Tạp chí Tin học và Điều khiển học, T.23, S.2 (2007), 121–

MỞ RỘNG PHỤ THUỘC HÀM VÀ PHỤ THUỘC ĐA TRỊ

HỒ THUẦN[1], HOÀNG THỊ LAN GIAO[2]

[1] *Viện Công nghệ thông tin, Viện Khoa học và Công nghệ Việt Nam*
[2] *Khoa Công Nghệ Thông Tin, Đại học Khoa học Huế*

Abstract. The aim of the paper is to give a generalization of functional and multivalued dependencies in an information system. The definitions are established under the assumption that there are some similarity relations between values of attributes. By using the so-called *generalized dependency matrices* we develop a necessary and sufficient condition for an extension dependency to be hold. Besides, some computational examples are given for illustration too.

Tóm tắt. Bài báo đã xây dựng các định nghĩa mở rộng về phụ thuộc hàm và phụ thuộc đa trịtrong hệ thống thông tin. Các định nghĩa mới này được thiết lập trên cơ sở thừa nhận việc tồn tại các quan hệ tương tự giữa các giá trị của những thuộc tính. Bằng cách sửdụng các *ma trận phụ thuộc mở rộng* chúng tôi đa ra được một điều kiện cần và đủđể một phụ thuộc mở rộng thoả mãn. Một số ví dụ minh hoạ cũng được trình bày trong bài báo.

1. MỞ ĐẦU

Cho $\mathcal{A} = (U, A)$ là một hệ thống thông tin với U là tập các đối tượng và A làtập các thuộc tính. Với mỗi $u \in U$ và $a \in A$ ta ký hiệu $u(a)$ là giá trị thuộc tính a của đối tượng u. Nếu $X \subseteq A$ là một tập các thuộc tính ta ký hiệu $u(X)$ là bộ gồm các giá trị $u(a)$ với $a \in X$. Vì vậy, nếu u và v là hai đối tượng thuộc U, ta sẽ nói $u(X) = v(X)$ nếu $u(a) = v(a)$ với mọi thuộc tính $a \in X$.

Nhắc lại rằng, nếu X, Y là các tập thuộc tính, thì Y được gọi là phụ thuộc hàm vào X trên U và ký hiệu $X \to Y$ nếu

$$\forall u, v \in U : u(X) = v(X) \Rightarrow u(Y) = v(Y). \tag{1}$$

Hơn nữa, bằng cách đặt $Z = A \setminus (X \cup Y)$, ta nói Y là phụ thuộc đa trị vào X trên U và ký hiệu $X \to\!\!\to Y$ nếu

$$\forall u, v \in U : u(X) = v(X) \Rightarrow \exists t \in U : \begin{cases} t(X) = u(X), \\ t(Y) = u(Y), \\ t(Z) = v(Z). \end{cases} \tag{2}$$

Tuy nhiên, trong thực tế, có những phụ thuộc hàm và phụ thuộc đa trị mà các đẳng thức trong (1) và (2) không đòi hỏi thực sự nghiêm ngặt nh vậy. Chẳng hạn, cho bảng dữ liệu sinh viên, đào tạo theo niên chế, với ba thuộc tính *lớp, họtên, môn học*. Dễ thấy thuộc tính *họ tên* là phụ thuộc đa trị vào thuộc tính *lớp*, nghĩa là mọi sinh viên trong cùng một lớp sẽ phải học các môn nh nhau. Bây giờ, nếu nhà trường đa ra một số học phần tự chọn, và với mỗi học phần nh vậy các sinh viên trong một lớp có thể học các môn khác nhau (nhng được xem là Tương đương), thì chúng ta vẫn có lý do để nói rằng thuộc tính *họ tên* phụ thuộc đa trị vào thuộc tính *lớp*, mặc dù lúc này điều kiện (2) không còn đúng nữa. Rõràng là trong trường hợp này các định nghĩa nh trên không còn phù hợp. Mặt khác, trong thực tế không hiếm khi những dữ liệu thu nhận được không còn chính xác nh vốn có. Điều này chắc chắn cũng làm cho chúng ta không phát hiện hết mọi phụ thuộc từ cơ sở dữ liệu.

Để góp phần phát hiện các phụ thuộc tiềm ẩn trong dữ liệu, trong bài này chúng tôi sẽ cốgắng đa ra một cách tiếp cận mở rộng của các khái niệm phụ thuộc hàm và phụ thuộc đa trị. Các khái niệm mở rộng này được thiết lập dựa trên một hàm đánh giá độ Tương tự giữa các giá trị trong bảng dữ liệu. Khi sự Tương tự giữa hai giá trị đạt đến một mức độ nhất định, chúng ta có thể xem hai giá trị này là "đồng nhất". Với cách tiếp cận này, các phụ thuộc thực ra là phụ thuộc xấp xỉ. Để kiểm chứng một phụ thuộc xấp xỉ nào đó, chúng tôi cũng sẽ sử dụng một ma trận Tương tự *ma trận phụ thuộc* được sử dụng trong [**?**].

2. CÁC KHÁI NIỆM

Cho hệ thống thông tin $\mathcal{A} = (U, A)$. Với mỗi $V \subseteq U$ và $X \subseteq A$, ta gọi miền giá trị của V trên X là tập hợp

$$\mathrm{dom}(V, X) := \{u(X) \mid u \in V\}.$$

Khi $V = U$ ta viết $\mathrm{dom}(X)$ thay cho $\mathrm{dom}(U, X)$ và hơn nữa, nếu $X = \{a\}$, ta chỉviết một cách đơn giản: $V_a = \mathrm{dom}(\{a\})$.

Để mở rộng các khái niệm phụ thuộc, trên các tập giá trị V_a, ngoài quan hệ bằng nhau thông thờng ta giả định là tồn tại một hàm Tương tự, phản ánh độ gần nhau giữa các giá trị. Một cách chính xác, một ánh xạ $s : V_a \times V_a \to [0, 1]$ được gọi là hàm tương tự trên tập V_a nếu hai điều kiện sau thỏa mãn:

1. $s(a_1, a_2) = s(a_2, a_1)$, với mọi $a_1, a_2 \in V_a$,
2. $s(a_1, a_2) = 1 \Leftrightarrow a_1 = a_2$.

$s(a_1, a_2)$ được gọi là mức tương tự giữa hai giá trị a_1 và a_2. Cho $\alpha \in [0, 1]$, hai giá trị a_1 và a_2 được gọi là $\alpha-$tương tự, và ký hiệu $a_1 =_\alpha a_2$, nếu $s(a_1, a_2) \geqslant \alpha$. Rõ ràng khi hàm s chỉ nhận hai giá trị 0 và 1, thì, với mọi $\alpha > 0$, $a_1 =_\alpha a_2$ khi và chỉ khi $a_1 = a_2$.

Cho tập thuộc tính $B = \{b_1, b_2, \cdots, b_m\} \subseteq A$ và $\beta, \gamma \in \mathrm{dom}(B)$. Khi đó β và γ có thể xem là hai dãy $(\beta_i)_i$ và$(\gamma_i)_i$, với $\beta_i, \gamma_i \in \mathrm{dom}(b_i)$, $1 \leqslant i \leqslant m$. Độ tương tự giữa β và γ được định nghĩa là giá trị:

$$S(\beta, \gamma) = \min\{s(\beta_i, \gamma_i) \mid 1 \leqslant i \leqslant m\}. \quad (3)$$

Bây giờ, giả sử $\gamma \in \mathrm{dom}(B)$ và $D \subseteq \mathrm{dom}(B)$, ta gọi độ thuộc của γ vào D là độ tương tự lớn nhất giữa γ với các giá trị trong D. Cụthể, giá trị này được xác định bởi:

$$b(\gamma, D) = \max_{\beta \in D} S(\gamma, \beta).$$

Một cách tự nhiên, ta tiếp tục dùng ký hiệu $\beta =_\alpha \gamma$ nếu $S(\beta, \gamma) \geqslant \alpha$ và nói rằng β và γ là $\alpha-$tương tự. Cũng vậy, ta nói γ thuộc vào tập D với mức α, và ký hiệu $\gamma \in_\alpha D$, nếu

$b(\gamma, D) \geqslant \alpha$. Mệnh đề dưới đây cho ta một số tính chất cơ bản của các khái niệm này màviệc chứng minh có thể suy ra trực tiếp từ định nghĩa.

Mệnh đề 2.1. *Cho* $B \subseteq A$, $D \subseteq \mathrm{dom}(B)$, $\beta, \gamma \in \mathrm{dom}(B)$, *ta có*

1. $S(\beta, \gamma) = S(\gamma, \beta)$; $S(\beta, \gamma) = 1 \Leftrightarrow \beta = \gamma$.
2. $0 \leqslant b(\gamma, D) \leqslant 1$; $b(\gamma, D) = 1 \Leftrightarrow \gamma \in D$.
3. *Nếu* $D \subseteq D' \subseteq \mathrm{dom}(B)$, *thì* $b(\gamma, D) \leqslant b(\gamma, D')$.
4. $\gamma \in_\alpha D \Leftrightarrow \exists \beta \in D, \gamma =_\alpha \beta$.

Ví dụ 2.1. Xét hệ thống với ba thuộc tính: a (tên lập trình viên), b (trình độ chuyên môn), c (ngôn ngữ lập trình sử dụng) được cho trong Bảng 1.

Bảng 1

a	b	c
A	Bc	PASCAL
A	Bc	FORTRAN
A	Bc	COBOL
A	Dip	PASCAL
A	Dip	C
A	Dip	FORTRAN
A	Ms	COBOL
A	Ms	PASCAL
B	Bc	C
B	Bc	PASCAL

Giả sử hàm tương tự giữa những giá trị trong từng thuộc tính được cho bởi các bảng sau

Bảng 2. Hàm tương tự trên V_b

b	Bc	Dip	Ms
Bc	1	0.6	0.8
Dip	0.6	1	0.3
Ms	0.8	0.3	1

Bảng 3. Hàm tương tự trên V_c

c	FORTRAN	COBOL	PASCAL	C
FORTRAN	1	0.9	0.7	0.8
COBOL	0.9	1	0.7	0.6
PASCAL	0.7	0.7	1	0.8
C	0.8	0.6	0.6	1

Đặt $B = \{b, c\}$, $\beta = (Bc, FOTRAN), \gamma = (Ms, COBOL) \in \mathrm{dom}(B)$, ta có

$$S(\beta, \gamma) = \min\{s(Bc, Ms), s(FORTRAN, COBOL)\} = \min\{0.8, 0.9\} = 0.8.$$

Mặt khác, với $D = \{Bc, Dip\} \subseteq \mathrm{dom}(b)$ và $Ms \in \mathrm{dom}(b)$, ta có

$$b(Ms, D) = \max\{s(Ms, Bc), s(Ms, Dip)\} = \max\{0.8, 0.3\} = 0.8.$$

Như vậy, $\beta =_{0.8} \gamma$ và $Ms \in_{0.8} D$.

3. PHỤ THUỘC MỞ RỘNG VÀ CÁC TÍNH CHẤT

Dựa trên quan hệ $\alpha-$tương tự trên các tập giá trị, chúng ta sẽ đa ra các khái niệm phụ thuộc hàm và phụ thuộc đa trị mở rộng. Một cách chính xác, chúng ta có các định nghĩa sau.

Định nghĩa 3.1. Cho $X, Y \subseteq A$ và $\alpha \in [0,1]$. Ta nói Y $\alpha-$phụ thuộc hàm vào X trên U và ký hiệu $X \to_\alpha Y$ nếu

$$\forall u, v \in U : u(X) = v(X) \Rightarrow u(Y) =_\alpha v(Y).$$

Khi $\alpha = 1$ ta nhận được định nghĩa phụ thuộc hàm nguyên thủy.

Định nghĩa 3.2. Cho $X, Y \subseteq A$ và $\alpha \in [0,1]$. Đặt $Z = A \setminus (X \cup Y)$. Ta nói Y là $\alpha-$phụ thuộc đa trị vào X trên U, và ký hiệu $X \twoheadrightarrow_\alpha Y$, nếu với mọi cặp đối tượng $u, v \in U$ sao cho $u(X) = v(X) = x$, tồn tại đối tượng $t \in U$ sao cho $t(X) = x$, đồng thời thỏa mãn một trong hai điều kiện sau:

1. $t(Y) = u(Y)$ và $t(Z) =_\alpha v(Z)$,
2. $t(Y) =_\alpha u(Y)$ và $t(Z) = v(Z)$.

Rõ ràng, khi $\alpha = 1$ hai điều kiện trên tương đương và cũng tương đương với (2), nên ta cũng nhận được khái niệm phụ thuộc đa trị nguyên thuỷ. Từ các định nghĩa mở rộng trên dễ kiểm tra được rằng, nếu $0 \leqslant \beta \leqslant \alpha \leqslant 1$, thì $X \to_\alpha Y$ ($X \twoheadrightarrow_\alpha Y$) kéo theo $X \to_\beta Y$ ($X \twoheadrightarrow_\beta Y$). Ngoài ra, một số tính chất của phụ thuộc hàm và phụ thuộc đa trị vẫn còn đúng đối với các phụ thuộc mở rộng. Điều đó được khẳng định trong mệnh đề dưới đây

Mệnh đề 3.1. *Cho $X, Y, Z \subseteq A$, $\alpha \in [0,1]$. Khi đó*

1. *Nếu $Y \subseteq X$ thì $X \to_\alpha Y$.*
2. *Nếu $X \to_\alpha Y$ thì $X \cup Z \to_\alpha Y \cup Z$.*
3. *Nếu $X \twoheadrightarrow_\alpha Y$ thì $X \twoheadrightarrow_\alpha A \setminus (X \cup Y)$.*
4. *Nếu $X \to_\alpha Y$ thì $X \twoheadrightarrow_\alpha Y$.*

Chứng minh.

1) Hiển nhiên đúng vì ta đã biết, nếu $Y \subseteq X$ thì $X \to_1 Y$.

2) Với mọi cặp đối tượng $u, v \in U$ nếu $u(X \cup Z) = v(X \cup Z)$ thì $u(Z) = v(Z)$ và$u(X) = v(X)$. Vì $X \to_\alpha Y$ nên $u(Y) =_\alpha v(Y)$. Do đó $u(Y \cup Z) =_\alpha v(Y \cup Z)$. Vậy $X \cup Z \to_\alpha Y \cup Z$.

3) Không mất tính tổng quát, giả sử $X \cap Y = \emptyset$. Khi đó, đặt $Z = A \setminus (X \cup Y)$, thì $Y = A \setminus (X \cup Z)$. Từ $X \twoheadrightarrow_\alpha Y$ suy ra với mọi cặp đối tượng $u, v \in U$ mà $u(X) = v(X) = x$ thì tồn tại đối tượng $t \in U$ sao cho $t(X) = x$ và $t(Y) = u(Y), t(Z) =_\alpha v(Z)$ hoặc $t(Y) =_\alpha u(Y), t(Z) = v(Z)$. Do đó $X \twoheadrightarrow_\alpha Z$.

4) Đặt $Z = A \setminus (X \cup Y)$. Do $X \to_\alpha Y$ nên với mọi cặp đối tượng $u, v \in U$ mà $u(X) = v(X) = x$ ta có $v(Y) =_\alpha u(Y)$. Bằng cách chọn $t = v$ ta nhận được $t(X) = x$, $t(Z) = v(Z)$ và $t(Y) =_\alpha u(Y)$. Vậy $X \twoheadrightarrow_\alpha Y$. ■

Ví dụ 3.1. Xét hệ thống $\mathcal{A} = (U, \{X, Y, Z\})$ được cho trong Bảng 4, các hàm tương tự trên V_Y và V_Z được cho trên Bảng 5.

Khi đó, dễ thấy $X \twoheadrightarrow_{0.8} Y$. Nhng $X \not\twoheadrightarrow_{0.9} Y$ vì có hai đối tượng t_4 vàt_5 cùng bằng x_1 trên thuộc tính X, mặt khác $t_4(Y) = \beta_1$ và$t_5(Z) = \gamma_3$, nhng không có đối tượng v nào mà $v(X) = x_1$ đồng thời thoả mãn ($v(Y) = \beta_1$ và $v(Z) =_{0.9} \gamma_3$) hoặc ($v(Y) =_{0.9} \beta_1$ và $v(Z) = \gamma_3$).

Bảng 4

U	X	Y	Z
t_1	x_1	β_1	γ_1
t_2	x_1	β_2	γ_1
t_3	x_1	β_3	γ_2
t_4	x_1	β_1	γ_2
t_5	x_1	β_3	γ_3
t_6	x_1	β_2	γ_3
t_7	x_2	β_1	γ_1
t_8	x_2	β_1	γ_2

Bảng 5. Các hàm tương tự trên V_Y và V_Z

Y	β_1	β_2	β_3
β_1	1	0.5	0.7
β_2	05	1	0.9
β_3	0.7	0.9	1

Z	γ_1	γ_2	γ_3
γ_1	1	0.6	0.7
γ_2	06	1	0.8
γ_3	0.7	0.8	1

4. KIỂM TRA β - PHỤ THUỘC ĐA TRỊ

4.1. Điều kiện tồn tại - phụ thuộc đa trị

Trong một bài báo trớc [?], để nghiên cứu phụ thuộc đa trị, các tác giả đã thiết lập *ma trận phụ thuộc* dựa vào phân hoạch trên các giá trị thuộc tính và đã chứng minh được rằng, $X \twoheadrightarrow Y$ đúng khi và chỉ khi *ma trận phụ thuộc là dầy đặc*, tức làmọi phần tử của ma trận đều có giá trị 1. Trong trường hợp ma trận phụ thuộc là *gần đặc* (tức là chứa phần lớn các số 1), thì ta cũng nhận được một phụthuộc đa trị xấp xỉ(tức là bỏ đi một số ít bộ nào đó của bảng dữ liệu thì nhận được phụ thuộc đúng). Trên cơ sở các kết quả này, một thuật toán kiểm chứng phụ thuộc và phụ thuộc xấp xỉ dựa vào ma trận phụ thuộc cũng đã được thiết lập. Phát triển ý tởng đó, ở đây chúng ta sẽ xây dựng một ma trận có vai trò tương tự trong việc xác định $\alpha-$phụthuộc đa trị.

Trên U ta xác định quan hệ $\mathrm{IND}(X)$ xác định bởi

$$u\,\mathrm{IND}(X)v \Leftrightarrow u(X) = v(X); \quad u, v \in U.$$

Dễ kiểm chứng được rằng $\mathrm{IND}(X)$ là một quan hệ tương đương trên U. Ta ký hiệu họ các lớp tương đương của U theo quan hệ này bởi $[X] = \{X_1, \cdots, X_m\}$. Rõ ràng, $Y \twoheadrightarrow_\alpha Z$ đúng trên U khi và chỉ khi $Y \twoheadrightarrow_\alpha Z$ đúng trên mọi X_i. Do đó, ở đây ta chỉ hạn chế việc kiểm tra phụ thuộc trên mỗi X_i cố định. Ký hiệu $Z = A \setminus (X \cup Y)$. Giả sử $\mathrm{dom}(X_i, Y) = \{\beta_1, \beta_2, \cdots, \beta_p\}$ và $\mathrm{dom}(X_i, Z) = \{\gamma_1, \gamma_2, \cdots, \gamma_q\}$. Với mỗi β_j, γ_k ta ký hiệu

$$E_j := \{t(Z) \mid t \in X_i;\ t(Y) = \beta_j\} \subseteq \mathrm{dom}(Z);$$
$$F_k := \{t(Y) \mid t \in X_i;\ t(Z) = \gamma_k\} \subseteq \mathrm{dom}(Y).$$

Ta gọi *ma trận phụ thuộc mở rộng*, tương ứng với lớp X_i, là $D^i = (d_{jk})_{p\times q}$, với các thành phần d_{jk} được xác định bởi:

$$d_{jk} := \max\{b(\beta_j, F_k), b(\gamma_k, E_j)\}.$$

Ma trận D^i được gọi là $\alpha-$dầy đặc nếu với mọi j, k ta đều có $d_{jk} \geqslant \alpha$, hay, một cách tương đương: hoặc $\beta_j \in_\alpha F_k$ hoặc $\gamma_k \in_\alpha E_j$.

tương tự nh phụ thuộc đa trị nguyên thủy, $\alpha-$phụ thuộc đa trị cũng có thể được đặc trưng hoàn toàn bằng họ các ma trận phụ thuộc mở rộng D^i. Điều đó được khẳng định trong định lý sau.

Định lý 4.1. *Y là $\alpha-$phụ thuộc đa trị vào X khi và chỉ khi D^i là$\alpha-$dầy đặc, với mọi $1 \leqslant i \leqslant m$.*

Chứng minh. Giả sử $X \twoheadrightarrow_\alpha Y$. Chúng ta sẽ chứng minh mọi D^i đều là ma trận $\alpha-$dầy đặc. Thật vậy, với mọi $1 \leqslant j \leqslant p$ và $1 \leqslant k \leqslant q$, tồn tại hai đối tượng $u, v \in X_i$ sao cho $u(Y) = \beta_j$ và $v(Z) = \gamma_k$. Vì u vàv cùng thuộc lớp X_i nên $u(X) = v(X)$. Theo định nghĩa của $\alpha-$phụ thuộc đa trị, tồn tại bộ $t \in X_i$ thoả mãn một trong hai điều kiện sau

1. $t(Y) = u(Y) = \beta_j$ và $t(Z) =_\alpha v(Z) = \gamma_k$,
2. $t(Y) =_\alpha u(Y) = \beta_j$ và $t(Z) = v(Z) = \gamma_k$.

Nếu trường hợp a) xãy ra thì $t(Z) \in E_j$ và $\gamma_k =_\alpha t(Z)$. Do đó, $b(\gamma_k, E_j) \geqslant \alpha$. Tương tự, nếu b) xãy ra thì $b(\beta_j, F_k) \geqslant \alpha$. Cả hai trường hợp đó đều dẫn đến $d_{jk} \geqslant \alpha$. Vì điều này đúng với mọi d_{jk} nên D^i là ma trận $\alpha-$dầy đặc.

Ngược lại, giả sử mọi D^i đều là ma trận $\alpha-$ dầy đặc. Cho hai đối tượng tuỳ ý$u, v \in U$ thoả mãn $u(X) = v(X)$. Lúc đó, u và v phải thuộc cùng một lớp tương đương X_i nào đó. Đặt $\beta_j = u(Y)$ và $\gamma_k = v(Z)$. Do $d_{jk} \geqslant \alpha$ nên ta có

1. hoặc $b(\beta_j, F_k) \geqslant \alpha$,
2. hoặc $b(\gamma_k, E_j) \geqslant \alpha$.

Nếu i) đúng thì tồn tại $t \in X_i$ sao cho $t(Z) = \gamma_k = v(Z)$ và $t(Y) =_\alpha \beta_j = u(Y)$, còn nếu ii) đúng thì tồn tại $t \in X_i$ sao cho $t(Y) = \beta_j = u(Y)$ và $t(Z) =_\alpha \gamma_k = v(Z)$. Trong cả hai trường hợp, t đều thoả mãn điều kiện của Định nghĩa **??**. Vậy $X \twoheadrightarrow_\alpha Y$ và định lý đã được chứng minh. ■

Ví dụ 4.1. Xét lại hệ thống cho trong Ví dụ **??**. Họ các lớp tương đương theo quan hệIND(X) là $[X] = \{X_1, X_2\}$, với $X_1 = \{t_1, t_2, t_3, t_4, t_5, t_6\}$ và$X_2 = \{t_7, t_8\}$.
Trên lớp X_1 ta có $\mathrm{dom}(X_1, Y) = \{\beta_1, \beta_2, \beta_3\}$, $\mathrm{dom}(X_1, Z) = \{\gamma_1, \gamma_2, \gamma_3\}$ và

$E_1 = \{t(Z) \mid t \in X_1,\ t(Y) = \beta_1\} = \{\gamma_1, \gamma_2\}$;
$E_2 = \{t(Z) \mid t \in X_1,\ t(Y) = \beta_2\} = \{\gamma_1, \gamma_3\}$;
$E_3 = \{t(Z) \mid t \in X_1,\ t(Y) = \beta_3\} = \{\gamma_2, \gamma_3\}$;
$F_1 = \{t(Y) \mid t \in X_1,\ t(Z) = \gamma_1\} = \{\beta_1, \beta_2\}$;
$F_2 = \{t(Y) \mid t \in X_1,\ t(Z) = \gamma_2\} = \{\beta_1, \beta_3\}$;
$F_3 = \{t(Y) \mid t \in X_1,\ t(Z) = \gamma_3\} = \{\beta_2, \beta_3\}$.

Từ đó các phần tử của ma trận D^1 được xác định bởi:

$d_{11} = \max\{b(\beta_1, F_1),\ b(\gamma_1, E_1)\} = \max\{1; 1\} = 1;$
$d_{12} = \max\{b(\beta_1, F_2),\ b(\gamma_2, E_1)\} = \max\{1; 1\} = 1;$
$d_{13} = \max\{b(\beta_1, F_3),\ b(\gamma_3, E_1)\} = \max\{0.7; 0.8\} = 0.8;$
$d_{21} = \max\{b(\beta_2, F_1),\ b(\gamma_1, E_2)\} = \max\{1; 1\} = 1;$
$d_{22} = \max\{b(\beta_2, F_2),\ b(\gamma_2, E_2)\} = \max\{0.9; 0.8\} = 0.9;$
$d_{23} = \max\{b(\beta_2, F_3),\ b(\gamma_3, E_2)\} = \max\{1; 1\} = 1;$
$d_{31} = \max\{b(\beta_3, F_1),\ b(\gamma_1, E_3)\} = \max\{0.9; 0.7\} = 0.9;$
$d_{32} = \max\{b(\beta_3, F_2),\ b(\gamma_2, E_3)\} = \max\{1; 1\} = 1;$
$d_{33} = \max\{b(\beta_3, F_3),\ b(\gamma_3, E_3)\} = \max\{1; 1\} = 1.$

Tương tự, trên lớp X_2 ta có $\mathrm{dom}(X_2, Y) = \{\beta_1\}$, $\mathrm{dom}(X_2, Z) = \{\gamma_1, \gamma_2\}$ và bằng tính toán đơn giản ta thu được các phần tử của ma trận D^2 là$d_{11} = d_{12} = 1$. Tóm lại, ta được

$$D^1 = \begin{pmatrix} 1 & 1 & 0.8 \\ 1 & 0.9 & 1 \\ 0.9 & 1 & 1 \end{pmatrix}, \qquad D^2 = \begin{pmatrix} 1 & 1 \end{pmatrix}.$$

Rõ ràng với $\alpha \leqslant 0.8$ thì cả hai ma trận D^1 và D^2 đều $\alpha-$dầy đặc. Do đó $X \twoheadrightarrow_\alpha Y$, với mọi $\alpha \leqslant 0.8$. Trong khi đó, nếu $\alpha > 0.8$ thì D^1 không $\alpha-$dầy đặc nên $X \not\twoheadrightarrow_\alpha Y$.

TÀI LIỆU THAM KHẢO

[1] David Maier, *The Theory of Relational Databases* (Computer Science Press, Rockville, Maryland, 1983).

[2] Jeffrey D. Ullman, *Principles of Database and Knowledge-Base Systems, Volume I* (Computer Science Press, Rockville, Maryland, 1988).

[3] Hoàng Thị Lan Giao, Hồ Thuần, "Khám phá phụ thuộc đa trị dựa vào ma trận phụ thuộc", *Tin học và điều khiển* (2006).

[4] Hồ Thuần, "Contribution to the theory of relational databases", *Tanulmanyok 184/1986, Studies 184/1986.* Budapest, Hungary.

[5] Iztok Savnik and Peter A. Flash, "Discovery of multivalued dependencies from ralations" *Intelligent data Analysis*, 4(3,4) 195 - 211, 2000

[6] Mannila, H., Toivonen, H. and Verkamo, A.I. , "Discovery of frequency episodes in event sequences" *Data Mining and Knowledge Discovery* (1997), I, 259 - 289.

[7] Shrabonti Ghosh, Ranjit Biswas, S.S. Alam, "Reduction of the Decision Table: A Rouhg Approach", *International Journal of Intelligent Systems*, Vol 19, 2004.

[8] Yka Huhtala, Juha Karkkainen, Pasi Porkka and Hannu Toivonen, "Tane: An Efficient Algorithm for Discovering Functional and Approximate Dependencies" *The Computer Journal*, Vol 42, No 2. 1999.

[9] Yka Huhtala, Juha Karkkainen, Pasi Porkka and Hannu Toivonen, "Efficient Discovery of Functional and Approximate Dependencies using Partitions" *In Proc, 14th Int. Conf. on Data Engineering, IFFE Computer Society Press* ('98).

Nhận bài ngày 5 - 6 - 2006