

Quan hệ giữa phụ thuộc hàm nói lỏng và phụ thuộc Boole dương tổng quát

Nguyễn Xuân Huy¹, Nguyễn Thị Vân², Trương Thị Thu Hà³

¹Viện Hàn lâm Khoa học và Công nghệ Việt Nam

²Trường Cao đẳng Công đồng Hà Nội

³Trường Đại học Kinh doanh và Công nghệ

Tác giả liên hệ: Nguyễn Thị Vân, van.cdcd@gmail.com

Ngày nhận bài: 08/05/2020, ngày sửa chữa: 26/06/2020

Định danh DOI: 10.32913/mic-ict-research-vn.vyyyy.nx.xyz

Tóm tắt: Mục đích chính của bài báo là khảo sát và đặc tả các biến thể khác nhau của phụ thuộc Boole dương tổng quát. Tìm ra một đặc tả cho phụ thuộc nói lỏng tổng quát và chỉ ra rằng phụ thuộc nói lỏng nói chung và phụ thuộc hàm nói lỏng nói riêng chỉ là những trường hợp riêng của phụ thuộc Boole dương tổng quát.

Kết quả này được dùng làm cơ sở để xây dựng mô hình tổng quát về các lớp phụ thuộc khác nhau trong khai thác tri thức bằng các công cụ AI và học sâu.

Từ khóa: phụ thuộc Boole dương tổng quát, phụ thuộc hàm nói lỏng.

Title: Relationship between Relaxed Functional Dependencies and Generalized Positive Boolean Dependencies

Abstract: The main purpose in this paper is studying and describing the variety of the generalized positive Boolean dependencies. A class of generalized relaxed dependencies is introduced and pointed out that this class and class of relaxed functional dependencies are partial of the generalized positive Boolean dependencies. The result can be used to construct a model of different dependencies for knowledge mining with the AI approach and deep learning.

Keywords: Generalized Boolean Dependencies, Relaxed Functional Dependencies.

I. ĐẶT VẤN ĐỀ

Phụ thuộc dữ liệu đóng vai trò quan trọng trong thiết kế các hệ cơ sở dữ liệu và khai thác tri thức. Việc phát hiện các loại phụ thuộc khác nhau trong dữ liệu là một vấn đề nghiên cứu thú vị, có ý nghĩa và cũng là một trong những mục tiêu của lĩnh vực đảm bảo ngữ nghĩa, tính nhất quán và khai thác tri thức trên văn bản.

Phụ thuộc dữ liệu hay ràng buộc dữ liệu là những mệnh đề qui định các tính chất của dữ liệu cần được tôn trọng trong quá trình xử lý và lưu trữ nhằm phản ánh đúng hiện thực khách quan. Sau công bố năm 1970 của Codd về phụ thuộc hàm trong cơ sở dữ liệu, các nhóm nghiên cứu đã phát triển và đề xuất nhiều loại phụ thuộc dữ liệu ngày càng tinh tế hơn nhằm mở rộng và nâng cao khả năng mô tả dữ liệu trong thế giới thực. Một trong những hướng nghiên cứu quan trọng là những phụ thuộc hàm nói lỏng $X \rightarrow Y$, trong đó X và Y là các tập thuộc tính, \rightarrow thể hiện sự phụ thuộc của tập thuộc tính Y vào tập thuộc tính X với các điều kiện nói lỏng khác nhau [5].

Ví dụ, trong giao thông ta thường có các phụ thuộc dữ liệu sau:

Bảng I
BẢNG GIÁ TAXI

Xe	Điểm đi	Điểm đến	Khoảng cách	Thành tiền
201	18 Hoàng Quốc Việt	20 Nguyễn Trãi	7.3 km	63.000
67	20 Hoàng Văn Thái	125 Nguyễn Phong Sắc	5.5 km	50.000
120	18 Hoàng Quốc Việt	18 Nguyễn Trãi	7 km	63.000
45	42 Vương Thừa Vũ	78 Trần Đăng Ninh	5.3 km	50.000

Theo phụ thuộc hàm kinh điển: *Khoảng cách* \rightarrow *Thành tiền* thể hiện ngữ nghĩa sau: Nếu hai chuyến taxi chở khách đi cùng một khoảng cách thì hai xe nhận được cùng một số tiền.

Phụ thuộc hàm nói lỏng: *Khoảng cách (0.2)* \rightarrow *Thành*

tiền (0) thể hiện ngữ nghĩa sau: Nếu hai chuyến taxi chở khách có khoảng cách chênh lệch không quá 0.3 km thì hai xe nhận được số tiền như nhau.

Năm 2016, nhóm tác giả Loredana Caruccio, Vincenzo Deufemia, Giuseppe Polese đã tập hợp và phân loại các phụ thuộc nới lỏng (relaxed functional dependencies) và phân tích các đặc trưng của một số dạng nới lỏng [5].

Năm 1992, nhóm tác giả Nguyễn Xuân Huy và Lê Thị Thanh đã đề xuất một loại phụ thuộc dữ liệu mới là phụ thuộc Boole dương tổng quát (PTBDTQ) và chứng minh phụ thuộc hàm và một số biến thể của phụ thuộc hàm như phụ thuộc yếu, phụ thuộc mạnh, phụ thuộc đối ngẫu là những trường hợp riêng của phụ thuộc Boole dương tổng quát [1], [3]. Tiếp đến trong các năm từ 2012 đến 2016 nhóm nghiên cứu về phụ thuộc Boole dương tổng quát đã tiếp tục chứng minh rằng các phụ thuộc sai khác [6] và các biến thể như các phụ thuộc hàm xấp xỉ [4] cũng là những trường hợp riêng của phụ thuộc Boole dương tổng quát.

Trong bài này, tiếp tục mạch nghiên cứu trên, sẽ chỉ ra rằng các phụ thuộc hàm nới lỏng cũng là những trường hợp riêng của phụ thuộc Boole dương tổng quát.

II. CÁC KHÁI NIỆM VÀ QUY ƯỚC

Cho $U = \{x_1, \dots, x_n\}$ là tập hữu hạn các biến Boole nhận giá trị trong tập trị logic $\mathbb{B} = \{0, 1\}$. Tập các công thức Boole (CTB), kí hiệu $L(U)$, bao gồm các biểu thức được xây dựng từ các biến trong U , các hằng 0/1 và các phép toán logic $\wedge, \vee, \neg, \rightarrow$

Mỗi vector 0/1, $v = (v_1, \dots, v_n)$ trong không gian \mathbb{B}^n được gọi là một phép gán trị. Khi đó với mỗi CTB $f = L(U)$, $f(v)$ là trị của công thức f đối với phép gán trị v . Kí hiệu e là phép gán trị đơn vị, $e = (1, 1, \dots, 1)$. Công thức $f = L(U)$ được gọi là công thức Boole dương (CTBD) nếu $f(e) = 1$.

Kí hiệu $P(U)$. Với mỗi công thức Boole $f \in L(U)$, kí hiệu $T_f = \{v \in \mathbb{B}^n \mid f(v) = 1\}$ là bảng chân lí của f . Mỗi tập công thức $F \subseteq L(U)$ được hiểu là một hội logic của các công thức thành phần, $F = \wedge f \mid f \in F$. Khi đó, $T_F = \bigcap T_f \mid f \in F$ là bảng chân lí của tập công thức F . Với hai CTBD f và g trên U , ta đã biết CTBD g được suy dẫn logic từ CTBD f và được ký hiệu là $f \rightarrow g$ khi và chỉ khi $T_f \subseteq T_g$. Tương tự, nếu F là tập các CTBD trên U thì $F \rightarrow g$ khi và chỉ khi $T_F \subseteq T_g$. Hai tập CTBD F và G trên U là tương đương nhau khi và chỉ khi $T_F = T_G$.

Theo truyền thống của lý thuyết cơ sở dữ liệu, ta chấp nhận các ký hiệu sau đây:

Hợp của hai tập X và Y được viết XY ; giao của hai tập X và Y được viết $X \cap Y$; phép trừ hai tập X và Y được ký hiệu là $X - Y$.

Phép hội hai công thức logic X và Y được kí hiệu là $X \wedge Y$ hoặc XY hoặc X và Y (trong trường hợp cần chỉ rõ dấu phép toán) và được gọi là tích (logic); phép tuyển hai công thức logic X và Y được ký hiệu $X \vee Y$ hoặc $X + Y$ và được gọi là tổng (logic); dấu ' thay cho phép phủ định \neg . Ví dụ, biểu thức $ab + c + bd$ là một biểu diễn tương đương với công thức logic truyền thống $a \wedge b \vee c \vee b \wedge (\neg d)$.

Ví dụ, cho tập $U = \{a, b, c\}$ khi đó công thức $f: ab \rightarrow c$ là một công thức Boole dương vì $f(1, 1, 1) = 1$.

Bảng II
BẢNG CHÂN LÝ T_f

a	b	c	ab	$f=ab \rightarrow c$
0	0	0	0	1
0	0	1	0	1
0	1	0	0	1
0	1	1	0	1
1	0	0	0	1
1	0	1	0	1
1	1	0	1	0
1	1	1	1	1

Nếu $U = \{x_1, x_2, \dots, x_n\}$ là tập các thuộc tính và v là một bộ trong quan hệ r trên U , $x \in U$ thì $v.x$ được ký hiệu là trị của thuộc tính x trong bộ v . Nếu $X \subseteq U$ thì ta định nghĩa $v.X = \{v.x \mid x \in X\}$. Với tập thuộc tính U cho trước, quan hệ trên tập thuộc tính U được kí hiệu là r , khi cần chỉ rõ tập thuộc tính ta có thể sử dụng kí hiệu $r(U)$.

Các qui ước và kí pháp truyền thống khác được giới thiệu trong các tài liệu kinh điển về cơ sở dữ liệu và được trích dẫn trong [1] và [2].

III. PHỤ THUỘC BOOLE DƯƠNG TỔNG QUÁT

Cho $U = \{x_1, \dots, x_N\}$ là tập hữu hạn các biến Boole nhận giá trị trong tập trị logic $\mathbb{B} = \{0, 1\}$.

Ta quy ước mỗi miền trị d_x của thuộc tính x trong U có chứa ít nhất hai phần tử. Với mỗi miền trị d_x , xét ánh xạ $\alpha_x: d_x^2 \rightarrow \mathbb{B}$ thoả các tiên đề sau [2][3]:

$$\forall a, b \in d_x$$

$$A1) \text{ Tiên đề phản xạ } \alpha_x(a, a) = 1$$

$$A2) \text{ Tiên đề đối xứng } \alpha_x(a, b) = \alpha_x(b, a)$$

$$A3) \text{ Tiên đề bộ phận } \exists c \in d_x: \alpha_x(a, c) = 0$$

α_x chính là quan hệ (hai ngôi) bộ phận thực sự, thoả các tính chất phản xạ và đối xứng trên miền trị d_x . Việc xác định α_x được hiểu là thiết lập một phép sánh trị trên miền trị d_x cho thuộc tính x

Quan hệ bằng $=_x$ (tạm gọi là quan hệ đẳng thức) được định nghĩa: $\forall a, b \in d_x: =_x(a, b) = 1$, khi và chỉ khi $a = b$, là trường hợp riêng của phép sánh trị và được ngầm định trong rường hợp không định nghĩa tường minh phép sánh trị cho thuộc tính x .

Ta gọi lược đồ dữ liệu là một cặp $p = (U, F)$, trong đó U là tập thuộc tính với các miền trị tương ứng và các phép sánh trị trên mỗi miền trị, F là tập các phụ thuộc trên U [1][2][3].

Cho lược đồ $p = (U, F)$ và quan hệ r trên U . Với mỗi cặp bộ $u = (u_1, u_2, \dots, u_n), v = (v_1, v_2, \dots, v_n)$ trong r , ta đặt tương ứng một vector 0/1 $t = (t_1, t_2, \dots, t_n) \in \mathbb{B}_n$ và kí hiệu là $t = \alpha(u, v)$, trong đó thành phần $t.x$ ứng với thuộc tính x trong U chính là ảnh của ánh xạ $t.x = \alpha_x(u.x, v.x)$.

Khi đó mỗi quan hệ r sẽ được đặt tương ứng với tập các vector 0/1, $T_r = \{ \alpha(u, v) \mid u, v \in r \}$, và được gọi là *bảng trị* của quan hệ r trên LDBDTQ p [2][3].

Quan hệ r trên tập thuộc tính U thỏa PTBĐTQ f (tập PTBĐTQ F) và viết $r(f)$ ($r(F)$) nếu $T_r \subseteq T_f$ ($T_r \subseteq T_F$).

Mỗi công thức Boole dương f trong $P(U)$ với các phép sánh trị cho trước được gọi là một phụ thuộc Boole dương tổng quát (PTBĐTQ), lược đồ thu được trong trường hợp này được gọi là lược đồ với phụ thuộc boole dương tổng quát.

IV. PHỤ THUỘC HÀM NÓI LÔNG

Phụ thuộc hàm nói lông [5] được xây dựng dưới dạng thức chung: $f: X(\lambda) \rightarrow Y(\gamma); X, Y \subseteq U$ với các điều kiện nói lông λ và γ như sau:

Quan hệ $r(U)$ thỏa PTHNL $f: X(\lambda) \rightarrow Y(\gamma); X, Y \subseteq U$ nếu $T_r \subseteq T_f$

Nói lông các phép sánh trị trên một vài thuộc tính: ngoài sánh trị đẳng thức có thể xét các phép xấp xỉ theo độ đo, hoặc các phép so sánh $<, \leq, >, \geq, =, \neq$. Quan hệ r thỏa phụ thuộc hàm nói lông theo phép sánh trị $X(\rho) \rightarrow Y$ và được viết là $r(X(\rho) \rightarrow Y)$ khi và chỉ khi hai bộ bất kỳ trong r sai khác nhau không quá ngưỡng ρ trên X thì hai bộ đó cũng sai khác nhau không quá ngưỡng ρ trên Y .

Nói lông theo lực lượng:

$$r(X(\delta) \rightarrow Y) = \frac{\neq r(X \rightarrow Y)}{\neq r} \leq \delta$$

Trong đó $\neq S$ là số phần tử có trong tập S . Biểu thức trên cho biết quan hệ r thỏa phụ thuộc hàm nói lông theo lực lượng $X(\delta) \rightarrow Y$ và được viết là $r(X(\delta) \rightarrow Y)$ khi và chỉ khi tỷ lệ giữa số lượng các bộ còn lại thỏa phụ thuộc hàm truyền thống $X \rightarrow Y$ và các bộ của quan hệ r đạt trên ngưỡng δ . Nói cách khác, ta có thể xóa khỏi quan hệ r một số bộ với tỷ lệ tỷ lệ δ để quan hệ còn lại thỏa phụ thuộc hàm truyền thống $X \rightarrow Y$.

Định nghĩa chung về phụ thuộc hàm nói lông tổng quát như sau:

Cho U là tập thuộc tính, X và Y là hai tập thuộc tính trong U . PTH nói lông có dạng:

$$f(X(\gamma) \rightarrow Y(\vartheta), X, Y \subseteq U$$

Ta nói phụ thuộc hàm nói lông f thỏa trong quan hệ $r(U)$ nếu:

$$f(X(\gamma) \rightarrow Y(\vartheta), X, Y \subseteq U$$

Với mọi cặp bộ $u, v \in r$, nếu tần từ $\gamma(u.X, v.X)$ suy ra được tần từ $\vartheta(u.X, v.Y)$.

Phụ thuộc hàm nói lông được hiểu là phụ thuộc hàm với điều kiện kèm theo nhằm giảm nhẹ các điều kiện của phụ thuộc hàm chính thống. Các điều kiện giảm nhẹ được phát biểu thông qua các tần từ γ và ϑ

V. CÁC LỚP BIẾN THỂ CỦA PHỤ THUỘC BOOLE DƯƠNG TỔNG QUÁT

Các kết quả chủ yếu của mục này cho thấy, tùy thuộc vào các dạng của công thức Boole dương cụ thể và phép sánh trị alpha cụ thể ta có thể nhận được các phụ thuộc nói lông khác nhau ứng với các lược đồ dữ liệu tương ứng sau đây.

1. Lớp IE (Implication Formula và Equal Comparison)

Lớp IE là lớp các lược đồ phụ thuộc hàm kinh điển, được xây dựng trên cơ sở phép toán suy dẫn và phép sánh trị đẳng thức.

Cho tập các thuộc tính $U = (x_1, x_2, \dots, x_n), n \geq 1$. Giả sử $X, Y \subseteq U$. Một phụ thuộc thuộc lớp IE là biểu thức dạng $f: X \rightarrow Y$.

Dựa trên biểu thức logic X và Y ta phân biệt các dạng phụ thuộc hàm sau:

IE-1: Biểu thức logic có dạng $X \rightarrow Y$, ta có phụ thuộc hàm truyền thống: Cho LĐQH (U, F) Cho LĐQH (U, F) và một phụ thuộc hàm $f: X \rightarrow Y$ trên U . Ta nói quan hệ r thỏa phụ thuộc hàm f và ký hiệu $r(f)$, nếu hai bộ tùy ý trong r giống nhau trên X thì cũng giống nhau trên Y

IE-2: biểu thức logic có dạng $\forall X \rightarrow Y$ cho ta lược đồ phụ thuộc hàm mạnh: Cho quan hệ r trên tập thuộc tính U , quan hệ R thỏa phụ thuộc hàm mạnh $\forall X \rightarrow Y$ trên U nếu với hai bộ u, v bất kỳ trong r : u và v bằng nhau tại một thuộc tính nào đó trên X thì u và v cũng bằng nhau trên Y .

IE-3: Biểu thức logic có dạng $X \rightarrow \forall Y$, ta có lược đồ phụ thuộc hàm yếu: Cho quan hệ r trên tập thuộc tính U , quan hệ r thỏa phụ thuộc hàm yếu $X \rightarrow \forall Y$ trên U nếu với hai bộ u, v bất kỳ trong r : u và v bằng nhau trên X thì u và v cũng bằng nhau tại một thuộc tính nào đó trên Y .

IE-4: Biểu thức logic có dạng $\forall X \rightarrow \forall Y$, ta có lược đồ phụ thuộc hàm đối ngẫu: Cho quan hệ r trên tập thuộc tính U , quan hệ r thỏa phụ thuộc hàm đối ngẫu $\forall X \rightarrow \forall Y$ trên U nếu với hai bộ u, v bất kỳ trong r : u và v bằng nhau

tại một thuộc tính nào đó trên X thì u và v cũng bằng nhau tại một thuộc tính nào đó trên Y .

Bảng dưới đây tóm lược các đặc tả cho các lớp con IE1-4 của lớp IE

Dạng phụ thuộc	Tên gọi	Đặc điểm
$X \rightarrow Y$	phụ thuộc hàm	ràng buộc chặt
$\forall X \rightarrow Y$	phụ thuộc hàm mạnh	nới lỏng về trái
$X \rightarrow \forall Y$	phụ thuộc hàm yếu	nới lỏng về phải
$\forall X \rightarrow \forall Y$	phụ thuộc hàm đối ngẫu	nới lỏng
$X(\delta) \rightarrow Y$	phụ thuộc xấp xỉ	nới lỏng

Như vậy, các dạng phụ thuộc hàm IE-1, IE-2, IE-3, IE-4 là các phụ thuộc hàm nới lỏng và chúng là những trường hợp riêng của phụ thuộc Boole dương tổng quát.

2. Lớp LA (Logic Formula và Alpha Comparison)

Dựa theo điều kiện nới lỏng, ta chia lớp LA thành các lớp con sau đây:

LA-1: Lớp LA-1 là lớp các phụ thuộc được xây dựng trên cơ sở phép toán suy dẫn và phép sánh trị alpha.

Cho tập các thuộc tính $U = (x_1, x_2, \dots, x_n), n \geq 1$. Giả sử $X, Y \subseteq U$. Một phụ thuộc thuộc lớp LA-1 là biểu thức dạng $f: \alpha_X \rightarrow \alpha_Y$ với các phép sách trị α như đã trình bày ở phần III.

Ta nói quan hệ r thỏa LA-1: $\alpha_X \rightarrow \alpha_Y$ và viết $r(X(\alpha) \rightarrow Y(\alpha))$, nếu với hai bộ bất kỳ $u, v \in r$, thỏa các ràng buộc được đặc tả bởi α_X trên tập thuộc tính X , thì u và v cũng thỏa các ràng buộc được đặc tả bởi hàm sai khác α_Y trên tập thuộc tính Y :

$$r(X(\alpha) \rightarrow Y(\alpha)) \stackrel{def}{\iff} \forall u, v \in r: \alpha(u.X, v.X) \Rightarrow \alpha(u.Y, v.Y)$$

Trong lớp LA-1, ta có các phụ thuộc đại diện sau:

- Nếu biểu thức logic dạng $f: \alpha_X \rightarrow \alpha_Y$, ta có phụ thuộc sai khác [6], với α_X và α_Y là các hàm sai khác được định nghĩa :

Cho quan hệ r trên tập thuộc tính $U, a \in U$ và độ sai khác m_a . Hàm sai khác α_a trên thuộc tính α đặc tả ràng buộc của độ sai khác m_a : Với hai trị $x, y \in d_a$, ta định nghĩa $\alpha_a(x, y) = 1$ khi và chỉ khi $m_a(x, y)$ thỏa điều kiện cho α_a dưới dạng các biểu thức so sánh với các phép so sánh $=, \neq, <, \leq, >$ và \geq .

Cho tập thuộc tính $X \subseteq U$. Hàm sai khác \emptyset_X trên tập thuộc tính X là hội logic của các hàm sai khác trên mọi thuộc tính $a \in X: \emptyset_X = \bigwedge_{a \in X} \emptyset_a$

Trong [5], đã chỉ ra rằng phụ thuộc hàm sai khác là trường hợp riêng của phụ thuộc hàm nới lỏng, trong phần này chỉ ra rằng, phụ thuộc sai khác thuộc lớp LA-1.

- Nếu biểu thức logic có dạng $f: X(\delta) \rightarrow Y, 0 \leq \delta \leq 1$ ta có phụ thuộc nới lỏng theo lực lượng.

Ví dụ, khi khảo sát các khoa của một trường đại học người ta đánh giá tổng thể môi trường đào tạo của khoa theo ý kiến của các chuyên gia (thuộc tính A). Có 4 mức cho A lần lượt là a, b, c và d , trong đó a là mức tốt nhất. Thuộc tính B cho biết mức độ hài lòng của các sinh viên về trường đó. B gồm 4 mức là 1, 2, 3 và 4, trong đó 1 là mức cao nhất. Giả sử kết quả khảo sát được thể hiện trong quan hệ $r(A, B)$. Người ta muốn biết giữa môi trường học tập (A) và mức độ hài lòng (B) có đạt trên ngưỡng 0.6 hay không, tức là xác định $f: A(0.6) \rightarrow B$?

r	
A	B
a	1
b	2
a	1
a	2 *
b	2

Ta thấy, sau khi xóa bộ thứ tư khỏi quan hệ r thì quan hệ con còn lại, ký hiệu là r' sẽ thỏa phụ thuộc hàm truyền thống $A \rightarrow B$. Tỷ lệ thu được lúc này sẽ là

$$\frac{\# r'}{\# r} = \frac{4}{5} = 0.8 > 0.6$$

- Phép sánh trị theo khoảng

Miền trị d_A được phân hoạch thành k khoảng không giao nhau:

$$d_A = \bigcup_{i=1}^k [a_i; b_i]$$

$$[a_i; b_i] \cap [a_j; b_j] = \emptyset, 1 \leq i, j \leq k, i \neq j$$

$\forall x, y \in d_A: \alpha_A(x, y) = 1$ khi và chỉ khi x và y thuộc cùng một khoảng:

Ví dụ, điểm tổng kết của học sinh được chia thành 8 khoảng

- Điểm F : [0; 3.9]
- Điểm D : [4.0; 4.7]
- Điểm D+ : [4.8; 5.4]
- Điểm C : [5.5; 6.2]
- Điểm C+ : [6.3; 6.9]
- Điểm B : [7.0; 7.7]

Điểm B+ : [7.8; 8.4]

Điểm A : [8.5; 10]

Lớp LA-2: Lớp LA-2 là lớp các phụ thuộc được xây dựng trên cơ sở phép toán logic Boole và phép sánh trị đẳng thức. Đại diện cho lớp LA-2 là phụ thuộc Boole dương. [2][3]

Lớp LA-3: Lớp LA-3 là lớp các phụ thuộc được xây dựng trên cơ sở phép toán logic Boole và phép sánh trị alpha (mục 4.2). Đại diện cho lớp LA-3 là phụ thuộc Boole dương tổng quát (phần III), đây cũng là lớp phụ thuộc bao hàm các phụ thuộc logic trong cơ sở dữ liệu đã và đang được nghiên cứu bởi các nhóm tác giả trong và ngoài nước.

VI. KẾT LUẬN

Việc phân loại và đề xuất một mô hình chung cho các loại phụ thuộc dữ liệu là một trong những vấn đề đang được giới nghiên cứu dữ liệu lớn quan tâm. Các điều kiện cần và đủ để nhận biết các đặc trưng của một lớp phụ thuộc là cơ sở để phân loại và tạo ra mối liên hệ giữa các lớp phụ thuộc.

Bài báo mô tả mối quan hệ giữa phụ thuộc hàm nói lỏng và phụ thuộc Boole dương tổng quát trong cơ sở dữ liệu. Các kết quả chủ yếu bao gồm:

1. Xây dựng các lớp phụ thuộc trong cơ sở dữ liệu dựa vào hai đặc trưng cơ bản là công thức suy dẫn và phép sánh trị.

2. Chỉ ra mối quan hệ giữa phụ thuộc hàm nói lỏng với phụ thuộc Boole dương tổng quát. Từ các loại phụ thuộc hàm nói lỏng có thể tổng quát hóa thành phụ thuộc Boole dương tổng quát và ngược lại, từ lý thuyết phụ thuộc Boole dương tổng quát có thể thực tế hóa về các phụ thuộc hàm nói lỏng tùy thuộc từng trường hợp biến thể của yêu cầu thực tế.

TÀI LIỆU THAM KHẢO

- [1] Nguyễn Xuân Huy (2006), "*Các phụ thuộc logic trong cơ sở dữ liệu*", NXB Thống Kê.
- [2] Berman J., Blok W. J. (1988), "*Positive Boolean dependencies*" Inf. Processing Letters, 27, p.147 – 150.
- [3] Huy Nguyen Xuan, Thanh Le Thi (1992), "*Generalized Positive Boolean Dependencies*", J. Inf. Process. Cybern. EIK, vol. 28, p. 363 – 370.
- [4] Jalal Atoum (2009), "*Mining Approximate Functional Dependencies from Databases*", European Journal of Scientific Research, ISSN 1450 – 216X vol. 33 no. 2, p.338 – 346.
- [5] Loredana Caruccio, Vincenzo Deufemia, Giuseppe Polese (2016), "*Relaxed Functional Dependencies - A Survey of Approaches*", IEEE Transactions on Knowledge and Data Engineering, vol. 28, p. 147–165.

- [6] Song S. and Chen L. (2011), "*Differential Dependencies: Reasoning and Discovery*", *ACM Trans. Datab. Syst*, vol.9, no 4, Article 39.

SƠ LƯỢC VỀ TÁC GIẢ

Nguyễn Xuân Huy



Sinh năm 1944 tại Hải Phòng. Cử nhân Toán, Đại học Sư phạm Leningrad (Liên Xô) năm 1973. Tiến sỹ CNTT năm 1982, tiến sỹ khoa học CNTT năm 1990, Viện Hàn lâm Khoa học Liên Xô. Nguyên trưởng Phòng Cơ sở dữ liệu và Lập trình, Viện Công nghệ Thông tin, Viện Hàn lâm Khoa học và Công nghệ Việt Nam (1997-2009). Lĩnh vực nghiên cứu: Cơ sở dữ liệu và Công nghệ phần mềm. Email: nxhuy564@gmail.com.

Nguyễn Thị Vân



Sinh năm 1985 tại Hà Tĩnh. Cử nhân CNTT tại Trường Đại học Kinh doanh và Công nghệ Hà Nội năm 2011. Thạc sĩ ngành Khoa học máy tính tại Trường Đại học Công nghệ thông tin và Truyền thông năm 2014. Hiện đang công tác tại Khoa Công nghệ thông tin, Trường Cao đẳng Cộng đồng Hà Nội. Lĩnh vực nghiên cứu: Các phụ thuộc logic trong cơ

sở dữ liệu, mô hình dữ liệu và cơ sở dữ liệu. Email: van.cdcd@gmail.com

Trương Thị Thu Hà



Sinh năm 1979 tại Nghệ An. Cử nhân CNTT Trường Đại học Sư phạm Hà Nội năm 2000. Thạc sỹ CNTT Trường Đại học Công nghệ, Đại học Quốc Gia Hà Nội năm 2006. Tiến sỹ CNTT, Học viện Kỹ thuật Quân sự năm 2018. Hiện công tác tại Trường Đại học Kinh doanh và Công nghệ Hà Nội. Lĩnh vực nghiên cứu: Cơ sở dữ liệu, Công nghệ phần mềm. Email:

thuha.bh@gmail.com