việc thiết kế các lược đồ cơ sở dữ liệu đó là phụ thuộc hàm. Phụ thuộc hàm mô tả mối liên hệ giữa các thuộc tính và là một trong các khái niệm chính được sử dụng trong quá trình chuẩn hóa. Trên cơ sở mô hình cơ sở hướng đối tượng mờ theo cách tiếp cận đại số gia tử được tác giả đề xuất, trong bài báo này, chúng tôi sẽ trình bày nghiên cứu về các dạng phụ thuộc dữ liệu, bao gồm phụ thuộc hàm mờ trong lớp đối tượng, phụ thuộc hàm mờ đầy đủ, phụ thuộc hàm mờ bộ phận, phụ thuộc hàm mờ bắc cầu và phụ thuộc hàm mờ với lượng từ ngôn ngữ. Dựa trên các khái niệm phụ thuộc hàm mờ, bài báo cũng đề xuất các dạng chuẩn đối tượng mờ của lớp trong mô hình.

Từ khóa: cơ sở dữ liệu hướng đối tượng mờ, đại số gia tử, phụ thuộc hàm mờ, các dạng chuẩn lớp đối tượng mờ.

1. ĐẶT VẤN ĐỀ

Phụ thuộc dữ liệu là nền tảng lí thuyết để xác định các dạng chuẩn của lược đồ cơ sở dữ liệu (CSDL), nhằm hạn chế đến mức thấp nhất sự dư thừa dữ liệu, nguyên nhân chính phá vỡ tính toàn vẹn dữ liệu trong các hệ thống CSDL. Khi ngữ nghĩa của CSDL được mở rộng, như cho phép lưu trữ các thông tin không chắc chắn, không đầy đủ (*gọi là thông tin mờ*), thì ngữ nghĩa của các phụ thuộc dữ liệu cũng thay đổi, nghĩa là phải mở rộng các dạng phụ thuộc dữ liệu. Đã có nhiều công trình tập trung nghiên cứu mở rộng các dạng phụ thuộc dữ liệu, đặc biệt là trong mô hình CSDL mờ [1, 2, 3 ,4, 5] và CSDL hướng đối tượng (HĐT) mờ [6, 7].

Một số mô hình CSDL HĐT mờ cơ bản đã được nghiên cứu chủ yếu dựa vào lí thuyết tập mờ [8], lí thuyết khả năng, quan hệ tương tự, đại số gia tử (ĐGST),...Đối với các mô hình đã nghiên cứu,các tác giả xem các giá trị mờ là tập mờ (như *trẻ*, *cao*, ….), khi biểu diễn và đối sánh các giá trị mờ này phải xây dựng các hàm thuộc và hoàn toàn phụ thuộc vào ý kiến chuyên gia. Hơn nữa, khi cần chuyển đổi giá trị mờ này thành giá trị rõ, phải sử dụng phương pháp khử mờ, tuy nhiên, mỗi phương pháp khử mờ sẽ cho một kết quả khác nhau nên dẫn đến sai số.

Đối với mô hình CSDL HĐT mờ được xây dựng sử dụngĐSGT là một hướng tiếp cận mới, giúp xử lí thông tin mờ một cách hiệu quả, đơn giản, trực quan và không cần phương pháp khử mờ. Với cách tiếp cận này, một mô hình mới của CSDL HĐT mờ đã được xây dựng, trong đó ngữ nghĩa ngôn ngữ được lượng hóa bằng các ánh xạ định lượng của ĐSGT. Khi đó, giá trị ngôn ngữ là dữ liệu, không phải là nhãn của các tập mờ biểu diễn ngữ nghĩa của các giá trị ngôn ngữ, và ưu điểm cơ bản của nó cho phép đánh giá ngữ nghĩa của thông tin mờ cũng như dữ liệu kinh điển một cách thống nhất. Dựa trên mô hình theo cách tiếp cận ĐSGT, chúng tôi đề xuất một số dạng phụ thuộc dữ liệu mờ, các dạng chuẩn trong lớp đối tượngmờ và các vấn đề liên quan.

2. MỘT SỐ KIẾN THỨC CƠ SỞ

2.1. Đại số gia tử

ĐSGT là một trong những cách tiếp cận để phát hiện cấu trúc đại số của miền giá trị của các biến ngôn ngữ. Theo quan điểm đại số,mỗi miền giá trị của biến ngôn ngữ X có thể được hiểu như là một đại số A$X = (X, G, H, \leq)$, trong đó *Dom*(X) = X là miền các giá trị ngôn ngữ của biến ngôn ngữ X được sinh ra từ tập các phần tử sinh $G = \{c^-, c^+\}$ bởi sự tác động của các gia tử trong $H=H^- \cup H^+$;*W* là phần tử trung hòa;$\leq$ là một mối quan hệ thứ tự có ngữ nghĩa trên X, nó được cảm sinh từ ý nghĩa định tính tự nhiên của từ ngữ. Cấu trúc thứ tự cảm sinh trực tiếp như vậy chính là điểm khác biệt so với các cách tiếp cận khác.Khi thêm một số phần tử đặc biệt, thì ĐSGT trở thành một đại số trừu tượng $\underline{\boldsymbol{X}}=(\boldsymbol{X}, G, H,\Sigma,\Phi,\leq)$, trong đó Σ, Φ là các toán tử lấy giới hạn của tập các phần tử được sinh ra khi chịu tác động liên tiếp của các gia tử trong *H*. Một cách khác, nếu ký hiệu $H(x)=\{h_1\ldots h_p x/h_1, \ldots h_p\in H\}$ *thì* Φx=infimumH(x) và Σx=supremumH(x). Như vậy, ĐSGT $\underline{\boldsymbol{X}}$ được xây dựng trên nền của một ĐSGT AX = $(X, G, H,\leq)$, *ở đây* $X = H(G)$, bằng cách bổ sung hai toán tửΣ, Φ. Khi đó X = X∪Lim(G) với Lim(G) là tập các phần tử giới hạn: với mọi x$\in$*Lim(G), tồn tại u*$\in$*X sao cho x* = Φu hoặc x = Σu. Các phần tử giới hạn này được bổ sung vào ĐSGT $\underline{\boldsymbol{X}}$ để làm cho các phép tính mới có nghĩa và vì vậy $\underline{\boldsymbol{X}} = (\boldsymbol{X}, G, H,\Sigma, \Phi, \leq)$ được gọi là ĐSGT đầy đủ. Các hàm định lượng ngữa nghĩa (υ), hàm độ đo tính mờ (*fm*), hàm dấu (Sgn) và các tính chất của ĐSGT có thể tham khảo trong các tài liệuliên quan [9].

2.2. Độ tương tự mức k

Khi định nghĩa lân cận mức *k* chúng ta mong muốn các giá trị đại diện như vậy phải là điểm trong của lân cận mức *k*. Vì vậy ta định nghĩa độ tương tự mức *k* như sau: Chúng ta luôn luôn giả thiết rằng mỗi tập H^- và H^+ chứa ít nhất 2 gia tử . Xét X_k là tập tất cả các phần tử độ dài *k*. Dựa trên các khoảng mờ mức *k* và các khoảng mờ mức *k + 1* chúng ta mô tả không hình thức việc xây dựng một phân hoạch của miền [0;1] như sau:Với *k* = 1, các khoảng mờ mức 1 gồm $I(c^-)$ và $I(c^+)$. Các khoảng mờ mức 2 trên khoảng $I(c^-)$ là $I(h_pc^-) = I(h_{p-1}c^-) = \ldots = I(h_2c^-) = I(h_1c^-) = \upsilon_A(c^-) = I(h_{-1}c^-) = I(h_{-2}c^-) = \ldots = I(h_{-q+1}c^-) = I(h_{-q}c^-)$. Khi đó, ta xây dựng phân hoạch về độ tương tự mức 1 gồm các lớp tương đương sau: $S(0) = I(h_pc^-)$,$S(c^-) = I(c^-)\backslash[I(h_{-q}c^-)\cup I(h_pc^-)]$; $S(W) = I(h_{-q}c^-) \cup I(h_{-q}c^+)$; tương tự ta có $S(c^+) = I(c^+)\backslash[I(h_{-q}c^+)\cup I(h_pc^+)]$ và $S(1) = I(h_pc^+)$.

Tương tự, với *k* = 2, ta có thể xây dựng phân hoạch các lớp tương tự mức 2. chẳng hạn trên khoảng mờ mức 2, $I(h_ic^+) = (\upsilon_A(\Phi h_ic^+), \upsilon_A(\sum h_ic^+)]$ với hai khoảng mờ kề là $I(h_{i-1}c^+)$ và $I(h_{i+1}c^+)$ chúng ta sẽ có các lớp tương đương dạng sau: $S(h_ic^+) = I(h_ic^+)\backslash[I(h_ph_ic^+)\cup I(h_{-q}h_ic^+)]$, $S(\Phi h_ic^+) = I(h_{-q}h_{i-1}c^+)\cup I(h_{-q}h_ic^+)$ và $S(\sum h_ic^+) = I(h_ph_{i-1}c^+)\cup I(h_ph_ic^+)$, với I sao cho $-q\leq i\leq p$ và $I \neq 0$. Bằng cách tương tự như vậy ta có thể xây dựng các phân hoạch các lớp tương tự mức *k* bất kì. Tuy nhiên, trong thực tế ứng dụng chúng ta có thể giới hạn các gia tử tác động liên tiếp lên các phần tử nguyên thủy c^- và c^+ là một số nguyên k^* nào đó. Các giá trị kinh điển và các giá trị mờ gọi là có độ tương tự mức *k* nếu các giá trị đại diện của chúng cùng nằm trong một lớp tương tự mức *k*.

Ví dụ 2.1. Ta xét lược đồ quan hệ *U* = { *MASO*, *TENCN*, *SONLV*, *THUNHAP* } với ý nghĩa: Mã số công nhân (*MASO*), Tên công nhân (*TENCN*) là 2 thuộc tính kinh điển, Số ngày làm việc trong tháng (*SONLV*), Thu nhập (*THUNHAP*) là 2 thuộc tính mờ. Trong đó D_{SONLV} = [0; 30] và $D_{THUNHAP}$ = [0; 100]. LD_{SONLV} và $LD_{THUNHAP}$ có cùng tập các xâu giống nhau với tập các phần tử sinh là {***0***, *thấp*, ***W***, *cao*, ***1***} và tập các gia tử là {*ít*, *khả năng*, *hơn*, *rất*}. Mặc dù các thuộc tính ngôn ngữ đang xét có cùng tập các xâu, nhưng ngữ nghĩa định lượng của chúng khác nhau.

Đối với thuộc tính *SONLV*: *fm*(*cao*) = 0,35; *fm*(*thấp*) = 0,65; μ(*khả năng*) = 0,25; μ(*ít*) = 0,20; μ(*hơn*) = 0,15 và μ(*rất*) = 0,40. Ta phân hoạch đoạn [0; 30] thành 5 khoảng tương tự mức 1 là: *fm*(*rất cao*) × 30 = 0,35 × 0,35 × 30 = 3,675. Vậy *S*(***1***) × 30 = (26,325; 30];

(*fm*(*khả năng cao*) + *fm*(*hơn cao*)) × 30 = (0,25 × 0,35 + 0,15 × 0,35) × 30 = 4,2 và *S*(*cao*) × 30 = (22,125; 26,325]; (*fm*(*ít thấp*) + *fm*(*ít cao*)) × 30 = (0,25 × 0,65 + 0,25 × 0,35) × 30 = 7,5 và *S*(***W***) × 30 = (14,625; 22,125]; (*fm*(*khả năng thấp*) + *fm*(*hơn thấp*)) × 30 = (0,25 × 0,65 + 0,15 × 0,65) × 30 = 7,8 và *S*(*thấp*) × 30 = (6,825; 14,625]; *S*(***0***) × 30 = [0; 6,825].

Đối với thuộc tính*THUNHAP*: *fm*(*cao*) = 0,6; *fm*(*thấp*) = 0,4; μ(*khả năng*) = 0,15; μ(*ít*) = 0,25; μ(*hơn*) = 0,25 và μ(*rất*) = 0,35. Ta phân hoạch đoạn [0; 100] thành 5 khoảng tương tự mức 1 là: *fm*(*rất cao*) × 100 = 0,35 × 0,6 × 100 = 21. Vậy *S*(***1***) × 100 = (79; 100];

(*fm*(*khả năng cao*) + *fm*(*hơn cao*)) × 100 = (0,25 × 0,6 + 0,15 × 0,6) × 100 = 24 và *S*(*cao*)) × 100 = (55; 79]; (*fm*(*ít thấp*) + *fm*(*ít cao*)) × 100 = (0,25 × 0,6 + 0,25 × 0,4) × 100 = 25 và *S*(***W***) × 100 = (30; 55]; (*fm*(*khả năng thấp*) + *fm*(*hơn thấp*)) × 100 = (0,25 × 0,4 + 0,15 × 0,4) × 100 = 16 và *S*(*thấp*) × 100 = (14; 30]; *S*(***0***) × 100 = [0; 14].

3. CƠ SỞ DỮ LIỆU HƯỚNG ĐỐI TƯỢNG MỜ SỬ DỤNG ĐẠI SỐ GIA TỬ

3.1. Đối tượng mờ

Các thực thể trong thế giới thực hay các khái niệm trừu tượng thường là các đối tượng phức tạp. Các đối tượng này chứa một tập nhất định các thông tin về đối tượng và các hành vi dựa trên các thông tin đó. Thông tin về đối tượng được gọi là thuộc tính đối tượng và được xác định bởi giá trị cụ thể, giá trị này có thể là giá trị rõ (giá trị chính xác) hoặc vì một lí do nào đó mà ta không xác định được giá trị chính xác của nó. Ví dụ, giá trịthuộc tính "tuổi" của một đối tượng được cho là "*khoảng 18*", hoặc có thể là một giá trị ngôn ngữ "*rất trẻ*", đây là những thông tin thông tin mờ. Về mặt hình thức, một đối tượng có ít nhất một thuộc tính cho phép nhận giá trị mờ là đối tượng mờ.

3.2. Lớp mờ

Các đối tượng có các thuộc tính giống nhau được đưa vào các lớp được tổ chức thành hệ thống phân cấp. Về mặt lí thuyết, một lớp có thể được xem xét từ hai quan điểm khác nhau: (a) lớp mở rộng, được định nghĩa bởi danh sách các đối tượng của nó, và (b) lớp khái niệm, được xác định bởi một tập các thuộc tính và các giá trị chấp nhận được của nó. Ngoài ra, một lớp con được xác định từ lớp cha bằng cách thừa kế trong CSDL HĐT có thể được xem như là trường hợp đặc biệt (b) ở trên.

Vì vậy, một lớp được xem là mờ vì một số lí do sau: *Thứ nhất*, một số đối tượng của một lớp được xác định có thể là mờ,những đối tượng này thuộc về lớp với độ thuộc nhất định. *Thứ hai*, khi một lớp được định nghĩa, miền trị của một thuộc tính nào đó có thể là mờ và như vậy một lớp mờ được hình thành. Ví dụ, một lớp *Hình ảnh* là mờ vì miền giá trị thuộc tính *năm* của nó sử dụng yếu tố thời gian là một tập hợp các giá trị mờ như *lâu*, *rất lâu* và *khoảng 50 năm*. *Thứ ba*, một lớp con được kế thừa một hoặc nhiều lớp cha, trong đó có ít nhất một lớp chalà lớp mờ, thì lớp con này cũng là lớp mờ.Trong CSDL HĐT mờ, các lớp là mờ vì miền trị thuộc tính của chúng chứa các giá trị mờ. Vấn đề một đối tượng thuộc về một lớp với một độ thuộc mức *k* ($k \in Z$) nào đó xảy ra vì lớp hoặc đối tượng đó có thể là mờ. Tương tự như vậy, một lớp là lớp con của một lớp khác cũng với mộtđộ thuộc mức *k* nào đó vì đó là lớp mờ . Do vậy, các đánh giá của *mốiquan hệ lớp đối tượng mờ* và *phân cấp thừa kế mờ* là cốt lõi của mô hình CSDL HĐT mờ.

3.3. Quan hệ lớp đối tượng mờ

Trong CSDL HĐT mờ, bốn trường hợp sau đây có thể được dùng để phân biệt cho các mối quan hệ lớp đối tượng: (a) *Lớp rõ và đối tượng rõ*: trường hợp này giống như trong CSDL HĐT, nghĩa là đối tượng thuộc hay không thuộc lớp một cách chắc chắn; (b) *Lớp rõ và đối tượng mờ*: lớp được xác định chính xác và có ranh giới chính xác, còn đối tượng là mờ vì giá trị thuộc tính của nó có thể mờ. Trong trường hợp này, đối tượng có thể là thành viên của lớp với độ thuộc nào đó; (c) *Lớp mờ và đối tượng rõ:* giống như trường hợp ở (b), các đối tượng có thể thuộc về lớp với độ thuộc mức*k*. Ví dụ một đối tượng học viên cao học và một lớp sinh viên trẻ; (d) *Lớp mờ và đối tượng mờ*: trong trường hợp này, đối tượng cũng thuộc về lớp với độ thuộcmức *k*.Các mối quan hệ lớp đối tượng trong (b), (c) và (d) trên đây được gọi là quan hệ lớp đối tượng mờ. Trong thực tế, trường hợp (a) có thể được xem như là trường hợp đặc biệt của mối quan hệ lớp đối tượng mờ, với độ thuộc của đối tượng vào lớp là 1. Với mỗi giá trị ngôn ngữ mờx, chúng ta sẽ định nghĩa một biểu diễn khoảng cho x. Trong thực tế, số gia tử trong các giá trị ngôn ngữ là hữu hạn nên tồn tại một số nguyên dương k^* sao cho $0<|x|\leq k^*$, $\forall x \in \boldsymbol{X}$. Với bất kì $x \in \boldsymbol{X}$, đặt $j = |x|$, với mỗi số nguyên *k* cho trước với $1 \leq k \leq k^*$, lân cận tối thiểu k của x kí hiệu là $O_{min,k}(x)$ được định nghĩa như sau: nếu k=j thì $O_{min,k}(x) = I_{k+1}(h_{-1}x) \cup I_{k+1}(h_1x)$, nếu $1 \leq k < j$ thì $O_{min,k}(x) = I_j(x)$ và nếu $j+1 \leq k \leq k^*$thì $O_{min,k}(x) = I_{k+1}(h_{-1}y) \cup I_{k+1}(h_1y)$. Từ đó, ta biểu diễn dữ liệu ngôn ngữ mờ theo định nghĩa sau:

Định nghĩa 3.1. Cho $x \in \boldsymbol{X} \cup C$, một biểu diễn khoảng của x là một tập *IRp(x)* các khoảng được xác định:

$$IRp(x) = \{O_{min,k}(x) \mid 1 \leq k \leq k^*\}$$

Cách biểu diễn dữ liệu ngôn ngữ mờ như trên có thể sử dụng để biểu diễn các dạng dữ liệu khác. Đối với giá trị số, đây là loại dữ liệu rõ, độ mờ của dữ liệu bằng 0, khi đó mỗi giá trị số a được biểu diễn bằng [a, a], và $O_{min,k}(a) = \{[a, a]\}$, với mọi $1 \leq k \leq k^*$ và $IRp(a) = \{[a, a]\}$. Còn mỗi giá trị khoảng a được biểu diễn bằng $[a-\varepsilon, a+\varepsilon]$, với với ε được xem là bán kính với tâm a. Vì $[a-\varepsilon, a + \varepsilon]$ là dữ liệu rõ nên $O_{min,k}([a - \varepsilon, a + \varepsilon]) = \{[a - \varepsilon, a + \varepsilon]\}$, với mọi $1 \leq k \leq k^*$ và $IRp([a - \varepsilon, a + \varepsilon]) = \{[a - \varepsilon, a + \varepsilon]\}$. Khi x có độ dài bé hơn *k* thì giá trị $\upsilon(x)$ là điểm đầu mút của một lớp tương đương I(u) trong P_k. điều này dẫn đến có những giá trị trong lân cận của x lại không tương tự mức *k*. Vì vậy, chúng ta sẽ xây dựng một phân hoạch khác sao cho $\upsilon(x)$ là điểm tôpô của phân hoạch với mọi x, $|x| \leq k$, như sau:

Xét $\underline{\boldsymbol{X}}$ là ĐSGT tuyến tính đầy đủ, với $H^+ = \{h_1,..., h_p\}$ và $H^- = \{h_{-1}, ..., h_{-q}\}$, trong đó p, q >1. Đặt H_1 là tập các gia tử yếu, H_2 là tập các gia tử mạnh theo nghĩa khi tác động nó sẽ làm thay đổi nghĩa mạnh hơn số gia tử trong H_1, tức là các tập H_1và H_2 gồm: $H_1 = \{h_i, h_{-j} \mid 1\leq i \leq [p/2],\ 1 \leq j \leq [q/2]\}$,$H_2 = \{h_i, h_{-j} \mid [p/2] \leq i\leq p,\ [q/2] \leq j \leq q\}$. Đặt $P_{k+1}(H_n) = \{I(h_iy) \mid y\in X_k, h_i\in H_n\}$, với n=1,2. Hai khoảng I(x) và I(y) trong $P_{k+1}(H_n)$ được gọi là liên thông với nhau nếu tồn tại các khoảng thuộc $P_{k+1}(H_n)$ liên tiếp nhau xếp từ I(x) đến I(y). Quan hệ này sẽ phân $P_{k+1}(H_n)$ thành các thành phần liên thông. Ta lại có, với mỗi $y\in \boldsymbol{X}_k$, $P_{k+1}(H_1)$ được phân thành các cụm có dạng $\{I(h_iy) \mid h_i \in H_1\}$. Hơn nữa, do $I(h_{-1}y) \leq \upsilon(y) \leq I(h_1y)$ hoặc là $I(h_1y) \leq \upsilon(y) \leq I(h_{-1}y)$ nên bao giờ ta cũng có $\upsilon(y) \in \{I(h_iy) \mid h_i\in H_1\}$.Bây giờ ta phân cụm các khoảng mờ của $P_{k+1}(H_2)$. Giả sử $\boldsymbol{X}_k = \{x_s \mid s = 0,\dots, m - 1\}$ gồm m phần tử được sắp thành một dãy sao cho $x_i\leq x_j$ khi và chỉ khi i ≤ j. Kí hiệu $H_2^- = H_2 \cap H^-$ và $H_2^+ = H_2 \cap H^+$. Để ý rằng $h_{-q} \in H_2^-$ và $h_p \in H_2^+$. Các cụm được sinh ra từ các khoảng mờ thuộc $P_{k+1}(H_2)$ có ba loại là cụm nằm bên trái x_0: $\{I(h_ix_0) \mid h_i\in H_2^+\}$; cụm nằm bên phải x_{m-1}: $\{I(h_ix_{m-1}) \mid h_i\in H_2^+\}$; cụm nằm giữa x_s và x_{s+1} với s = 0,…, m - 2: phụ thuộc vào $Sgn(h_px_s)$ và $Sgn(h_px_{s+1})$ như sau:

$\boldsymbol{C} =\{ I(h_ix_s), I(h_j{}'x_{s+1}) \mid h_i\in H_2^+, h_j{}'\in H_2^-\}$, nếu $Sgn(h_px_s) =+1$ và $Sgn(h_px_{s+1}) = +1$

$\boldsymbol{C} =\{ I(h_ix_s), I(h_j{}'x_{s+1}) \mid h_i\in H_2^+, h_j{}'\in H_2^+\}$, nếu $Sgn(h_px_s) = +1$ và $Sgn(h_px_{s+1}) = -1$

$\boldsymbol{C} =\{ I(h_ix_s), I(h_j{}'x_{s+1}) \mid h_i\in H_2^-, h_j{}'\in H_2^-\}$, nếu $Sgn(h_px_s) = -1$ và $Sgn(h_px_{s+1}) = +1$

$\boldsymbol{C} =\{ I(h_ix_s), I(h_j{}'x_{s+1}) \mid h_i\in H_2^-, h_j{}'\in H_2^+\}$, nếu $Sgn(h_px_s) =-1$ và $Sgn(h_px_{s+1}) = -1$

Tập tất cả các cụm được kí hiệu là $\boldsymbol{C}$.Vì $\{ S_k(\boldsymbol{C}) \mid C\in \boldsymbol{C}\}$ là một phân hoạch trên miền trị tham chiếu nên nó xác định một quan hệ tương đương và chúng ta sẽ gọi là quan hệ tượng tự mức *k*. Do tính chất của phân hoạch nên với mỗi giá trị x của thuộc tính, tồn tại duy nhất một cụm $\boldsymbol{C}$ sao cho $\upsilon(x) \in S_k(\boldsymbol{C})$ vàta định nghĩa khoảng tương tự mức *k* như sau:

Định nghĩa 3.2. Với mỗi *C*thuộc $\boldsymbol{C}$, ta gọi khoảng tương tự mức *k* ứng với $\boldsymbol{C}$ là: $S_k(\boldsymbol{C}) = \cup\{I(u) \mid I(u)\in \boldsymbol{C}\}$, khi đó$S_k(x) = S_k(\boldsymbol{C})$.

Mệnh đề 3.1. Cho $\underline{\boldsymbol{X}}$ là ĐSGT tuyến tính đầy đủ, trong đó H^+ và H^- có ít nhất hai phần tử. Khi đó:

1. Với mỗi k, $\{S_k(u) \mid u \in \boldsymbol{X} \cup C\}$ được xác định duy nhất và là một phân hoạch của đoạn [0,1].

2. Với mọi $x,u \in \boldsymbol{X} \cup C$, nếu $\upsilon(x) \in S_k(u)$ thì lân cận bé nhất mức k của x nằm trong $S_k(u)$, tức là $O_{min,k}(x) \in S_k(u)$.

Định nghĩa 3.3. Cho một đối tượng bất kì o trên tập thuộc tính $\{A_1, A_2 ,..., A_n\}$ của lớp *C*, $\underline{\boldsymbol{X}}$ là một ĐSGT tuyến tính đầy đủ, với mỗi k, $1 \leq k\leq k^*$, S_k là quan hệ tương tự mức k trên miền trị thuộc tính A_i của lớp *C*. Khi đó, với mọi $u\in \boldsymbol{X}$, giá trị $o(A_i)$ và u được gọi là bằng nhau mức k, kí hiệu $o(A_i) =_k u$, khi và chỉ khi $O_{min,k}(o(A_i))\in S_k(u)$.

Định nghĩa 3.4. Cho hai đối tượng bất kì o_1, o_2 trên tập thuộc tính $\{A_1, A_2 ,..., A_n\}$ của lớp *C*, $\underline{\boldsymbol{X}}$ là một ĐSGT tuyến tính đầy đủ, với mỗi k, $1 \leq k\leq k^*$, S_k là quan hệ tương tự mức k trên miền trị thuộc tính A_i của lớp *C*. Khi đó:

1. Hai giá trị $o_1(A_i)$ và $o_2(A_i)$ được gọi là bằng nhau mức k, kí hiệu $o_1(A_i) =_k o_2(A_i)$, khi và chỉ khi tồn tại một lớp tương đương $S_k(u)$ của quan hệ tương tự S_k sao cho $O_{min,k}(o_1(A_i)) \in S_k(u)$ và $O_{min,k}(o_2(A_i)) \in S_k(u)$.

2. Hai giá trị $o_1(A_i)$ và $o_2(A_i)$ được gọi là khác nhau mức k, kí hiệu $o_1(A_i) \neq_k o_2(A_i)$, nếu không tồn tại một lớp tương đương $S_k(u)$ của quan hệ tương tự S_k sao cho $O_{min,k}(o_1(A_i)) \in S_k(u)$ và $O_{min,k}(o_2(A_i)) \in S_k(u)$.

Bổ đề 3.1. Quan hệ bằng nhau theo mức k ($=_k$) là một quan hệ tương đương.

Hệ quả 3.1. Cho o_1, o_2 là hai đối tượng bất kì trên tập thuộc tính $\{A_1, A_2, ..., A_n\}$ của lớp *C*, S_k là quan hệ tương tự mức k ($0 < k \leq k^*$) trên miền trị thuộc tính A_i của lớp *C*,

1. Nếu $o_1(A_i) =_k o_2(A_i)$ thì $o_1(A_i) =_{k'} o_2(A_i)$, $\forall k' < k$
2. Nếu $o_1(A_i) \neq_k o_2(A_i)$ thì $o_1(A_i) \neq_{k'} o_2(A_i)$, $\forall k' > k$

4. PHỤ THUỘC DỮ LIỆU TRONG MÔ HÌNH CƠ SỞ DỮ LIỆU HƯỚNG ĐỐI TƯỢNG MỜ

4.1. Phụ thuộc hàm mờ trong lớp đối tượng

Hai đối tượng o_1 và o_2 được gọi là bằng nhau mức k trên tập X, kí hiệu $o_1(X) =_k o_2(X)$, nếu với mọi $A \in X$, ta có $o_1(A) =_k o_2(A)$. Hai đối tượng o_1 và o_2 được gọi là khác nhau mức k trên tập X, kí hiệu $o_1(X) \neq_k o_2(X)$, nếu tồn tại $A \in X$: $o_1(A) \neq_k o_2(A)$.

Định nghĩa 4.1. Cho lớp đối tượng mờ *C* với tập thuộc tính U, $X, Y \subseteq U$, với mỗi số nguyên k và $1 \leq k \leq k^*$. Ta nói lớp *C* thỏa mãn phụ thuộc hàm mờ X xác định Y với mức k, kí hiệu $X \sim>_k Y$, nếu $\forall o_1, o_2 \in C$, $o_1(X) =_k o_2(X) \Rightarrow o_1(Y) =_k o_2(Y)$.

Thuật toán 4.1. Kiểm tra lớp *C* có thỏa mãn phụ thuộc hàm mờ $X \sim>_k Y$ không?

Vào: Lớp *C* cùng với tập thuộc tính X, Y, tập các đối tượng $\{o_i, i=1, \ldots, m\}$

Ra: *True* nếu *C* thỏa $X \sim>_k Y$, ngược lại *False*

Phương pháp:

1. Xây dựng các Đại số gia tử cho các thuộc tính mờ có trong hai tập thuộc tính X và Y
2. Xây dựng lân cận tối thiểu mức k của các giá trị các đối tượng trên X và Y
3. Duyệt lần lượt từng cặp đối tượng của lớp *C* để phát hiện cặp đối tượng không thỏa phụ thuộc hàm:

For (với mỗi đối tượng $o_i \in C$, i=1,…,m-1)

For (với mỗi đối tượng $o_j \in C$, j=i+1,…,m)

If ($o_i(X) =_k o_j(X)$) và ($o_i(Y) \neq_k o_j(Y)$)Return*False*;

Return *True*;

Thuật toán 4.1 đảm bảo tính dừng vì số thuộc tính (n) và số đối tượng (m) của lớp *C* hữu hạn và có độ phức tạp của thuật toán là $O(m^2 * n)$.

Ví dụ 4.1. Ta xét một lớp “*Nhanvien*” bao gồm các thuộc tính như sau:

Class **NhanVien** {

Oid: allID

TenNV : string

BoPhan: string

CongViec: string

KinhNghiem : [fuzzy] domain [0 .. 40]: float

Luong: [fuzzy] domain [2..30]: float

ThueTN: [fuzzy] domain [0 .. 4,15]: float

}

Trong đó, *TenNV*, *BoPhan* và *CongViec* là các thuộc tính kinh điển, còn *KinhNghiem*,*Luong* và *ThueTN* là các thuộc tính mờ. Trong thực tế, giữa các thuộc tính trên có thể tồn tại những quan hệ không chính xác, ví dụ như “*Các nhân viên trong cùng một bộ phận có công việc và kinh nghiệm giống nhau thì phải có mức lương gần bằng nhau*”, …, những mối quan hệ như vậy được gọi là phụ thuộc hàm mờ. Ta xét một số các đối tượng của lớp “*Nhanvien*” như trong Bảng 1

Bảng 1. Lớp “Nhanvien”.

Oid TenNV	BoPhan	CongViec	KinhNghiem	Luong	ThueTN
oid_1Bình	Kĩ thuật	Kĩ sư	25	15	0,36
oid_2Lan	Kế toán	Kế toán viên	khá thấp	khá thấp	0
oid_3Minh	Kĩ thuật	Quản lí	rất cao	rất cao	rất cao
oid_4Tuấn	Kĩ thuật	Kĩ sư	khoảng 26	khả năng cao	cao
oid_5Văn	Kế toán	Kế toán viên	5,5	5	0

Giả sử ta có yêu cầu kiểm tra lớp “*Nhanvien”* có thỏa phụ thuộc hàm mờ:*BoPhan,CongViec,KinhNghiem*$\sim>_2$*Luong* ?

Các thuộc tính *KinhNghiem*, *Luong* và *ThueTN* là các thuộc tính mờ, nhận các giá trị số, giá trị khoảng và giá trị ngôn ngữ, nên trước tiên, chúng ta sẽ xây dựng một đại số gia tử cho mỗi thuộc tính mờ:

- Đối với thuộc tính Kinh Nghiem, ta có: *G* = {*thấp*, *cao*}, H^-= *{ít, khả năng*} và H^+= {*khá*, *rất*}. Các tham số mờ: *fm(thấp)* = 0,35; *fm(cao*) = 0,65; *µ*(*rất*) = 0,3; *µ*(*khá*) = 0,25; *µ*(*khả năng*) = 0,2; *µ*(*ít*) = 0,25. cdom(*KinhNghiem*) = [0; 40]; ta sẽ dùng hệ số r = 40 để chuyển đổi từ [0; 1] qua [0; 40]

+ Với *k* = 1, ta có: $|I_r(\textit{thấp})| = fm(\textit{thấp}) \times 40 = 14$. Suy ra $I_r(\textit{thấp})$ = [0,0; 14], $fm_r(\textit{rất thấp}) = \mu(\textit{rất}) \times fm(\textit{thấp}) \times 40 = 4,2$; $fm_r(\textit{khá thấp}) = \mu(\textit{khá}) \times fm(\textit{thấp}) \times 40 = 3,5$. Vì $\{I_r(\textit{rất thấp}), I_r(\textit{khá thấp}), I_r(\textit{khả năng thấp}), I_r(\textit{ít thấp})\}$ là một phân hoạch của $I_r(\textit{thấp})$, ta suy ra $I_r(\textit{rất thấp})$ = [0,0; 4,2], $I_r(\textit{khá thấp})$ = (4,2; 7,7]. Vậy, $O_{min,1}(\textit{khá thấp}) = I_r(\textit{khá thấp})$ = (4,2; 7,7].Tương tự, ta có: $O_{min,1}(\textit{rất cao}) = I_r(\textit{rất cao})$ = (32,2; 40].

+ Với k = 2: $O_{min,2}$(*khá thấp*) = I_r(*khá khá thấp*) ∪ I_r(*khả năngkhá thấp*) = (5,25; 6,825], $O_{min,2}$(*rất cao*) = I_r(*khả năng rất cao*) ∪ I_r(*khá rất cao*) = (34,15; 37,66].

Đối với giá trị số và giá trị khoảng, ta có:

$O_{min,k}$(25) = [25; 25]; $O_{min,k}$(5,5) = [5,5; 5,5]; $O_{min,k}$(khoảng 26) = [25; 27], $\forall 1 \le k \le k^*$.

-Đối với thuộc tính Luong, ta có: G = {*thấp, cao*}, H^-= {*ít, khả năng*} và H^+= {*khá, rất*}. Các tham số mờ: *fm(thấp)* = 0,25; *fm(cao)* = 0,75; μ(*rất*) = 0,3; μ(*khá*) = 0,3; μ(*khả năng*) = 0,25; μ(*ít*) = 0,15; cdom(*Luong*) = [2; 30].

+ Với k = 1, ta có:$O_{min,1}$(*rất cao*) = I_r(*rất cao*) = (23,7; 30], $O_{min,1}$(*khả năng cao*) = I_r(*khả năng cao*) = (12,15; 17,4]; $O_{min,1}$(*khá thấp*) = I_r(*khá thấp*) = (4,1; 6,2].

+ Với k = 2, ta có:$O_{min,2}$(*rất cao*) = I_r((*khả năng rất cao*) ∪I_r(*khá rất cao*) = (24,645; 28,11], $O_{min,2}$(*khả năng cao*) = I_r((*khả năng khả năng cao*) ∪I_r(*khá khả năng cao*) = (12,9375; 15,825], $O_{min,2}$(*khá thấp*) =I_r(*khá khá thấp*)∪I_r(*khả năngkhá thấp*) = (4,73 ; 5,885].

Ta thấy lớp "*NhanVien*" thỏa phụ thuộc hàm mờ: *BoPhan,Congviec,KinhNghiem*$\sim>_2$*Luong* vì:

- *BoPhan* và *CongViec* là hai thuộc tính rõ, nên các đối tượng của lớp hoặc là bằng nhau theo mọi mức hoặc là khác nhau với mọi mức trên hai thuộc tính này. Do đó, ta chỉ cần xét các đối tượng bằng nhau trên hai thuộc tính rõ này:

- Xét đối tượng thứ nhất (oid_1) và thứ tư (oid_4), ta có: oid_1(*KinhNghiem*) $=_2$ oid_4(*KinhNghiem*) vì $\exists S_2$(cao)= I_r((*rất khả năng cao*)∪ I_r((*ít khá cao*) = (24,14; 25,7] ∪ (25,7; 27,325] = (24,14; 27,325] : $O_{min,1}$(oid_1(*KinhNghiem*)) = [25; 25] $\subseteq \exists S_1$(cao) và $O_{min,1}$(oid_4(*KinhNghiem*)) = [25; 27] $\subseteq \exists S_1$(cao). Ta lại có: oid_1(*Luong*) $=_2$ oid_4(*Luong*) vì $\exists S_2$(khả năng cao) = I_r((*khả năng khả năng cao*)∪ I_r((*khá khả năng cao*) = (12,9375; 14,25] ∪ (14,25; 15,825] = (12,9375; 15,825] : $O_{min,2}$(oid_1(*Luong*)) = [15; 15] ⊆ S_2(khả năng cao) và $O_{min,2}$(oid_4(*Luong*))= (12,9375; 15,825] ⊆ S_2(khả năng cao).

- Xét đối tượng thứ hai (oid_2) và thứ năm (oid_5), ta có: oid_2(*KinhNghiem*) $=_2$ oid_5(*KinhNghiem*) vì $\exists S_2$(*khá thấp*) = I_r((*khá khá thấp*)∪ I_r((*khả năngkhá thấp*) = (5,25; 6,825] : $O_{min,2}$(oid_2(*KinhNghiem*)) = (5,25; 6,825]⊆ S_2(*khá thấp*) và $O_{min,2}$(oid_4(*KinhNghiem*)) = [5,5; 5,5] ⊆ S_2(*khá thấp*). Ta lại có: oid_2(*Luong*) $=_2$ oid_5(*Luong*) vì $\exists S_2$(*khá thấp*) = I_r((*khá khá thấp*)∪ I_r((*khả năngkhá thấp*) = (4,73; 5,885]: $O_{min,2}$(oid_2(*Luong*)) = (4,73; 5,885]⊆ S_2(*khá thấp*) và $O_{min,2}$(oid_5(*Luong*))= [5; 5] ⊆ S_2(*khá thấp*).

Mệnh đề 4.1. Cho lớp đối tượng mờ C với tập thuộc tính U, và X, Y, Z ⊆ U. Ta có các luật suy diễn trên các phụ thuộc hàm mờ:

- Luật phản xạ: Nếu X ⊇ Y thì X$\sim>_k$ Y

- Luật gia tăng: Nếu X$\sim>_k$ Y thì XZ$\sim>_k$ YZ

- Luật bắc cầu: Nếu X$\sim>_k$ Y và Y$\sim>_k$Z thì X$\sim>_k$ Z

- Luật tựa bắc cầu: Nếu X$\sim>_{k1}$ Y và Y$\sim>_{k2}$ Z thì X$\sim>_{min(k1, k2)}$ Z

Các luật suy diễn trên là đúng đắn.

Chứng minh:

- Luật phản xạ: Vì X ⊇ Y, nên với hai đối tượng bất kì o_1, o_2 của lớp C, nếu $o_1(X) =_k o_2(X)$ thì $o_1(Y) =_k o_2(Y)$. Vậy theo định nghĩa phụ thuộc hàm mờ ta có X$\sim>_k$ Y.

- Luật gia tăng: Theo giả thiết C thỏa phụ thuộc hàm mờ $X\sim>_k Y$, nên ta có $o_1(X) =_k o_2(X)$ thì $o_1(Y) =_k o_2(Y)$, với mọi o_1, $o_1 \in C$ (1). Mặc khác, ta lại có $o_1(XZ) =_k o_2(XZ)$ nên $o_1(X) =_k o_2(X)$ và $o_1(Z) =_k o_2(Z)$ (2). Từ (1)&(2) ta có $o_1(YZ) =_k o_2(YZ)$. Vậy $XZ\sim>_k YZ$.

- Luật bắt cầu: Từ $X\sim>_k Y$ ta có nếu $o_1(X) =_k o_2(X)$ thì $o_1(Y) =_k o_2(Y)$. Và từ $Y\sim>_k Z$ ta lại có nếu $o_1(Y) =_k o_2(Y)$ thì $o_1(Z) =_k o_2(Z)$. Do đó, nếu $o_1(X) =_k o_2(X)$ thì $o_1(Z) =_k o_2(Z)$, vậy $X\sim>_k Z$

- Luật tựa bắc cầu:

+ Theo giả thiết ta có $X\sim>_{k1} Y : o_1(X) =_{k1} o_2(X)$ thì $o_1(Y) =_{k1} o_2(Y)$, và $X\sim>_{k1} Y : o_1(Y) =_{k2} o_2(Y)$ thì $o_1(Z) =_{k2} o_2(Z)$ (1).

+ Xét trường hợp $k_1>k_2$: nếu $o_1(X) =_{k1} o_2(X) \Rightarrow o_1(X) =_{k2} o_2(X)$, và $o_1(Y) =_{k1} o_2(Y) \Rightarrow o_1(Y) =_{k2} o_2(Y)$ (theo hệ quả 3.1) (2). Từ (1) và (2) ta được: $o_1(X) =_{k2} o_2(X) \Rightarrow o_1(Z) =_{k2} o_2(Z)$ (3).

+ Xét trường hợp $k_1<k_2$: cũng theo hệ quả 3.1 nếu $o_1(Y) =_{k2} o_2(Y) \Rightarrow o_1(Y) =_{k1} o_2(Y)$ và $o_1(Z) =_{k2} o_2(Z) \Rightarrow o_1(Z) =_{k1} o_2(Z)$ (4). Từ (1) và (4) ta được: $o_1(X) =_{k1} o_2(X) \Rightarrow o_1(Z) =_{k1} o_2(Z)$ (5). Từ (3) và (5): $o_1(X) =_{\min(k1,\ k2)} o_2(X) \Rightarrow o_1(Z) =_{\min(k1,\ k2)} o_2(Z)$. Vậy nếu $X\sim>_{k1} Y$ và $Y\sim>_{k2} Z$ thì $X\sim>_{\min(k1,\ k2)} Z$.

4.2. Phụ thuộc hàm mờ với lượng từ ngôn ngữ

Việc sử dụng các lượng từ ngôn ngữ như *một vài*, *hầu hết*, … vào trong phụ thuộc hàm mờ làm cho việc mô tả các phụ thuộc dữ liệu được mềm dẻo và gần với thực tế hơn, chẳng hạn như: “ *Hầu hết*các nhân viên trong cùng một bộ phận có công việc và kinh nghiệm giống nhau thì có mức lương xấp xỉ nhau”.Zadel chia lượng từ ngôn ngữ thành hai loại đó là: lượng từ tuyệt đối và lượng từ tỉ lệ.Gọi *Q* là lượng từ trong phụ thuộc hàm mờ, O là tập đối tượng ban đầu của lớp C, $\|O\| = m$ là số đối tượng của tập O, miền trị $D_C = [0..m]$. Chúng ta có thể chia lượng từ *Q* thành hai trường hợp:

1. *Trường hợp Q là lượng từ tuyệt đối:* Kí hiệu $\|Q\|$ là số lượng xác định của lượng từ *Q*, nếu *Q* đơn điệu tăng: Ta xây dựng một hàm $f_Q^A : D_C \to \{0,1\}$ sao cho $\forall x \in D_C, f_Q^A(x) = 1$ nếu $x \le |Q|$ và $f_Q^A(x) = 0$ nếu ngược lại.Nếu *Q* đơn điệu giảm: Ta xây dựng một hàm $f_Q^D : D_C \to \{0,1\}$ sao cho $\forall x \in D_C, f_Q^D(x) = 1$ nếu $x \ge |Q|$ và $f_Q^D(x) = 0$ nếu ngược lại.

2. *Trường hợp Q là lượng từ tỉ lệ*:

Vì $D_c = [0, m]$ là liên tục nên sử dụng sử dụng phép biến đổi tuyến tính chuyển về khoảng [0,1]. Khi đó ta xây dựng hai khoảng mờ của hai khái niệm nguyên thủy *nhỏ* và *lớn*, kí hiệu là I(*nhỏ*) và I(*lớn*) với độ dài tương ứng là *fm*(*nhỏ*) và *fm*(*lớn*) sao cho chúng tạo thành một phân hoạch của [0, 1]. Tiếp đến, ta đi xây dựng các lớp tương đương S(*1*), S(*lớn*), S(*W*), S(*nhỏ*), S(*0*). Từ đó, ta có thể khẳng định rằng tổng số đối tượng của lớp C thỏa điều kiện mờ với lượng từ *Q* nếu tổng số đối tượng thuộc về một trong các khoảng:*S*(*1*) ×m, *S*(*lớn*) ×m, *S*(*W*) ×m, *S*(*nhỏ*) × m hoặc *S*(*0*) ×m.

Định nghĩa 4.2. Cho lớp đối tượng mờ *C* với tập thuộc tính U, và X, Y$\subseteq$ U. Ta gọi O_{thoa} là tập các đối tượng của lớp *C* thỏa mãn tập X và Y với mức *k* và được xác định như sau:$O_{thoa} = \{o_i \in C: (\exists j \neq i,\ o_i(X)_k = o_j(X) \wedge o_i(Y) =_k o_j(Y)) \vee (\forall j \neq i,\ o_i(X)_k \neq o_j(X))\}$;$O_{khong}$ là tập các đối tượng của lớp *C* thỏa mãn tập X nhưng không thỏa mãn tập Y với mức *k* và được xác định: $O_{khong} = \{o_i \in C: \exists j \neq i, o_i(X)_k = o_j(X) \wedge o_i(Y) \neq_k o_j(Y)\}$.

Định nghĩa 4.3. Cho lớp đối tượng mờ *C* với tập thuộc tính U, và X, Y$\subseteq$ U. Ta nói lớp *C* thỏa mãn phụ thuộc hàm mờ X xác định Y với mức *k* và lượng từ ngôn ngữ *Q*, kí hiệu $X \sim>_{kQ} Y$, nếu :

1. *Q* là lượng từ tuyệt đối đơn điệu tăng thì $f_Q^A(\| O_{thoa} \|) = 1$
2. *Q* là lượng từ tuyệt đối đơn điệu giảm thì $f_Q^D(\| O_{thoa} \|) = 1$
3. *Q* là lượng từ "*một ít*" thì $\|O_{thoa}\|/m \in S(0)$
4. *Q* là lượng từ "*khoảng một nửa*" thì $\|O_{thoa}\|/m \in S(W)$
5. *Q* là lượng từ "*hầu hết*" thì $\|O_{thoa}\|/m \in S(1)$
6. *Q* là lượng từ "*với mọi*" thì $\|O_{thoa}\|/m = 1$

Thuật toán 4.2. Kiểm tra lớp *C* có thỏa mãn phụ thuộc hàm mờ $X \sim>_{kQ} Y$ không?

Vào: Lớp *C* cùng với tập thuộc tính X, Y, tập các đối tượng $\{o_i, i = 1, \ldots, m\}$, mức k và lượng từ Q

Ra: *True* nếu *C* thỏa $X \sim>_{kQ} Y$, ngược lại *False*

Phương pháp:

1. Xây dựng các Đại số gia tử cho các thuộc tính mờ có trong hai tập thuộc tính X và Y.
2. Xây dựng lân cận tối thiểu mức k của các giá trị các đối tượng trên X và Y.
3. $O_{thoa} = O$.
4. Duyệt lần lượt từng đối tượng của lớp C để tìm ra đối tượng thỏa mãn tập X nhưng không thỏa mãn tập Y với mức *k*

For (với mỗi đối tượng $o_i \in C$, i=1,…m-1)

If (tồn tại $j \neq i$: $(o_i(X) =_k o_j(X))$ và $(o_i(Y) \neq_k o_j(Y))$)

$O_{thoa} = O_{thoa} \backslash o_i$;

5. result = *False;*
6. if (*Q là lượng từ tuyệt đối*)
7. if ($f_Q^A(\| O_{thoa} \|) = 1$ or $f_Q^D(\| O_{thoa} \|) = 1$) then result = *True;*
8. if (*Q là lượng từ tỉ lệ*)
9. Xây dựng các khoảng *S*(*1*), *S*(*lớn*), *S*(*W*), *S*(*nhỏ*), *S*(*0*)
10. Case *Q* of
11. "*một ít*": if $\|O_{thoa}\|/m \in S(0)$ then result = *True;*
12. "*khoảng một nửa*" : if $\|O_{thoa}\|/m \in S(W)$ then result = *True;*
13. "*hầu hết*" : if $\|O_{thoa}\|/m \in S(1)$ then result = *True;*
14. "*với mọi*" : if $\|O_{thoa}\|/m = 1$ then result = *True;*
15. Return *result*;

Thuật toán 4.2 luôn dừng vì số thuộc tính (n) và số đối tượng (m) của lớp *C* hữu hạn và độ phức tạp của thuật toán là $O(m^2 * n)$

Ví dụ 4.2. Ta xét một lớp "*NhanVien*" có cấu trúc như ví dụ 4.1 và các đối tượng như trong Bảng 2.

Bảng 2. Lớp "NhanVien".

Oid	TenNV		BoPhan	CongViec ThueTN	KinhNghiem	Luong
oid_1	Bình	Kĩ thuật 0,36	Kĩ sư	25	15	
oid_2	Lan	Kế toán	Kế toán viên	khá thấp	khá thấp	0
oid_3	Minh	Kĩ thuật cao	Quản lí	rất cao	rất	cao
oid_4	Tuấn	Kĩ thuật thấp	Kĩ sư	khoảng 26	khả năng	cao
oid_5	Văn	Kế toán	Kế toán viên	5,5	5	0
oid_6	Hải thấp	Kĩ thuật	Quản lí	35	khoảng 26	ít
oid_7	Nhân	Kế toán	Kế toán viên	khoảng 10	khoảng 8	0
oid_8	Đệ	Kĩ thuật 0,05	Kĩ sư	8	10	
oid_9	Sương	Kế toán	Kế toán viên	10	7	0

Xét ràng buộc có sử dụng lượng từ: "*Hầu hết* các nhân viên trong cùng một bộ phận có công việc và kinh nghiệm giống nhau thì có lương gần bằng nhau". Ràng buộc trên tương ứng với phụ thuộc hàm mờ *BoPhan,CongViec,KinhNghiem* $\sim>_{2,Hầu\ hết}$ *Luong*. Ta sẽ sử dụng thuật toán 4.2 để kiểm tra lớp *C* có thỏa phụ thuộc hàm mờ với lượng từ ngôn ngữ trên hay không. Dựa vào ví dụ 4.1, ta có các kết quả tính toán:

- Đối vớithuộc tính KinhNghiem:

$O_{min,k}(oid_1(Kinhnghiem)) = O_{min,k}(25)=[25; 25]$, $\forall 1 \le k \le k^*$,$O_{min,2}(oid_2(KinhNghiem)) = O_{min,2}(khá\ thấp) = (5,25; 6,825]$; $O_{min,2}(oid_3(KinhNghiem)) = O_{min,2}(rất\ cao) = (34,15; 37,66]$; $O_{min,k}(oid_4(Kinhnghiem)) = O_{min,k}(khoảng\ 26) = [25; 27]$, $\forall 1 \le k \le k^*$; $O_{min,k}(oid_5(KinhNghiem)) = O_{min,k}(5,5) = [5,5; 5,5]$, $\forall 1 \le k \le k^*$,$O_{min,2}(oid_6(KinhNghiem)) = O_{min,k}(cao) = (24,14; 27,325]$; $O_{min,k}(oid_7(KinhNghiem)) = O_{min,k}(8) = [8; 8]$, $\forall 1 \le k \le k^*$; $O_{min,k}(oid_8(KinhNghiem)) = O_{min,k}(khoảng\ 10) = [9; 11]$, $\forall 1 \le k \le k^*$; $O_{min,k}(oid_9(KinhNghiem)) = O_{min,k}(10) = [10; 10]$, $\forall 1 \le k \le k^*$.

-Đối với thuộc tính Luong:

$O_{min,k}(oid_1(Luong)) = O_{min,k}(15) = [15;15]$, $\forall 1 \le k \le k^*$; $O_{min,2}(oid_2(Luong)) = O_{min,2}(khá\ thấp) = (4,73; 5,885]$; $O_{min,2}(oid_3(Luong)) = O_{min,2}(rất\ cao) = (24,645; 28,11]$; $O_{min,2}(oid_4(Luong)) = O_{min,2}(khả\ năng\ cao) = (12,9375; 15,825]$; $O_{min,k}(oid_5(Luong)) = O_{min,k}(5) = [5; 5]$, $\forall 1 \le k \le k^*$; $O_{min,k}(oid_6(Luong)) = O_{min,k}(khoảng\ 26) = [25; 27]$, $\forall 1 \le k \le k^*$; $O_{min,k}(oid_7(Luong)) = O_{min,k}(10) =$

[10; 10], $\forall 1 \le k \le k^*$; $O_{min,k}$(oid$_8$(*Luong*)) = $O_{min,k}$(khoảng 8) = [7; 9], $\forall 1 \le k \le k^*$; $O_{min,k}$(oid$_9$(*Luong*)) = $O_{min,k}$(7) = [7; 7], $\forall 1 \le k \le k^*$.

Theo bước 3 và 4 trong thuật toán 4.2, ta xây dựng được tập O_{thoa}= { oid$_1$, oid$_2$, oid$_3$, oid$_4$, oid$_5$, oid$_6$, oid$_8$}. Vì "*hầu hết*" là lượng từ tương đối nên ta đi xây dựng khoảng mờ S(*1*). Chọn *fm(lớn)* = 0,6; *fm(nhỏ)* = 0,4; μ(*rất*) = 0,4; μ(*hơn*) = 0,15; μ(*khả năng*) = 0,25; μ(*ít*) = 0,2. Ta có *fm(rất lớn)* = 0,4 × 0,6 = 0,24; do đó S(*1*) = (0,76; 1]. Ta lại có $||O_{thoa}||/m = 7/9 = 0{,}777 \in$ S(*1*), vậy theo định nghĩa lớp đã cho thỏa phụ thuộc hàm mờ *BoPhan,Congviec,KinhNghiem* ~> $_{2,\text{Hầu hết}}$*Luong*.

4.3. Các dạng phụ thuộc hàm mờ và các dạng chuẩn lớp đối tượng mờ

Mục đích chính của việc xây dựng một mô hình dữ liệu là tạo ra một dạng biểu diễn chính xác của dữ liệu, các mối liên kết các dữ liệu và các ràng buộc. Để đạt được mục tiêu này, chúng ta phải xác định một tập hợp các quan hệ thích hợp. Một cách tiếp cận để có thể xác định các quan hệ này được gọi là chuẩn hóa. Chuẩn hóa giúp cho người thiết kế cơ sở dữ liệu kiểm tra các quan hệ đã được chuẩn hóa hay chưa để tránh xảy ra các bất thường khi cập nhật.Trước hết, chúng ta tìm hiểu một số khái niệm về *k*-khóa, phụ thuộc hàm mờ đầy đủ, phụ thuộc hàm mờ bộ phận và phụ thuộc hàm mờ bắc cầuđể làm cơ sở xây dựng các dạng chuẩn đối tượng mờ.

4.3.1. Các dạng phụ thuộc hàm mờ

Định nghĩa 4.4. Cho lớp đối tượng mờ C với tập thuộc tính U, X, Y$\subseteq$ U, A$\in$U, với mỗi số nguyên k và $1 \le k \le k^*$. Khi đó:

1.A phụ thuộc hàm mờ đầy đủ vào X theo mức k, kí hiệu X~>$_{kF}$ A, nếu X ~>$_k$A và không tồn tại Y$\subset$X để cho Y~>$_k$A.

2. A phụ thuộc hàm mờ bộ phận vào X theo mức *k*, kí hiệu X~>$_{kP}$ A, nếu X ~>$_k$A và tồn tại Y$\subset$X để cho Y~>$_k$A.

3.A phụ thuộc hàm mờ bắc cầu vào X theo mức k, kí hiệu X~>$_{kT}$ A, nếu tồn tại Y$\subseteq$ U để cho X~>$_k$Y, Y!~>$_k$X, Y~>$_k$A, A$\notin$XY.

Ví dụ 4.3. Ta xét một lớp "*SinhVien*" bao gồm các thuộc tính {MaSV, Hoten, MaLop, TenLop, Monhoc, Diem} và tập phụ thuộc hàm mờ F = { MaSV~>$_1$ Hoten, MaLop; MaLop~>$_1$ TenLop; MaSV, Monhoc~>$_1$ Diem}.

Ta có MaSV, Monhoc~>$_1$ Diem là một phụ thuộc hàm mờ đầy đủ theo mức 1 vì Diem không phụ thuộc hàm mờ vào MaSV và cũng không phụ thuộc hàm mờ vào Monhoc.Ta có MaSV, MaLop~>$_1$ TenLop là một phụ thuộc hàm mờ bộ phận theo mức 1 vì TenLop phụ thuộc hàm mờ vào Malop.Ta có MaSV~>$_1$ TenLop là một phụ thuộc hàm mờ bắc cầu theo mức 1 vì MaSV~>$_1$ MaLop; MaLop~>$_1$ TenLop; TenLop$\notin$ { MaSV, MaLop}.

Trong mô hình CSDL HĐT, mỗi đối tượng có một sự tồn tại độc lập với giá trị của nó thông qua định danh đối tượng và các định danh này là ẩn đối với người dùng. Điều này có thể dẫn đến trường hợp tồn tại các đối tượng bằng nhau trên mọi giá trị của các thuộc tính, đây là một dạng dư thừa dữ liệu trong CSDL HĐT. Để khắc phục vấn đề này, khi thiết kế CSDL HĐT, người ta thường sử dụng một tập các thuộc tính của đối tượng mà giá trị của nó dùng để xác định duy nhất một đối tượng trong lớp Với khái niệm phụ thuộc hàm mờ theo mức *k*, khái niệm khóa của lớp mờ có thể được phát biểu như sau:

Định nghĩa 4.5. Cho lớp đối tượng mờ *C* với tập thuộc tính U, và K$\subseteq$ U. K được gọi là *k-khóa* của lớp đối tượng mờ *C* nếu thỏa mãn đồng thời hai điều kiện sau đây:

1. K $\sim>_k$ U

2. Không tồn tại K'$\subset$K sao cho K' $\sim>_k$ U

Thuộc tính A được gọi là thuộc tính *k-khóa* của *C* nếu A thuộc một *k-khóa* nào đó của *C*,ngược lại, A được gọi là thuộc tính không *k-khóa* của *C*.

Ví dụ 4.4.Xét lớp "*SinhVien*" như ví dụ 4.3 , ta có(MaSV,Monhoc)là một *1-khóa* vì MaSV,Monhoc$\sim>_1$U; còn MaSV!$\sim>_1$U và Monhoc!$\sim>_1$U.

4.3.2. Các dạng chuẩn lớp đối tượng mờ

Định nghĩa 4.6. Cho lớp mờ *C* với tập thuộc tính U. Khi đó:

1. Lớp mờ *C* được gọi là ở dạng chuẩn mờ 1 mức *k*, kí hiệu *k*-1FONF nếu với mọi thuộc tính A$\in$U, thì A hoặc là thuộc tính không có kiểu bộ nhận giá trị tập hoặc A chỉ chứa các thuộc tính không có kiểu bộ nhận giá trị tập.

2. Lớp mờ *C*được gọi là ở dạng chuẩn mờ 2 mức *k*, kí hiệu *k*-2FONF nếu C ở dạng chuẩn *k*-1FONF và mọi thuộc tính không *k*-khóa đều phụ thuộc mờ đầy đủ mức *k* vào mọi *k*-khóa.

3. Lớp mờ *C*được gọi là ở dạng chuẩn mờ 3 mức *k*, kí hiệu*k*-3FONF nếu C ở dạng chuẩn *k*-1FONF và không tồn tại thuộc tính không *k*-khóa phụ thuộc mờ bắc cầu mức *k* vào *k*-khóa.

Ví dụ 4.5. Ta xét một lớp "*SinhVien*" ở dạng chuẩn *1*-1FONF bao gồm các thuộc tính U={MaSV, Hoten, MaLop, TenLop, Monhoc, Diem} và tập phụ thuộc hàm mờ F={MaSV$\sim>_1$ Hoten, MaLop ; MaLop$\sim>_1$ TenLop; MaSV, Monhoc$\sim>_1$ Diem}.

Tập thuộc tính (MaSV, Monhoc) là một *1*-khóa của lớp vì MaSV,Monhoc$\sim>_1$U và MaSV!$\sim>_1$U; Monhoc!$\sim>_1$U. Ta thấy lớp trên không ở dạng chuẩn *1*-2FONF vì Hoten, TenLop phụ thuộc hàm mờ bộ phận vào vào *k*-khóa (MaSV, Monhoc). Đây chính là nguyên nhân của các dị thường về dữ liệu trong lớp. Ta có thể cấu trúc lại lớp "*SinhVien*" trên thành hai lớp là "*SV-Lop*" và "*SV-Diem*" ở dạng chuẩn *1*-2FONF như sau:

- Lớp "*SV-Lop*" gồm các thuộc tính {MaSV, Hoten, MaLop, TenLop} và tập phụ thuộc hàm mờ F_1={ MaSV$\sim>_1$ Hoten, MaLop ; MaLop$\sim>_1$ TenLop}.

- Lớp "*SV-Diem*" gồm các thuộc tính {MaSV,Monhoc,Diem} và tập phụ thuộc hàm mờ F_2={ MaSV, Monhoc $\sim>_1$Diem}.

Lớp "*SV-Lop*" như ví dụ trên đã ở dạng chuẩn *1*-2FONF nhưng không phải là hết những dị thường, do tồn tại phụ thuộc hàm bắc cầu mờ MaSV$\sim>_{1T}$ Tenlop. Còn lớp "SV-Diem" như trên thuộc dạng chuẩn *1*-3FONF.

5. KẾT LUẬN

Trên cơ sở mô hình CSDL HĐT mờ theo cách tiếp cận ĐSGT đã xây dựng, chúng tôi đãnghiên cứu và xây dựng một số dạng phụ thuộc dữ liệu trên mô hình này. Bài báo đã đề xuất phụ thuộc hàm mờ trong lớp đối tượng mờ, các hệ luật suy diễn trên các phụ thuộc hàm mờ và chứng minh tính đúng đắn của chúng. Bên cạnh đó, một số dạng phụ thuộc hàm mờ khác cũng được đề cập làm cơ sở cho việc xây dựng các dạng chuẩn của mô hình CSDL HĐT mờ. Vấn đề lượng từ ngôn ngữ cũng được chúng tôi đề cập theo cách tiếp cận riêng và đưa vào trong phụ thuộc hàm mờ, làm cho phụ thuộc hàm được mềm dẻo và gần gũi với thực tế. Một số dạng phụ thuộc dữ liệu đặc biệt trong lớp đối tượng mờ sẽ được chúng tôi nghiên cứu trong những bài báo sau.

TÀI LIỆU THAM KHẢO

1. Nguyễn Công Hào - Phụ thuộc hàm mờ với lượng từ ngôn ngữ theo cách tiếp cận đại số gia tử, Tạp chí Công nghệ thông tin và Truyền thông **22** (2) (2009) 87-93.
2. Nguyễn Công Hào - Về một số dạng chuẩn mờ theo cách tiếp cận đại số gia tử, Tạp chí Công nghệ thông tin và Truyền thông **21** (17) (2008) 101-107.
3. Thuan H., Thanh T. T. - Fuzzy Functional Dependencies with Linguistic Quantifiers, Tạp chí Tin học và Điều khiển học **18** (2) (2002) 97-108.
4. Ho Cam Ha, Vu Duc Quang - Fuzzy function dependencies in fuzzy object-oriented databases, Journal of HNUE **7** (2011) 23-31.
5. Zadeh L. A. - A Computational approach to fuzzy quantifiers in natural languages, Computers and Mathematics with Applications **9** (1) (1983) 149-184.
6. Nguyễn Công Hào, Trương Thị Mỹ Lê - Mô hình cơ sở dữ liệu hướng đối tượng mờ theo cách tiếp cận đại số gia tử, Tạp chí Tin học và Điều khiển học **20 (**3) (2012) 129-140.
7. Cristina-Maria Vladarean - Extending object-oriented databases for fuzzy information modeling, S.C. WATERS Romania S.R.L, Romai J. **2** (1) (2006) 225-237.
8. Ma Z. M. - Advances in Fuzzy Object-Oriented Databases: Modeling and Applications, Idea Group Publishing, 2004.
9. Nguyễn Cát Hồ, Lê Xuân Vinh, Nguyễn Công Hào - Thống nhất dữ liệu và xây dựng quan hệ tương tự trong cơ sở dữ liệu ngôn ngữ bằng đại số gia tử, Tạp chí Tin học và Điều khiển học **25** (4) (2009) 314-332.
10. Nguyễn Kim Anh - Chuẩn hóa sơ đồ cơ sở dữ liệu hướng đối tượng, Tạp chí Tin học và Điều khiển học **19** (2) (2003) 125-130.
11. Raju H. V. S. N, Mazumdar A. K. - Fuzzy function dependencies and lossless join decomposition of fuzzy relational database systems, ACM Transaction of Database Systems **13** (2) (1998) 129-166.
12. Sadok Ben Yahia, Habib Ounalli, Ali Jaoua - An Extension of Classical Functional Dependency, Information Science **119** (3-4) (1999) 219-234.

ABSTRACT

DATA DEPENDENCIES IN FUZZY OBJECT ORIENTED DATABASES MODEL BASED ON HEDGE ALGEBRA

Nguyen Cong Hao[1, *], Truong Thi My Le[2]

[1]*Hue University Information Technology Center, 2 Le Loi, Hue City*

[2] *Quang Trung University, 130 Tran Hung Dao, Quy Nhon City*

[*]Email: *nchao@hueuni.edu.vn*

The single most important concept in database schemas design is that of a functional dependency. A functional dependency describes the relationship among attributes and is one of the key concepts used in the normalization. On the basic of fuzzy object-oriented databases model with hedge algebra has been proposed, in this paper, we introduce the study of the form of data dependencies, including fuzzy functional dependency in object class, fully fuzzy functional dependency, partial fuzzy functional dependency, transitive fuzzy functional dependency and fuzzy functional dependency with quantifier in natural languages. Based on the concepts of fuzzy functional dependencies, the paper also propose the normal form of fuzzy object-class in the model.

Keywords: fuzzy object-oriented databases, hedge algebra, fuzzy functional dependency, normal form of fuzzy object-class.