

BỘ GIÁO DỤC VÀ ĐÀO TẠO

**VIỆN HÀN LÂM KHOA HỌC
VÀ CÔNG NGHỆ VIỆT NAM**

HỌC VIỆN KHOA HỌC VÀ CÔNG NGHỆ

.....***.....

NGUYỄN THỊ VÂN

**PHÁT TRIỂN PHỤ THUỘC BOOLE DƯƠNG XẤP XỈ
TRONG CƠ SỞ DỮ LIỆU QUAN HỆ**

**Ngành: Hệ thống thông tin
Mã số: 9 48 01 04**

TÓM TẮT LUẬN ÁN TIẾN SĨ HỆ THỐNG THÔNG TIN

Hà Nội - 2023

**Công trình này được hoàn thành tại: Học viện khoa học và Công nghệ-
Viện Hàn lâm Khoa học và Công nghệ Việt nam**

Người hướng dẫn khoa học học: PGS. TSKH. Nguyễn Xuân Huy

Phản biện 1:.....

Phản biện 2:.....

Phản biện 3:.....

Luận án sẽ được bảo vệ trước Hội đồng đánh giá luận án tiến sỹ cấp Học viện, họp tại Học viện khoa học và Công nghệ-Viện Hàn lâm Khoa học và Công nghệ Việt nam vào hồi giờ.....ngày.....tháng..... năm 2023

Có thể tìm luận án tại:

-Thư viện Học viện khoa học và Công nghệ

-Thư viện Quốc gia Việt Nam

MỞ ĐẦU

1. Tính cấp thiết của đề tài luận án

Tình hình nghiên cứu ngoài nước:

- Năm 1970, Codd [15], [16] giới thiệu mô hình cơ sở dữ liệu quan hệ và khái niệm phụ thuộc hàm (PTH) để phản ánh ngữ nghĩa của dữ liệu trong thế giới thực.

- Năm 1983 [7] J. Demetrovics và O. Gyepesi đề xuất phụ thuộc đối ngẫu, phụ thuộc mạnh, phụ thuộc yếu.

- Từ năm 1977 đến năm 2003, R. Fagin và Zaniolo và một số nhóm tác giả khác đã đề xuất phụ thuộc đa trị [18] [19] [20], phụ thuộc đa trị mở rộng tập trị không chỉ nhận hai giá trị $\{0, 1\}$ mà bao gồm k giá trị thực trong đoạn $[0, 1]$.

- Năm 1981 - 1985, các nhóm nghiên cứu của Berman, Blok và Sagiv, Delobel [21] [22] đã phát triển khái niệm phụ thuộc hàm sang khái niệm phụ thuộc Boole dương (PTBD), bao gồm những ràng buộc dữ liệu được mô tả thông qua các công thức Boole dương (CTBD), nhưng vẫn giữ nguyên phép sánh trị đẳng thức.

- Năm 1995 Jyrki Kivinen và các đồng nghiệp đề xuất phụ thuộc hàm xấp xỉ (PTHXX) [24]. Các nhóm Hultala Y. và các đồng nghiệp [25], Ronald S. K. và Janes J. L. [26] đã phát triển thêm một số thuật toán cho loại phụ thuộc hàm xấp xỉ này.

- Năm 2004, Ilyas và các đồng nghiệp đã nghiên cứu phụ thuộc hàm mềm, là phụ thuộc hàm mà giá trị của X xác định giá trị của Y với độ không chắc chắn cho trước.

- Năm 2007 Bohannon cùng các đồng nghiệp đề xuất phụ thuộc hàm có điều kiện để làm sạch dữ liệu.

- Năm 2011 nhóm nghiên cứu Song S. và Chen L. đề xuất phụ thuộc sai khác (PTSK) để giải quyết một số vấn đề như đảm bảo tính toàn vẹn, tối ưu truy vấn [34].

- Hiện nay, một số hướng phát triển mới về phụ thuộc dữ liệu đang được các nhóm tập trung nghiên cứu như phụ thuộc so sánh (Song S., Chen L., and Yu P.S - 2013) [32]; phân tích các ràng buộc theo cấu trúc mẫu (Baixeries J., Kaytoue M., and Napoli A. - 2015) [35], [36] mở rộng phụ thuộc hàm xấp xỉ.

- Năm 2016 các tác giả Loredana Caruccio, Vincenzo Deufemia, Giuseppe Polese đã tổng kết 35 loại phụ thuộc mở rộng của các nhóm nghiên cứu trên thế giới được gọi là *nhóm các phụ thuộc hàm nói lỏng*

Tình hình nghiên cứu trong nước:

- Tại Việt Nam, một số nhóm nghiên cứu trong nước đã mở rộng các loại phụ thuộc nhằm tạo các mối ràng buộc chặt chẽ hơn trong cơ sở dữ liệu: như các tác giả Đàm Gia Mạnh [8]; Vũ Ngọc Loan [3], Bùi Đức Minh [13], Nguyễn Hoàng Sơn [10], Lương Nguyễn Hoàng Hoa [12] [39] Lê Xuân Vinh [11]; Trương Thị Thu Hà [14]... đã khảo sát các lớp Boole dương tổng quát, phụ thuộc yếu, phụ thuộc sai khác, phụ thuộc đối ngẫu, ... dưới nhiều góc độ khác nhau.

- Nguyễn Xuân Huy và Lê Thị Thanh [23] mở rộng phụ thuộc Boole dương thành phụ thuộc Boole dương tổng quát (PTBDTQ), phụ thuộc Boole dương đa trị, phụ thuộc Boole dương theo nhóm bộ.

Trên cơ sở khảo sát và phân tích NCS thu được một số đặc trưng cơ bản sau đây:

(1) Các phụ thuộc dữ liệu có thể được mô tả thông qua các mệnh đề logic phản ánh tương quan giữa các thuộc tính trong cơ sở dữ liệu.

(2) Tất cả phụ thuộc dữ liệu trong cơ sở dữ liệu đều dựa trên cơ sở nhận thức của thế giới thực, đó là cố gắng biểu diễn ngữ nghĩa của dữ liệu trong thế giới thực.

(3) Phần lớn các kết quả tập trung vào các khái niệm cơ bản, những tính chất đặc trưng, ứng dụng và cũng như các thuật toán cơ sở quan trọng của lý thuyết cơ sở dữ liệu. Một số nghiên cứu hiện đại, sâu hơn xuất hiện trong thời gian gần đây của lý thuyết cơ sở dữ liệu theo hướng tổ hợp như tập đóng, khóa, phản khóa, chuyển dịch lược đồ quan hệ, họ các tập tối tiểu của thuộc tính, mở rộng phụ thuộc hàm hay tìm các mô tả tương đương của phụ thuộc hàm cũng được giới thiệu.

Luận án tiếp tục nghiên cứu và mở rộng phụ thuộc Boole dương tổng quát để thu được một dạng phụ thuộc mới là phụ thuộc Boole dương xấp xỉ, phụ thuộc boole dương xấp xỉ tổng quát và phụ thuộc yếu xấp xỉ với mục đích tạo một mô hình tổng quát cho các lớp phụ thuộc nói trên bằng cách chỉ ra mối tương quan giữa các loại phụ thuộc, đề xuất một lớp phụ thuộc thống nhất bao hàm các lớp phụ thuộc dữ liệu hiện đang được quan tâm trong nghiên cứu và triển khai.

2. Mục tiêu nghiên cứu

(1) Đề xuất một số loại phụ thuộc mới, bao gồm: phụ thuộc Boole dương xấp xỉ, tiếp tục mở rộng biến thể của phụ thuộc hàm là phụ thuộc Boole dương xấp xỉ thành phụ thuộc Boole dương xấp xỉ tổng quát, phụ thuộc yếu thành phụ yếu xấp xỉ tổng quát.

(2) 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. 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.

(3) Luận án đề xuất một lớp phụ thuộc mới là phụ thuộc logic rộng hơn các phụ thuộc đã biết. Với lớp phụ thuộc logic này, luận án đã thu được các kết quả sau đây:

- Đề xuất qui trình giải bài toán suy dẫn theo ba tiếp cận hình thức là chứng minh trực tiếp theo thuật giải Vương Hạo, chứng minh trực tiếp theo chuẩn hội và chứng minh phản chứng theo hợp giải và các kết quả mới về việc vận dụng các phương pháp chứng minh tautology.

- Xây dựng và đánh giá thuật toán tìm bao đóng của một tập thuộc tính cho lớp các phụ thuộc logic.

- Xây dựng và đánh giá thuật toán tìm khóa cho lớp các phụ thuộc logic.

3. Đối tượng và phạm vi nghiên cứu

- *Đối tượng nghiên cứu* của luận án là các khái niệm và tính chất của các phụ thuộc logic: Phụ thuộc Boole dương, phụ thuộc Boole dương tổng quát, phụ thuộc yếu, phụ thuộc Boole dương xấp xỉ, phụ thuộc Boole dương xấp xỉ tổng quát... Mối quan hệ giữa các phụ thuộc logic. Các bài toán kinh điển trong lý thuyết các phụ thuộc.

- *Phạm vi nghiên cứu* của luận án là các biến thể của phụ thuộc hàm, phụ thuộc Boole dương tổng quát, mối quan hệ giữa các phụ thuộc logic và phương pháp giải một số bài toán kinh điển trong phụ thuộc logic.

4. Phương pháp nghiên cứu

Phương pháp nghiên cứu chủ yếu của luận án là:

(1) *Nghiên cứu lý thuyết*: Phương pháp suy luận, diễn giải, hình thức hóa từ các kết quả nghiên cứu để trình bày các khái niệm về các phụ thuộc logic cơ bản đã được nghiên cứu theo tuần tự: nêu định nghĩa, các đặc trưng, các bài toán,... đã được chứng minh và sử dụng; phân tích, tổng hợp, chứng minh để đưa ra kết quả dự kiến.

(2) *Nghiên cứu thực nghiệm*: Các thuật toán đề xuất được so sánh, đánh giá với các thuật toán khác nhằm minh chứng về tính hiệu quả của các nghiên cứu về lý thuyết.

5. Nội dung nghiên cứu

(1) Nghiên cứu các biến thể của phụ thuộc Boole nhằm thu được các lớp phụ thuộc tổng quát hơn và mở rộng các ứng dụng trong quản lý và khai thác các cơ sở dữ liệu.

(2) Nghiên cứu các phép sánh trị giữa các bộ trong cơ sở dữ liệu. Đề xuất hàm Lambda định lượng cho các thuộc tính và vận dụng khái niệm độ đo trong việc so sánh các bộ của quan hệ, thu được các dạng phụ thuộc mới là phụ thuộc Boole dương xấp xỉ, Boole dương xấp xỉ tổng quát, phụ thuộc yếu xấp xỉ.

(3) Đề xuất khái niệm về phụ thuộc Boole dương xấp xỉ, phụ thuộc Boole dương xấp xỉ tổng quát trong mô hình dữ liệu quan hệ, phát biểu và chứng minh các định lý, các tính chất của Phụ thuộc Bool dương xấp xỉ và phụ thuộc Boole dương xấp xỉ tổng quát, khẳng định mối quan hệ giữa phụ thuộc Bool dương xấp xỉ và phụ thuộc Bool dương tổng quát, giữa phụ thuộc Boole dương xấp xỉ tổng quát và phụ thuộc Boole dương tổng quát.

(4) Đề xuất khái niệm về phụ thuộc yếu xấp xỉ, phát biểu và chứng minh các định lý, các tính chất của Phụ thuộc yếu xấp xỉ chỉ ra mối quan hệ giữa phụ thuộc yếu xấp xỉ và phụ thuộc Bool dương tổng quát.

(5) So sánh, đánh giá các thuật toán đề xuất với các thuật toán khác đã được công bố.

6. Ý nghĩa khoa học và thực tiễn

Ý nghĩa khoa học

Về mặt khoa học luận án đã phát triển, mở rộng một số phụ thuộc dữ liệu và xây dựng lớp các phụ thuộc logic với các tính chất và đặc trưng chung của các lớp phụ thuộc.

Ý nghĩa thực tiễn

Kết quả của luận án được dùng làm cơ sở để xây dựng mô hình tổng quát về các lớp phụ thuộc khác nhau trong khai thác dữ liệu và tri thức bằng các công cụ AI và học sâu theo một khung nhìn tổng quan về mối quan hệ giữa các phụ thuộc logic trong cơ sở dữ liệu. Từ đó, có thể lựa chọn lý thuyết của các phụ thuộc phù hợp hơn cho việc thiết kế cơ sở dữ liệu cụ thể đáp ứng được các nhu cầu thường xuyên biến động của thực tiễn, có khả năng hỗ trợ cho các ứng dụng đa phương tiện, đáp ứng được nhu cầu về thu thập, tổ chức, quản lý cơ sở dữ liệu.

7. Bố cục của luận án

Chương 1. Trình bày các kiến thức nền tảng liên quan đến luận án, cụ thể: Trình bày tổng quan về phụ thuộc hàm, các khái niệm về quan hệ, thuộc tính, bộ, định lý tương đương... và tập trung trình bày nội dung về phụ thuộc hàm nói lỏng, phụ thuộc hàm xấp xỉ, phụ thuộc boole dương tổng quát.

Chương 2. Trình bày kết quả nghiên cứu của bản thân NCS về phụ thuộc Boole dương xấp xỉ, phụ thuộc Boole dương xấp xỉ tổng quát, phụ thuộc yếu xấp xỉ.

Chương 3. Trình bày bài toán suy dẫn, thuật toán bao đóng, thuật toán tìm khoá trong lược đồ quan hệ, phương pháp để chuyển một công thức logic bất kỳ về dạng chuẩn hội, các thuật toán chứng minh hằng đúng và một số kết quả liên quan của NCS.

Kết luận. Tổng kết các kết quả đã đạt được, những điểm còn tồn tại và hướng nghiên cứu tiếp theo.

CHƯƠNG 1. CÁC LỚP PHỤ THUỘC DỮ LIỆU TRONG CƠ SỞ DỮ LIỆU

1.1. Mở đầu

Các kết quả của chương này bao gồm:

(1) Trình bày các vấn đề quan trọng và tinh túy nhất liên quan đến khái niệm về phụ thuộc hàm, phụ thuộc Boole dương tổng quát, phụ thuộc hàm xấp xỉ, phụ thuộc hàm nói lỏng.

(2) 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.

(3) 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.

1.2. Một số khái niệm và quy ước

Khái niệm thuộc tính, miền trị, bộ và quan hệ :

Cho $U = \{a_1, a_2, \dots, a_n\}$, $n \geq 1$, U gọi là *tập các thuộc tính*. Mỗi phần tử $a \in U$ là tập *miền trị* V_x . Kí hiệu \mathcal{D} là *hợp các miền trị* V_x của các thuộc tính trong U , $\mathcal{D} = \bigcup_{a \in U} V_x$.

Quan hệ r với tập thuộc tính U , ký hiệu $R(U)$, là tập các ánh xạ $t: U \rightarrow \mathcal{D}$ sao cho mỗi thuộc tính $a \in U$ ta có $t.a \in d_x$, trong đó $t.a$ là ảnh của thuộc tính a qua ánh xạ t . Mỗi ánh xạ t được gọi là một *bộ* trong quan hệ r [1] [40] [41] [42].

- Phép gán trị e cho biến x : $x := e$

- Kí hiệu $x := (e) ? a : b$,

- Miền trị của thuộc tính a được viết là d_a .

Các bộ trong quan hệ r được ký hiệu là t, u, v, \dots hoặc kèm chỉ số t_i, u_j, v_k . Trị của thuộc tính a trong bộ t là $t.a$, trị của tập con các thuộc tính X trong bộ t là $t.X = \{t.a \mid a \in X\}$.

Hợp của hai tập được viết $X \cup Y$; giao là: được viết $X \cap Y$; phép trừ là $X - Y$.

Hội hai công thức logic kí hiệu là $X \wedge Y$ hoặc XY hoặc $X \cdot Y$; phép tuyển: $X \vee Y$ hoặc $X + Y$, dấu ' thay cho phép phủ định \neg . Phân hoạch của tập M (thành các tập con rời nhau) X_1, X_2, \dots, X_k được ký hiệu $M = X_1 \mid X_2 \mid \dots \mid X_k$ với ý nghĩa $M = X_1 \cup X_2 \cup \dots \cup X_k$ và $X_i \cap X_j = \emptyset$, $1 \leq i, j \leq k$, $i \neq j$.

1.2. Phụ thuộc hàm

Cho tập thuộc tính U . Một *phụ thuộc hàm (PTH)* trên U là biểu thức dạng

$$f: X \rightarrow Y; X, Y \subseteq U$$

Cho quan hệ $r(U)$ và một PTH $f: X \rightarrow Y$ trên U . Ta nói quan hệ r *thỏa PTH* f và viết $r(f)$, nếu hai bộ tùy ý trong r giống nhau trên X thì chúng cũng giống nhau trên Y ,

$$R(X \rightarrow Y) \Leftrightarrow (\forall u, v \in R): (u.X = v.X) \Rightarrow (u.Y = v.Y)$$

Ký hiệu $X \not\rightarrow Y$ nghĩa là tập thuộc tính Y *không phụ thuộc hàm* vào tập thuộc tính X .

Cho tập PTH F trên tập thuộc tính U . Ta nói quan hệ $r(U)$ *thỏa tập PTH* F ,

$$R(F) \Leftrightarrow (\forall f \in F): R(f)$$

Bao đóng của tập thuộc tính

Cho tập PTH F trên tập thuộc tính U và một tập con các thuộc tính X trong U . *Bao đóng của tập thuộc tính* X , ký hiệu X^+ , là tập thuộc tính $X^+ = \{A \in U \mid X \rightarrow A \in F^+\}$. *Một số tính chất của bao đóng*

Cho $LDQH$ $a = (U, F)$. Khi đó $\forall X, Y \subseteq U$ ta có:

$$(1) \text{ Tính phản xạ: } X \subseteq X^+$$

(2) *Tính đồng biến*: $X \subseteq Y \Rightarrow X^+ \subseteq Y^+$

(3) *Tính lũy đẳng*: $(X^+)^+ = X^+$

Lược đồ quan hệ

Lược đồ quan hệ (LĐQH) là một cặp (U, F) , U là tập hữu hạn các thuộc tính, F là tập các ràng buộc.

Khóa của lược đồ quan hệ

Cho LĐQH $a = (U, F)$. Tập thuộc tính $K \subseteq U$ được gọi là *khóa* của LĐQH a nếu

(i) $K^+ = U$

(ii) $\forall A \in K: (K - \{A\})^+ \neq U$

Nếu K thỏa điều kiện (i) (hoặc (i')) thì K được gọi là một *siêu khóa*.

Cho LĐQH $a = (U, F)$. Ta ký hiệu U_K là tập các *thuộc tính khóa* của a và U_0 là tập các *thuộc tính không khóa* của a . Dễ thấy $U_K | U_0$ là một phân hoạch của U .

Định lý 1.1: *Định lý tương đương cho phụ thuộc hàm* [17][20]

Cho tập PTH F và một PTH f trên tập thuộc tính U , ba loại suy dẫn sau là tương đương:

- Suy dẫn logic: $F \models f$
- Suy dẫn theo quan hệ: $F \vdash f$
- Suy dẫn theo quan hệ có không quá hai bộ: $F \vdash_2 f$

Cho LĐQH (U, F) , F là tập phụ thuộc hàm trên tập thuộc tính U .

Bao đóng của tập phụ thuộc hàm, ký hiệu F^+ , là tập tất cả các phụ thuộc hàm trên U được suy dẫn logic từ F . $F^+ = \{f \mid F \models f\}$

1.3. Phụ thuộc hàm nói lỏng

Phụ thuộc hàm nói lỏng được xây dựng dưới dạng thức chung: $f: X(\lambda) \rightarrow Y(\gamma)$; $X, Y \subseteq U$ với các điều kiện nói lỏng 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 lỏng các phép sánh trị trên một vài thuộc tính: Quan hệ r thỏa phụ thuộc hàm nói lỏng theo phép sánh trị $r(X(\rho) \rightarrow Y)$ 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 .

Định nghĩa chung về *phụ thuộc hàm nói lỏng tổng quát*:

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

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

Ta nói phụ thuộc hàm nói lỏng 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 tân từ $\gamma(u, X, v, X)$ suy ra được tân từ $\vartheta(u, X, v, Y)$.

Phụ thuộc hàm nói lỏng 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 tân từ γ và ϑ .

1.4. Phụ thuộc Boole dương

1.4.1. Công thức Boole

Định nghĩa 1.1

Cho $U = \{x_1, \dots, x_n\}$ là tập hữu hạn các *biến Boole*, \mathcal{B} là tập *trị Boole*, $\mathcal{B} = \{0, 1\}$. Khi đó các *công thức Boole* (CTB) hay còn gọi là các *công thức logic* được xây dựng như sau:

- (i) Mỗi trị 0/1 trong \mathcal{B} là một CTB,
- (ii) Mỗi biến nhận giá trị trong U là một CTB,
- (iii) Nếu a là một CTB thì (a) là một CTB,

- (iv) Nếu a và b là các CTB thì $a \vee b$, $a \wedge b$, $\neg a$ và $a \rightarrow b$ là một CTB,
 (v) Chỉ có các công thức được tạo bởi các quy tắc (i) - (iv) là các CTB.
 Ký hiệu $L(U)$ là tập các CTB xây dựng trên tập các biến U .

Định nghĩa 1.2

Mỗi vector các phần tử 0/1, $v = (v_1, \dots, v_n)$ trong không gian $\mathcal{B}^n = \mathcal{B} \times \mathcal{B} \times \dots \times \mathcal{B}$ được gọi là một *phép gán trị*.

Khi đó với mỗi CTB $f \in L(U)$ ta có $f(v) = f(v_1, \dots, v_n)$ là *trị của công thức f* đối với phép gán trị v , và $f(v)$ được tính như sau:

1. Thay toàn bộ biến x_i trong f bằng trị v_i tương ứng, $i=1, 2, \dots, n$ để thu được một mệnh đề logic b .

2. Trị của $f(v)$ chính là trị của b .

Với mỗi tập con $X \subseteq U$, ta quy ước viết *hội logic* (\wedge) của các biến trong X là dãy ký hiệu của X . X đồng thời biểu diễn cho các đối tượng sau đây:

- một tập thuộc tính trong U ,
- một tập biến logic trong U ,
- một công thức Boole được lập bởi hội logic các biến trong X .

Ta gọi công thức $f: Z \rightarrow V$ là

- công thức suy dẫn nếu Z và V có dạng hội, tức là $f: \wedge Z \rightarrow \wedge V$
- công thức suy dẫn mạnh nếu Z có dạng tuyển và V có dạng hội, tức là $f: \vee Z \rightarrow \wedge V$
- công thức suy dẫn yếu nếu Z có dạng hội và V có dạng tuyển, tức là $f: \wedge Z \rightarrow \vee V$
- công thức suy dẫn đối ngẫu nếu Z và V đều có dạng tuyển, tức là $f: \vee Z \rightarrow \vee V$

Hai phép gán trị đặc biệt là *phép gán trị đơn vị*, $e = (1, 1, \dots, 1)$ và *phép gán trị không*, $z = (0, 0, \dots, 0)$.

Với mỗi tập hữu hạn các CTB, $F = \{f_1, f_2, \dots, f_m\}$ trong $L(U)$, F là một công thức dạng $F = f_1 \wedge f_2 \wedge \dots \wedge f_m$. Khi đó với mỗi phép gán trị v , giá trị chân lý của công thức F được tính là:

$$F(v) = f_1(v) \wedge f_2(v) \wedge \dots \wedge f_m(v)$$

1.4.2. Bảng trị và bảng chân lý

Khái niệm *bảng trị* và *bảng chân lý* được định nghĩa như sau:

- Với mỗi công thức f trên U , *bảng trị của f* . *Bảng chân lý của f* , ký hiệu là T_f , là tập các phép gán trị v sao cho $f(v)$ nhận giá trị 1, $T_f = \{v \in \mathcal{B}^n \mid f(v) = 1\}$

Bảng chân lý T_F của tập hữu hạn các công thức F trên U , chính là *giao* của các bảng chân lý của mỗi công thức thành viên trong F

$$T_F = \bigcap_{f \in F} T_f$$

Ta có, $v \in T_F$ khi và chỉ khi $\forall f \in F: f(v) = 1$.

1.5. Phụ thuộc Boole dương tổng quát

- $U = \{x_1, \dots, x_n\}$ là tập hữu hạn các biến Boole nhận giá trị trong tập trị logic $\mathcal{B} = \{0, 1\}$.
- Quy ước mỗi miền trị V_x của thuộc tính x trong U có chứa ít nhất hai phần tử. Với mỗi miền trị V_x , xét ánh xạ $\alpha_x: V_x^2 \rightarrow \mathcal{B}$ thoả các tiên đề sau:

$$\forall a, b \in V_x$$

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

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

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

α_x chính là quan hệ bộ phận, thoả các tính chất phản xạ và đối xứng trên miền trị V_x .

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

- 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.

1.6. Phân loại các lớp phụ thuộc Boole dương tổng quát

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.

1.6.1. Lớp IE (Implication Formula & 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 lược đồ *phụ thuộc hàm truyền thống*.

IE-2: Biểu thức logic có dạng $\forall X \rightarrow Y$ cho ta lược đồ *phụ thuộc hàm mạnh*.

IE-3: Biểu thức logic có dạng $X \rightarrow \forall Y$, ta có lược đồ *phụ thuộc hàm yếu*.

IE-4: Biểu thức logic có dạng $\forall X \rightarrow \forall Y$, ta có lược đồ *phụ thuộc hàm đối ngẫu*.

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	nói lỏng về trái
$X \rightarrow \forall Y$	phụ thuộc hàm yếu	nói lỏng về phải
$\forall X \rightarrow \forall Y$	phụ thuộc hàm đối ngẫu	nói lỏng
$X(\delta) \rightarrow Y$	phụ thuộc xấp xỉ	nói lỏng

Bảng 2. 2. Bảng đặc tả các lớp con IE1-4

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

1.6.2. Lớp LA (Logic Formula & Alpha Comparison)

Dựa theo điều kiện nói 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 sánh trị α .

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 α_X trên tập X , thì u và v cũng thỏa các ràng buộc được đặc tả bởi hàm sai khác α_Y trên tập 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:

- $f: \alpha_X \rightarrow \alpha_Y$, 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 a 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 $=, \neq, <, \leq, >, \geq$.

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

Phụ thuộc hàm sai khác là trường hợp riêng của phụ thuộc hàm nói 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 nói lỏng theo lực lượng*.

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.

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. Đại diện cho lớp LA-3 là phụ thuộc Boole dương tổng quát, 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.

1.7. Kết luận chương 1

Trong chương 1 luận án đã trình bày các khái niệm cơ sở về phụ thuộc dữ liệu trong cơ sở dữ liệu quan hệ và phân tích để thấy rằng việc so sánh các bộ theo đẳng thức là chưa đủ để cơ sở dữ liệu phản ánh tính đa dạng của ngữ nghĩa của dữ liệu trong thực tiễn.

Chương 1 cũng trình bày các nghiên cứu liên quan đến các định hướng nghiên cứu của luận án: nghiên cứu phát triển, mở rộng lớp phụ thuộc Boole dương xấp xỉ. Có thể tóm lược các điểm nhấn quan trọng của Chương 1 như sau:

(1) Với phép so sánh đẳng thức ta có thể phát triển các loại phụ thuộc dữ liệu dạng hàm và một số biến thể của dạng này như phụ thuộc hàm nói lỏng, phụ thuộc hàm xấp xỉ.

(2) Định lý tương đương cho phụ thuộc hàm kinh điển cho phép giải bài toán thành viên trên cơ sở của phép suy dẫn logic $X \rightarrow Y$.

(3) Kỹ thuật xây dựng bảng trị của quan hệ theo phép so sánh đẳng thức. Kỹ thuật kiểm tra tính thỏa của quan hệ đối với phụ thuộc hàm theo phép so sánh đẳng thức.

(4) 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.

CHƯƠNG 2. CÁC LỚP PHỤ THUỘC XẤP XỈ TRONG CƠ SỞ DỮ LIỆU

2.1. Mở đầu

Trong chương này, luận án đề xuất các *dạng phụ thuộc mới là phụ thuộc Boole dương xấp xỉ, phụ thuộc boole dương xấp xỉ tổng quát và phụ thuộc yếu xấp xỉ*. Nội dung chương sẽ bao gồm các kết quả cụ thể sau:

(1) Đề xuất cách xây dựng một lược đồ Boole dương tổng quát từ một lược đồ xấp xỉ dạng mở rộng cho trước.

(2) Đề xuất hàm Lambda trên miền trị của thuộc tính và chứng minh các định lý đảm bảo toán học cho việc vận dụng hàm Lambda nhằm phát triển khái niệm phụ thuộc Boole dương xấp xỉ và phụ thuộc Boole dương xấp xỉ tổng quát.

(3) Đề xuất các tính chất và định lý để xây dựng cách tiếp cận mới cho phụ thuộc Boole dương xấp xỉ và phụ thuộc Boole dương xấp xỉ tổng quát theo giá trị chứ không theo số bộ. Cụ thể là chứng minh định lý về điều kiện cần và đủ để một quan hệ thỏa tập phụ thuộc Boole dương xấp xỉ.

(4) Đề xuất khái niệm tổng quát về phụ thuộc xấp xỉ trên tập các thuộc tính số và phi số, không nhất thiết theo dạng hàm mà theo một biểu thức Boole dương tùy ý với các phép sánh trị tổng quát hơn phép so sánh đẳng thức.

Các kết quả trong chương này được công bố trong các CT1-CT6, phần “*Danh mục các công trình đã công bố*” của NCS.

2.2. Xây dựng hàm lambda và độ đo

2.2.1. Hàm lambda

Định nghĩa 2.1

Cho U là tập thuộc tính. Với mỗi thuộc tính A trong U ta định nghĩa một ánh xạ λ_A từ miền trị của thuộc tính sang không gian số thực:

$$\lambda_A: V_A \rightarrow \mathbb{R}$$

- Tính đơn trị: $\forall a, b \in V_A: a = b \Rightarrow \lambda_A(a) = \lambda_A(b)$
- Tính khác biệt: $\exists a, b \in V_A: \lambda_A(a) \neq \lambda_A(b)$

Gọi λ_A là hàm định lượng cho thuộc tính A .

Định lý 2.1

Biết λ_A , khi đó hàm α_A được định nghĩa qua hệ thức (*) dưới đây

$$\forall a, b \in V_A: \alpha_A(a, b) \stackrel{\text{def}}{\iff} (\lambda_A(a) = \lambda_A(b)) \quad (*)$$

thỏa ba tiên đề A1, A2 và A3 trong định nghĩa ánh xạ α .

2.2.2. Độ đo

Độ đo d trên tập V là một ánh xạ $d: V \times V \rightarrow \mathbb{R}^+$ với \mathbb{R}^+ là tập số thực không âm, thỏa các tiên đề sau đây:

$$\forall a, b, c \in V$$

- Tiên đề không âm: $d(a, b) \geq 0, d(a, a) = 0$
- Tiên đề đối xứng: $d(a, b) = d(b, a)$
- Tiên đề tam giác: $d(a, b) + d(b, c) \geq d(a, c)$.

Định lý 2.2

Cho quan hệ r trên tập thuộc tính U và các hàm Lambda trên mỗi thuộc A trong U . Cho d là một độ đo tùy ý trong không gian vector n chiều. Khi đó hàm d^* dưới đây là một độ đo:

$$\forall u, v \in r: d^*(u, v) = d(\lambda(u), \lambda(v))$$

Ta gọi d^* là *độ đo cảm sinh từ độ đo d* .

2.3. Đề xuất phụ thuộc hàm xấp xỉ tổng quát

Trong phần này sẽ mở rộng khái niệm *phụ thuộc hàm xấp xỉ* trên các tập thuộc tính số, sau đó sẽ mở rộng khái niệm *phụ thuộc hàm xấp xỉ* trên các tập thuộc tính tùy ý số hoặc phi số.

Khái niệm phụ thuộc hàm xấp xỉ tổng quát được định nghĩa như sau:

Cho hai tập con thuộc tính X và Y trong tập thuộc tính U và hai ngưỡng $\varepsilon_1, \varepsilon_2$ là hai số thực không âm. Phụ thuộc hàm xấp xỉ của tập thuộc tính Y theo tập thuộc tính X là biểu thức dạng

$$X \varepsilon_1 \rightarrow \varepsilon_2 Y$$

Ta nói tập thuộc tính Y phụ thuộc xấp xỉ vào tập thuộc tính X và quan hệ $r(U)$ thỏa phụ thuộc hàm xấp xỉ $X \varepsilon_1 \rightarrow \varepsilon_2 Y$ nếu với mọi cặp bộ u và v trong quan hệ r , hai bộ u và v có độ đo lệch nhau không quá ε_1 trên tập thuộc tính X thì hai bộ này cũng lệch nhau không quá ε_2 trên tập thuộc tính Y .

trong đó d là hàm độ đo trên các bộ của quan hệ.

$$\forall u, v \in r(U): d(u.X, v.X) \leq \varepsilon_1 \Rightarrow d(u.Y, v.Y) \leq \varepsilon_2$$

2.4. Xây dựng lược đồ quan hệ xấp xỉ thông qua độ đo

Lược đồ quan hệ xấp xỉ là một cặp $p = (U, F, d)$, trong đó U là tập thuộc tính, F là tập các PTHXX theo độ đo d .

Với mỗi độ đo d trên D ta liên kết một giá trị $\varepsilon \geq 0$ và gọi là ngưỡng của d trên D . Nếu trong D tồn tại hai giá trị a và b thỏa tính chất $d(a, b) > \varepsilon$ thì ε được gọi là *ngưỡng không tầm thường*; ngược lại, nếu với mọi cặp tri $a, b \in D$, ta đều có $d(a, b) \leq \varepsilon$ thì ε được gọi là *ngưỡng tầm thường*.

2.5. Phụ thuộc yếu

Cho tập thuộc tính U . Một công thức suy dẫn yếu được gọi là một phụ thuộc yếu. Như vậy mỗi phụ thuộc yếu có dạng

$$f: X \rightarrow Y$$

trong đó X là một hội, Y là một tuyển của các thuộc tính trong U .

Ta nói quan hệ $r(U)$ thỏa phụ thuộc yếu $f: X \rightarrow Y$ và viết $r(X \rightarrow Y)$ nếu với mọi cặp bộ u và v trong quan hệ r , hai bộ u và v bằng nhau trên X sẽ bằng nhau trên một thuộc tính nào đó của Y .

$$r(X \rightarrow Y) \Leftrightarrow (\forall u, v \in R): (u.X = v.X) \Rightarrow \exists B \in Y: (u.B = v.B)$$

Cho tập phụ thuộc yếu F trên tập thuộc tính U . Ta nói quan hệ $r(U)$ thỏa tập phụ thuộc yếu F , và viết $r(F)$, nếu r thỏa mọi phụ thuộc yếu trong F ,

$$r(F) \Leftrightarrow (\forall f \in F): r(f)$$

Nếu quan hệ r thỏa phụ thuộc yếu f ta cũng nói phụ thuộc yếu f đúng trong quan hệ r .

Vì phụ thuộc yếu là trường hợp riêng của phụ thuộc Boole dương nên ta cũng có:

Quan hệ r thỏa phụ thuộc yếu f khi và chỉ khi $T_r \subseteq T_f$.

Quan hệ r thỏa tập phụ thuộc yếu F khi và chỉ khi $T_r \subseteq T_F$.

2.6. Đề xuất phụ thuộc yếu xấp xỉ

Cho tập thuộc tính U và một công thức suy dẫn yếu trên $U, f: X \rightarrow Y$. Cho hai ngưỡng $\varepsilon_1, \varepsilon_2$ là hai số thực không âm. Phụ thuộc yếu xấp xỉ (PTYXX) của tập thuộc tính Y theo tập thuộc tính X là biểu thức dạng

$$X \varepsilon_1 \rightarrow \varepsilon_2 Y$$

Ta nói quan hệ $r(U)$ thỏa *phụ thuộc yếu xấp xỉ* $X \varepsilon_1 \rightarrow \varepsilon_2 Y$ nếu với mọi cặp bộ u và v trong quan hệ r , hai bộ u và v có độ đo lệch nhau không quá ε_1 trên tập thuộc tính X thì hai bộ này cũng lệch nhau không quá ε_2 trên *một thuộc tính nào đó* của Y :

$\forall u, v \in r: d(u.X, v.X) \leq \varepsilon_1 \Rightarrow \exists B \in Y: d(u.B, v.B) \leq \varepsilon_2$, trong đó, d là một độ đo.

Ta gọi bộ sáu $p = (U, \varepsilon_1, \varepsilon_2, \lambda, d, F)$ là một lược đồ quan hệ với phụ thuộc yếu xấp xỉ, trong đó

- U là tập n thuộc tính, mỗi thuộc tính được trang bị một hàm định lượng Lambda λ tương ứng,

- d là một độ đo tùy ý,

- F là tập các công thức suy dẫn yếu $X \rightarrow Y$, X là một hội, Y là một tuyển khác rỗng các thuộc tính trong U ,

- $\varepsilon_1 \geq 0, \varepsilon_2 \geq 0$ là các ngưỡng xấp xỉ.

Cho quan hệ r trên lược đồ quan hệ với phụ thuộc yếu xấp xỉ, $p = (U, \varepsilon_1, \varepsilon_2, \lambda, d, F)$. Với mỗi cặp bộ $u = (u_1, u_2, \dots, u_n), v = (v_1, v_2, \dots, v_n)$

trong quan hệ r , ta định nghĩa

$t(u, v) = (t_1, t_2, \dots, t_n)$

$$t_i = (d(\lambda(u_i), \lambda(v_i)) \leq \varepsilon) ? 1 : 0 ; 1 \leq i \leq n$$

$$\varepsilon = \min(\varepsilon_1, \varepsilon_2)$$

$$T_r = \{t(u, v) \mid u, v \in r\}$$

và gọi T_r là bảng trị của quan hệ r theo lược đồ p .

Định lý 2.3

Cho lược đồ quan hệ với tập $PTYXX F$, $p = (U, \varepsilon_1, \varepsilon_2, \lambda, d, F)$ và một $PTYXX f$. Cho quan hệ r trên U . Khi đó,

(i) Quan hệ r thỏa $PTYXX f$ khi và chỉ khi $T_r \subseteq T_f$.

(ii) Quan hệ r thỏa tập $PTYXX F$ khi và chỉ khi $T_r \subseteq T_F$.

Chứng minh

2.7. Đề xuất phụ thuộc Boole dương xấp xỉ

- Việc tiếp tục mở rộng khái niệm phụ thuộc Boole dương sang phụ thuộc Boole dương xấp xỉ có thể giúp ta quản lí các cơ sở dữ liệu phức tạp hơn. Đặc biệt là cho phép mở rộng khả năng tìm kiếm dữ liệu.

Các hướng nghiên cứu cơ sở dữ liệu mờ và thô cũng cho phép ta tìm kiếm với các truy vấn thuộc loại trên. Tuy nhiên, sự phụ thuộc giữa các thuộc tính dưới dạng logic tổng quát kèm theo độ đo phản ánh tính xấp xỉ thì còn là vấn đề mở, đáng được quan tâm nghiên cứu.

Khái niệm về phụ thuộc Boole dương xấp xỉ được định nghĩa như sau:

Cho tập U gồm n thuộc tính, một công thức Boole f dương trên U , và các hàm λ_i trên mỗi thuộc tính i trong U . Cho các ngưỡng ε_i là các số thực không âm trên mỗi thuộc tính i trong U . Ta gọi f là một phụ thuộc Boole dương xấp xỉ theo ngưỡng $\varepsilon = (\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n)$.

Ta nói phụ thuộc Boole dương f thỏa quan hệ $r(U)$ theo ngưỡng ε nếu với mọi cặp bộ $u, v \in r: f(t(u, v)) = 1$, trong đó $t(u, v)$ được xác định như sau:

$$t(u, v) = (t_1, t_2, \dots, t_n)$$

$$t_i = (|\lambda_i(u_i) - \lambda_i(v_i)| \leq \varepsilon_i) ? 1 : 0$$

Các điều kiện trên có ý nghĩa như sau:

Quan hệ r thỏa phụ thuộc Boole dương xấp xỉ nếu với mỗi cặp bộ có độ đo theo từng thuộc tính đạt dưới ngưỡng quy định thỏa công thức f .

Định nghĩa 2.6

Ta gọi bộ bốn $p = (U, \lambda, \varepsilon, F)$ là một *lược đồ quan hệ với phụ thuộc Boole dương xấp xỉ, trong đó*

- U là tập n thuộc tính,
- F là tập các công thức Boole dương trên U ,
- $\lambda = (\lambda_1, \lambda_2, \dots, \lambda_n)$ là các hàm định lượng cho các thuộc tính trong U ,
- $\varepsilon = (\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n)$ là các ngưỡng xấp xỉ cho các thuộc tính trong U ,
- $1 \leq i \leq n$

Cho quan hệ r trên lược đồ quan hệ với phụ thuộc Boole dương xấp xỉ, $p = (U, \lambda, \varepsilon, F)$. Ta ký hiệu, $T_r = \{t(u, v) : u, v \in r\}$

và gọi T_r là bảng chân lý của quan hệ r theo lược đồ p .

Định lý 2.4

Cho lược đồ quan hệ với phụ thuộc Boole dương xấp xỉ $p = (U, \lambda, \varepsilon, F)$ và một phụ thuộc Boole dương xấp xỉ $f \in F$. Cho quan hệ r trên U . Khi đó,

- (i) Quan hệ r thỏa PTBD xấp xỉ f khi và chỉ khi và chỉ khi $T_r \subseteq T_f$:

$$r(f) \Leftrightarrow T_r \subseteq T_f$$

- (ii) Quan hệ r thỏa tập PTBD xấp xỉ F khi và chỉ khi và chỉ khi $T_r \subseteq T_F$:

$$r(F) \Leftrightarrow T_r \subseteq T_F$$

2.8. Đề xuất thuộc Boole dương xấp xỉ tổng quát

2.8.1. Xây dựng phép sánh trị alpha

Ta có thể xây dựng hàm *Alpha* dựa trên hàm *Lambda* như sau:

Giả sử với mỗi thuộc tính A trong U đã có hàm λ_A . Ta định nghĩa

$$\forall a, b \in d_A: \alpha_A(a, b) \equiv (\lambda_A(a) = \lambda_A(b)) (*)$$

Định lý sau đây khẳng định rằng α_A thỏa các tính chất phản xạ, đối xứng và bộ phận.

Định lý 2.5

Hàm α_A được định nghĩa qua hệ thức thỏa ba tính chất $A1, A2$ và $A3$ trong định nghĩa ánh xạ α .

Định nghĩa 2.7

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], [6], [16].

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 $t = (t_1, t_2, \dots, t_n) \in \mathfrak{B}^n$ và ký 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 lược đồ p .

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* (PTBDTQ), lược đồ thu được trong trường hợp này được gọi là *lược đồ với phụ thuộc boole dương tổng quát*.

Ta nói quan hệ r trên tập thuộc tính U thỏa PTBDTQ f và ký hiệu $r(f)$, nếu $T_r \subseteq T_f$:

$$r(f) \Leftrightarrow T_r \subseteq T_f$$

Nếu $r(f)$ ta cũng nói PTBDTQ f đúng trong quan hệ r .

Quan hệ r thỏa tập PTBDTQ F và ký hiệu $r(F)$, nếu R thỏa mọi PTBDTQ trong F ,

$$r(F) \Leftrightarrow \forall f \in F: r(f) \Leftrightarrow T_r \subseteq T_F.$$

Cho tập PTBDTQ F và một PTBDTQ f . Ta nói

F dẫn ra được f theo quan hệ, và ký hiệu $F \vdash f$ nếu $\forall r \in REL(U): r(F) \Rightarrow r(f)$

F dẫn ra được f theo quan hệ có không quá hai bộ, ký hiệu $F \vdash_2 f$ nếu

$$\forall r \in REL_2(U): r(F) \Rightarrow r(f)$$

(Định lý tương đương)

[17] Cho tập PTBDTQ F và một PTBDTQ f trên U . Ba mệnh đề sau là tương đương,

(i) $F \vDash f$ (suy dẫn logic)

(ii) $F \vdash f$ (suy dẫn theo quan hệ)

(iii) $F \vdash_2 f$ (suy dẫn theo quan hệ có không quá 2 bộ)

2.8.2. Phụ thuộc Boole dương xấp xỉ tổng quát

Định nghĩa 2.8

Cho tập thuộc tính U và một công thức Boole dương f trên U . Cho hai ngưỡng $\varepsilon_1, \varepsilon_2$ là hai số thực không âm. Phụ thuộc Boole dương xấp xỉ tổng quát (PTBDXXTQ) là tập các phụ thuộc yếu xấp xỉ S tương đương với f theo ngưỡng $\varepsilon_1, \varepsilon_2$.

Ta nói quan hệ $r(U)$ thỏa PTBDXXTQ f nếu r thỏa mọi PTYXX trong S .

Do f tương đương với S nên $T_f = T_S = \cap \{T_g \mid g \in S\}$ nên ta có kết quả sau đây như là một hệ quả của định lý 4.1:

Hệ quả 2.3

Cho lược đồ quan hệ $p = (U, \varepsilon_1, \varepsilon_2, \lambda, d, F)$ và một PTBDTQXX f . Cho quan hệ r trên U . Khi đó,

(i) Quan hệ r thỏa PTBDXXTQ f khi và chỉ khi $T_r \subseteq T_f$.

(ii) Quan hệ r thỏa tập PTBDXXTQ F khi và chỉ khi $T_r \subseteq T_F$.

2.9. Kết luận chương 2

Trong Chương 2, luận án trình bày kết quả xây dựng hàm biến đổi lambda mới, dựa vào hàm lambda được xây dựng và kết hợp giữa phụ thuộc hàm xấp xỉ và phụ thuộc Boole dương tổng quát, luận án đề xuất các phụ thuộc dữ liệu mới bao gồm: phụ thuộc Boole dương xấp xỉ, phụ thuộc Boole dương xấp xỉ tổng quát và phụ thuộc yếu xấp xỉ và chứng minh sự tương đương giữa phụ thuộc Boole dương và chỉ ra và chứng minh mối quan hệ giữa phụ thuộc Boole dương tổng quát với phụ thuộc Boole dương xấp xỉ và phụ thuộc Boole dương tổng quát xấp xỉ.

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. Với một độ đo tùy ý, thông qua các hàm Lambda trên miền trị của các thuộc tính, ta có thể đánh giá độ xấp xỉ của các bộ trong quan hệ bao gồm các thuộc tính số và phi số như trang phục, kỹ năng biểu diễn,... để vận dụng vào hoạt động đánh giá và xếp loại các đối tượng cũng như tìm kiếm xấp xỉ các đối tượng trong cơ sở dữ liệu.

CHƯƠNG 3. CÁC THUẬT TOÁN XỬ LÝ LỢC ĐỒ QUAN HỆ

3.1. Mở đầu

Trong chương 3 NCS tập trung trình bày các nội dung sau:

(1) Đề xuất qui trình giải bài toán suy dẫn theo ba tiếp cận hình thức là chứng minh trực tiếp theo thuật giải Vương Hạo, chứng minh trực tiếp theo chuẩn hội và chứng minh phản chứng theo hợp giải và các kết quả mới về việc vận dụng các phương pháp chứng minh tautology để giải bài toán thành viên trong lớp các phụ thuộc logic và thu gọn các luật suy dẫn trong cơ sở tri thức

(2) Xây dựng và đánh giá thuật toán tìm bao đóng của một tập thuộc tính cho lớp các phụ thuộc logic.

(3) Xây dựng và đánh giá thuật toán tìm khóa cho lớp các phụ thuộc logic. Khóa trong một quan hệ được hiểu là một tập đủ nhỏ các thuộc tính có thể xác định không quá một bộ trong cơ sở dữ liệu. Khóa được vận dụng chủ yếu trong các thuật toán truy vấn cơ sở dữ liệu.

Các kết quả này là một trong những đóng góp cơ bản của NCS, được trình bày trong CT2, CT4 - *Danh mục các công trình công bố* của NCS.

3.2. Xây dựng phương pháp chuyển công thức logic về dạng chuẩn hội

Để có thể mở rộng khái niệm phụ thuộc Boole dương xấp xỉ, các phân lập luận dưới đây liên quan đến việc chuyển một công thức logic bất kỳ về dạng chuẩn hội (CNF).

Định nghĩa 3.1

Cho tập các biến Boole $U = \{a_1, a_2, \dots, a_n\}$. Công thức logic Ψ trên U có dạng chuẩn hội nếu Ψ được biểu diễn dưới dạng tích của các tuyến: $\Psi = g_1 g_2 \dots g_m$

trong đó mỗi g_i , $1 \leq i \leq m$ là một tuyến của các biến thành phần trong U .

Ví dụ, Cho tập biến $U = \{a, b, c, d\}$. Công thức Ψ sau thuộc dạng chuẩn hội: $\Psi = (a+b+c')(b+d)a'(b+c)$.

Định lý 3.3

Mọi công thức logic đều tương đương với một công thức CNF [32].

Theo các giáo trình về toán rời rạc và các tài liệu [28] [32] có hai phương pháp chuyển một công thức logic về dạng CNF như sau :

3.2.1. Phương pháp logic

Chuyển công thức logic Ψ về dạng CNF bằng cách áp dụng các luật dưới đây trong logic mệnh đề:

Các luật logic

Với mọi công thức X, Y, Z trên tập biến Boole U . Ta có:

(R1) Luật giao hoán: $X + Y \equiv Y + X$; $XY \equiv YX$.

(R2) Luật kết hợp: $X+(Y+Z) \equiv (X+Y)+Z$; $X(YZ) \equiv (XY)Z$.

(R3) Luật lũy đẳng: $X+X \equiv X$; $XX \equiv X$.

(R4) Luật trung hòa: $X+0 \equiv X$; $X+1 \equiv 1$; $X0 \equiv 0$; $X1 \equiv X$.

(R5) $XX' \equiv 0$; $X+X' \equiv 1$.

(R6) Phủ định của phủ định: $X'' \equiv X$.

(R7) $X \rightarrow Y \equiv X' + Y$.

(R8) $(X \rightarrow Y)' \equiv XY'$.

(R9) Luật de Morgan: $(XY)' \equiv X' + Y'$; $(X+Y)' \equiv X'Y'$.

(R10) Luật phân phối: $X + YZ \equiv (X+Y)(X+Z)$; $(X+Y)Z \equiv XZ + YZ$;

(R11) Luật nuốt: $X + XY \equiv X$; $X(X+Y) \equiv X$.

Lặp lại việc áp dụng các luật khử phép kéo theo, luật de Morgan, luật phân phối trong logic [46], [47], cho đến khi thu được công thức logic có dạng CNF.

3.2.2. Phương pháp lập bảng

Chuyển công thức logic Ψ trên tập thuộc tính U về dạng CNF theo phương pháp lập bảng như sau:

1. Lập bảng trị T_Ψ theo các biến trong U và công thức logic Ψ

2. Từ mỗi dòng $t = (v_1, \dots, v_n)$ thỏa $\Psi(t) = 0$ trong bảng trị T_Ψ tạo ra một tuyến cơ sở $g_t = (z_1 + \dots + z_n)$, với

$$z_i = \begin{cases} x_i & \text{nếu } v_i = 0 \\ x'_i & \text{nếu } v_i = 1 \end{cases} \quad 1 \leq i \leq n$$

3. Trả kết quả $M = g_1 \dots g_k$ là tích của các tuyến cơ sở.

Kết quả thu được sau bước 3 chính là dạng chuẩn hội của công thức Ψ .

Hai phương pháp có thể cho ra hai dạng CNF khác nhau, mặc dù các kết quả thu được là tương đương. Cả hai phương pháp đều thuộc lớp NPC và có độ phức tạp tính toán $O(2^n)$, trong đó n là số ký biến logic.

Mệnh đề 3.1

Mỗi công thức Boole tương đương với một tập các công thức suy dẫn yếu W .

Hệ quả 3.1

Mỗi CTB tương đương với một công thức suy dẫn yếu.

Mỗi tập CTB tương đương với một công thức suy dẫn yếu.

Hệ quả 3.2

Mỗi tập công thức Boole dương tương đương với một tập các công thức suy dẫn yếu.

3.3. Xây dựng phương pháp chứng minh công thức hằng đúng

Định nghĩa 3.2

Cho tập ký biến Boole U . Một công thức Boole f trên được gọi là *hằng đúng* (tautology) nếu $f(x) = 1$ với mọi phép gán trị x .

3.3.1. Phương pháp chứng minh trực tiếp theo CNF

Theo phương pháp này để chứng minh công thức f là hằng đúng ta thực hiện theo hai pha sau đây:

Pha 1. Chuyển f về dạng CNF. $f \equiv u_1 u_2 \dots u_k$

Pha 2. Kết luận: f là hằng đúng khi và chỉ khi $u_i = 1, 1 \leq i \leq k$.

3.3.2. Phương pháp Vương Hao

Phương pháp Vương Hao dùng để chứng minh tautology $T \rightarrow P$ theo các bước sau [5].

Bước 1 (CNF \rightarrow DNF).

Đưa vế trái T về dạng CNF.

Đưa vế phải i về dạng DNF là công thức logic có dạng chuẩn tuyến.

Để đưa công thức $T \rightarrow P$ về dạng CNF \rightarrow DNF ta có thể vận dụng các luật R1 – R11 để biến đổi hai vế T và P thành dạng chuẩn sau: $t_1 t_2 \dots t_u \rightarrow p_1 + p_2 + \dots + p_v$

Trong đó vế trái là một CNF, vế phải là một DNF.

Bước 2 (Khử phủ định). Chuyển vế các biến có dấu phủ định. Nếu x' xuất hiện tại vế trái thì chuyển x' sang vế phải thành x và ngược lại.

Bước 3 (Tách). Nếu dòng hiện hành có một trong hai dạng sau:

• Dạng 1: $t_1 \dots (a + b) \dots t_u \rightarrow p_1 + \dots + p_v$

Thì thay bằng 2 dòng:

$$\begin{cases} t_1 \dots a \dots t_u \rightarrow p_1 + \dots + p_v \\ t_1 \dots b \dots t_u \rightarrow p_1 + \dots + p_v \end{cases}$$

• Dạng 2: $t_1 \dots t_u \rightarrow p_1 + \dots + a \cdot b + \dots + p_v$

Thì thay bằng 2 dòng:

$$\begin{cases} t_1 \dots t_u \rightarrow p_1 + \dots + a + \dots + p_v \\ t_1 \dots t_u \rightarrow p_1 + \dots + b + \dots + p_v \end{cases}$$

Bước 4 (Kết luận). Một dòng được chứng minh khi và chỉ khi có một biến xuất hiện ở cả hai vế.

Bài toán được chứng minh nếu mọi dòng được chứng minh.

3.3.3. Phương pháp hợp giải

Trong logic, phương pháp hợp giải được sử dụng để giải bài toán suy dẫn $f \rightarrow g$.

Hợp giải còn được gọi là phương pháp chứng minh *bằng phản chứng*.

Cho Q dạng CNF. Phương pháp hợp giải thực hiện quá trình ước lược Q như sau: Thay mỗi cặp nhân tử $(x + B)$, $(x' + C)$ trong Q bằng nhân tử $B + C$, B và C là các công thức logic. Quá trình được lặp lại cho đến khi trong Q không còn tồn tại cặp nhân tử nào có dạng trên hoặc không còn ước lược được nữa. Nếu $Q = \emptyset$ thì ta nói là *hợp giải thành công*, tức là $Q \rightarrow \text{false}$, ngược lại ta nói *hợp giải không thành công*. Khi hợp giải thành công, do $Q \rightarrow \text{false}$ nên ta kết luận Q là hằng sai.

Thuật toán 3.1 (Thuật toán hợp giải)

```

1  Algorithm  Unif(H)
2  Input      CNF(H)
3  Output     unif(W)
4  Begin
5      W := H;
6      while there are terms (x + B) and
7          (x' + C) in W do
8          delete (x + B) in W;
9          delete (x' + C) in W;
10         (B + C) to W
11     endwhile
12     return W;
13 End Unif

```

Định lý 3.1

Các bài toán chứng minh tautology sau đây thuộc lớp NPC:

(i) Chứng minh theo phương pháp Vương Hao.

(ii) Chứng minh theo phương pháp hợp giải.

(iii) Chứng minh trực tiếp theo phương pháp CNF.

Phép gán trị đẳng thức và bảng chân lý của quan hệ

Định nghĩa 3.3

Cho quan hệ r với tập n thuộc tính. Quy ước rằng mỗi miền trị d_i của thuộc tính i , $1 \leq i \leq n$, có chứa ít nhất hai phần tử. Với mỗi cặp bộ

$$u = (u_1, u_2, \dots, u_n) \text{ và}$$

$$v = (v_1, v_2, \dots, v_n)$$

trong r ta xây dựng một vector Boole $t(u, v)$ như sau:

$$t(u, v) = (t_1, t_2, \dots, t_n)$$

$$t_i = (u_i = v_i) ? 1 : 0$$

Với mỗi quan hệ r trên tập n thuộc tính u ta xác định tập các vector Boole

$$T_r = \{t(u, v) : u, v \in r\} \text{ và gọi } T_r \text{ là bảng chân lý của quan hệ } r.$$

Theo định nghĩa ta thấy:

- Nếu r là quan hệ rỗng thì $T_r = \emptyset$.
- Nếu quan hệ r có chứa ít nhất một bộ u nào đó thì do $\alpha(u, u) = e$ nên $e \in T_r$.

Định nghĩa 3.4 [21]

Cho U là tập thuộc tính không rỗng. Mỗi công thức Boole dương (PTBD) trên U là một phụ thuộc Boole dương (PTBD).

Cho quan hệ r trên U . Ta nói quan hệ r thỏa PTBD f và ký hiệu $r(f)$ nếu với mọi cặp bộ u, v trong r :

$$r(f) \Leftrightarrow \forall u, v \in r: f(t(u, v)) = 1.$$

Cho quan hệ r trên U . Ta nói quan hệ r thỏa tập PTBD F và ký hiệu $r(F)$ nếu r thỏa mọi phụ thuộc trong f : $r(F) \Leftrightarrow \forall f \in F: r(f)$

Cho tập thuộc tính U và tập các công thức Boole dương trên U . Ta gọi $p = (U, F)$ là lược đồ quan hệ với tập phụ thuộc Boole dương F .

Định lý 3.2

Cho lược đồ quan hệ $p = (U, F)$ và một phụ thuộc Boole dương $f \in F$. Cho quan hệ r trên U . Khi đó,

(i) Quan hệ r thỏa PTBD f khi và chỉ khi và chỉ khi $T_r \subseteq T_f: r(f) \Leftrightarrow T_r \subseteq T_f$

(ii) Quan hệ r thỏa tập PTBD F khi và chỉ khi và chỉ khi $T_r \subseteq T_F: r(F) \Leftrightarrow T_r \subseteq T_F$

Các phụ thuộc sau đây thuộc lớp PTBD với các công thức Boole tương ứng như sau:

Phụ thuộc	Ký hiệu	Công thức Boole dương
Phụ thuộc hàm	$X \rightarrow Y$	$\wedge X \rightarrow \wedge Y$
Phụ thuộc mạnh	$X(s) \rightarrow Y$	$\vee X \rightarrow \wedge Y$
Phụ thuộc yếu	$X(w) \rightarrow Y$	$\wedge X \rightarrow \vee Y$
Phụ thuộc đối ngẫu	$X(d) \rightarrow Y$	$\vee X \rightarrow \vee Y$
Phụ thuộc đa trị	$X \rightarrow Y$	$\wedge X \rightarrow (\wedge Y) \vee (\wedge(U \setminus X \setminus Y))$

Bảng 2. 3. Các lớp phụ thuộc Boole dương

3.4. Xây dựng thuật toán suy dẫn trong lược đồ quan hệ

- Trong phần này và các phần tiếp theo sẽ phát biểu và chứng minh điều kiện cần và đủ để một công thức logic có thể biểu diễn dưới dạng *hội của các công thức suy dẫn*.

- Ý nghĩa của kết quả này là: phụ thuộc logic có thể được sử dụng để mô tả các loại ràng buộc dữ liệu đa dạng trong thực tiễn. Phụ thuộc hàm và các biến thể, chẳng hạn, phụ thuộc hàm nói lỏng và các phụ thuộc mạnh, yếu và đối ngẫu đều được biểu diễn qua các công thức suy dẫn $X \rightarrow Y$.

3.4.1. Suy dẫn trong lược đồ quan hệ với phụ thuộc hàm

Mỗi phụ thuộc giữa các tập thuộc tính trong cơ sở dữ liệu quan hệ được mô tả qua một biểu thức f . Các biểu thức mô tả sự phụ thuộc có thể có dạng giải tích

Ví dụ: Hai lần rút tiền của cùng một thẻ ATM tại hai địa điểm cách nhau trên 100 km không thể lệch nhau dưới 30 phút.

Cho tập thuộc tính U , phụ thuộc f và quan hệ r trên U . Ta nói quan hệ r thỏa phụ thuộc f , và viết $r(f)$ nếu: $\forall u, v \in r: \gamma(f, u, v) \rightarrow \lambda(f, u, v)$, trong đó γ và λ là các tân từ xác định trên f và cặp thuộc tính u, v .

Lược đồ quan hệ là một cặp $p = (U, \Sigma)$, trong đó U là tập thuộc tính, Σ là tập các phụ thuộc trên U .

Mọi quan hệ và phụ thuộc trong đều được hiểu là các phụ thuộc xây dựng trên tập thuộc tính U cho trước.

Mỗi quan hệ r trên LĐQH p được gọi là một *thể hiện* của LĐQH p . Mọi thể hiện r trên LĐQH p phải thỏa các phụ thuộc Σ . Ta nói, quan hệ r *thỏa tập phụ thuộc* Σ , và viết $r(\Sigma)$, nếu r thỏa mọi phụ thuộc trong Σ ,

$$r(\Sigma) \stackrel{def}{\iff} \forall f \in \Sigma: r(f)$$

PTH truyền thống và các biến thể của PTH truyền thống được đặc tả như sau

PTH (truyền thống)	$\forall u, v \in R: Eq(u.X, v.X) \Rightarrow Eq(u.Y, v.Y)$ $Eq(u.X, v.X) \stackrel{def}{\iff} \forall A \in X: u.A = v.A$
PT mạnh	$\forall u, v \in R: Eq1(u.X, v.X) \Rightarrow Eq(u.Y, v.Y)$ $Eq1(u.X, v.X) \stackrel{def}{\iff} \exists A \in X: u.A = v.A$
PT yếu	$\forall u, v \in R: Eq(u.X, v.X) \Rightarrow Eq1(u.Y, v.Y)$
PT đối ngẫu	$\forall u, v \in R: Eq1(u.X, v.X) \Rightarrow Eq1(u.Y, v.Y)$

Bảng 3. 1. Đặc tả các loại phụ thuộc PTH truyền thống, phụ thuộc mạnh, yếu và đối ngẫu

3.4.2. Các bài toán liên quan đến phụ thuộc dữ liệu

Khái niệm phụ thuộc dữ liệu được nghiên cứu nhằm giải quyết các bài toán sau đây một cách tốt nhất.

Bài toán cập nhật: Cho quan hệ r trên lược đồ $p = (U, \Sigma)$ và hai bộ t và t' trên U . Bài toán cập nhật liên quan đến ba thao tác thêm, xóa và sửa như sau:

Giả thiết: $r(\Sigma)$: quan hệ r thỏa tập phụ thuộc Σ .

Yêu cầu: $r'(\Sigma)$: Sau khi cập nhật (thêm, xóa hoặc sửa) quan hệ kết quả r' vẫn thỏa tập phụ thuộc Σ .

Cụ thể là, nếu trước đó r đã thỏa tập phụ thuộc Σ thì sau khi thực hiện các thao tác thêm, xóa, sửa để thu được quan hệ r' thì r' vẫn phải thỏa Σ . Ta nói Σ là *bất biến đối với các thao tác cập nhật*.

Ta thấy, thao tác sửa tương đương với dãy tuần tự hai thao tác xóa và thêm, do đó chỉ cần yêu cầu Σ bất biến đối với hai thao tác xóa và thêm. Mặt khác, theo định nghĩa về tính thỏa, nếu $r(\Sigma)$ thì với mọi quan hệ $r' \subseteq r$, ta có $r'(\Sigma)$. Từ đây suy ra chỉ cần Σ là bất biến đối với thao tác *thêm*.

Để thêm bộ t vào quan hệ r ta phải kiểm tra điều kiện sau đây.

$$\forall u \in r, \forall f \in \Sigma: \gamma(f, u, t) \rightarrow \lambda(f, u, t)$$

Thủ tục này đòi hỏi độ phức tạp tính toán lớn vì phụ thuộc vào kích thước của các tập Σ và r .

Khó khăn này được giải quyết bước đầu bằng khái niệm *chuẩn hóa*.

Bài toán chuẩn hóa: Thay quan hệ cho trước bằng tập các quan hệ nhỏ hơn, thuận tiện nhất cho việc cập nhật. Với PTH truyền thống, người thiết kế CSDL tách mỗi quan hệ thành các quan hệ thành phần có dạng chuẩn "tốt nhất" cho phép chỉ kiểm tra trên một khóa K của r .

Nếu $\exists u \in r: u.K = t.K$ thì không nạp t ; ngược lại: nạp t vào r .

Bài toán tìm kiếm: Cho quan hệ r trên lược đồ $p = (U, \Sigma)$ và một bộ t trên U . Các yêu cầu tìm kiếm có thể rất đa dạng.

Muốn tăng tốc độ tìm kiếm ta phải vận dụng khái niệm *khóa* như một tập con đủ nhỏ các thuộc tính cho phép xác định duy nhất một bộ trong quan hệ. Khái niệm khóa lại được xây dựng trên cơ sở *bài toán thành viên* dưới đây.

Bài toán thành viên [1]: Cho LĐQH $p = (U, \Sigma)$ và quan hệ r trên p , ta có $r(\Sigma)$, nghĩa là r thỏa mọi phụ thuộc trong Σ . Ngoài Σ , r có thể thỏa các phụ thuộc khác nữa. Ví dụ, $U = ABC$, $\Sigma = \{A \rightarrow B, B \rightarrow C\}$ thì mọi quan hệ r thỏa Σ , r còn thỏa phụ thuộc $A \rightarrow C$.

Bài toán thành viên được phát biểu như sau:

Giả thiết: Cho LĐQH $p = (U, \Sigma)$ và một phụ thuộc g trên U . Kết luận: $g \in \Sigma^+$?

Giải được bài toán thành viên ta có thể giải được bài toán khóa. Trong trường hợp nào thì có thể giải bài toán thành viên bằng công cụ logic, cụ thể là, nếu ta xem các biểu thức mô tả các phụ thuộc như là các biểu thức logic xây dựng trên các biến là các thuộc tính trên U thì trong điều kiện nào $\Sigma \vdash g$ tương đương với $\Sigma \models g$?

Vấn đề này được Armstrong giải quyết trọn vẹn với PTH truyền thống [3] như sau

Cho LĐQH $p = (U, \Sigma)$ và một phụ thuộc g trên U . Ta nói phụ thuộc g được suy dẫn theo quan hệ có không quá hai bộ từ tập phụ thuộc Σ và viết $\Sigma \vdash_2 g$ nếu $\forall r \in REL_2(U): r(\Sigma) \Rightarrow r(g)$.

Định lý tương đương [1], [2]

Cho LĐQH $p = (U, \Sigma)$ và một PT g trên U . Ba mệnh đề sau là tương đương:

- $\Sigma \models g$ (suy dẫn logic)
- $\Sigma \vdash g$ (suy dẫn theo quan hệ)
- $\Sigma \vdash_2 g$ (suy dẫn theo quan hệ có không quá hai bộ).

Armstrong cũng đề xuất ba tiên đề làm cơ sở cho suy dẫn logic đối với PTH truyền thống.

$\forall X, Y, Z \subseteq U:$

- Tiên đề phản xạ: Nếu $Y \subseteq X$ thì $X \rightarrow Y$
- Tiên đề gia tăng: Nếu $X \rightarrow Y$ thì $XZ \rightarrow YZ$
- Tiên đề bắc cầu: Nếu $X \rightarrow Y$ và $Y \rightarrow Z$ thì $X \rightarrow Z$

Theo định lý tương đương, muốn chứng minh $\Sigma \vdash g$ ta có thể vận dụng các tiên đề Armstrong, muốn phủ định $\Sigma \vdash g$ ta chỉ cần xây dựng một quan hệ r với 2 bộ sao cho $r(\Sigma)$ nhưng r không thỏa g .

Các công trình liên quan đến phụ thuộc hàm và các loại phụ thuộc Boole đương trước năm 1992 chỉ quan tâm phép sánh trị đẳng thức Eq được định nghĩa như sau.

$\forall X \subseteq U, \forall u, v \in r(U):$

$$Eq(u.X, v.X) \stackrel{def}{\iff} \forall A \in X: u.A = v.A$$

Năm 1992 nhóm nghiên cứu của Nguyễn Xuân Huy mở rộng khái niệm sánh trị và cung cấp mô hình phụ thuộc Boole đương tổng quát bảo toàn định lý tương đương làm cơ sở cho cơ chế suy dẫn và các bài toán liên quan như bài toán tìm kiếm, bài toán thành viên, xác định khóa... là những vấn đề không thể thiếu trong quản lý CSDL.

3.4.3. Thuật toán suy dẫn

Khái niệm chung

Trong lớp các phụ thuộc Boole dương, các công thức suy dẫn chính là *lớp các phụ thuộc hàm* được Codd hình thức hóa lần đầu tiên vào năm 1970 [15].

Ta định nghĩa $\forall v \in B^n$, $Set(v) = \{A \in U \mid v.A = 1\}$, khi đó giữa B^n và 2^U tồn tại một song ánh.

Nếu coi X là một hội các biến logic thì với mỗi phép gán trị v , ta có $X(v)=1$ khi và chỉ khi $Set(v) \supseteq X$.

Cho $f: X \rightarrow Y$ là một CTSD trên U , với mỗi phép gán trị v , ta có $f(v)=1$ khi và chỉ khi $Set(v) \supseteq X \Rightarrow Set(v) \supseteq Y$.

Ký hiệu $I(U)$ là tập các CTSD trên tập biến U .

Ta gọi một *hội suy dẫn* là tập F các CTSD, $F = \{f_1, \dots, f_m\}$. Với mỗi CTSD $f: X \rightarrow Y$, ta biết $f(e) = X(e) \rightarrow Y(e) = 1 \rightarrow 1 = 1$, nên các CTSD là các công thức Boole dương, tức là $I(U) \subseteq P(U)$.

Hội suy dẫn

Ta cũng biết mọi công thức logic đều có thể biểu diễn dưới dạng chuẩn tuyền (hội). Nói cách khác, mỗi bảng $T \subseteq B^n$ đều ứng với một công thức logic dạng chuẩn tuyền (hội). Vấn đề biểu diễn một công thức logic qua một tập các phép toán và hằng logic cho trước chưa có lời giải tổng quát [2]. Các phần trình bày dưới đây liên quan đến bài toán sau.

Bài toán

Xác định điều kiện cần và đủ để có thể biểu diễn một công thức logic dưới dạng hội suy dẫn [2].

Bổ đề 3.1

(Bổ đề về tính đóng của phép & trong T_f [2]) *Với mỗi CTSD f trên U , T_f chứa các phép gán trị đơn vị e , gán trị không z và đóng với phép &.*

Với mỗi HSD F trên U , vì bảng chân lý T_F của F là giao của các bảng chân lý của các công thức thành viên nên ta có các hệ quả sau đây.

Hệ quả 3.1

Với mỗi HSD F trong $I(U)$, T_F chứa các phép gán trị đơn vị e , gán trị không z và đóng với phép &.

Hệ quả dưới đây thiết lập điều kiện cần và đủ để một bảng $T \subseteq B^n$ là một bảng chân lý của một hội suy dẫn.

Hệ quả 3.2

Bảng $T \subseteq B^n$ là bảng chân lý của một HSD khi và chỉ khi T chứa các phép gán trị đơn vị e , gán trị không z và đóng với phép &.

3.5. Xây dựng thuật toán tìm bao đóng với phụ thuộc Boole dương tổng quát

Gọi U là tập các thuộc tính và Ψ là tập phụ thuộc logic trên U , $X \subseteq U$. Ta định nghĩa bao đóng X^+ của X là tập các thuộc tính sau:

$$X^+ = \{a \in U \mid X \models a \in \Psi^+\}$$

Bao đóng của tập thuộc tính X là tập toàn bộ các thuộc tính phụ thuộc vào X .

Thuật toán tìm bao đóng của tập các thuộc tính

Ý tưởng thuật toán: Để tìm bao đóng X^+ của X , ta khởi tạo $X^+ = X$, sau đó vận dụng thuật toán thành viên để xét cho từng thuộc tính $a \in U - X$, nếu $X \models a \in \Psi^+$ thì bổ sung a vào X^+ . Với Ψ đã được đưa về dạng chuẩn không dư thì điều kiện $X \models a \in \Psi^+$ sẽ cho $CNF(\Psi(X \models a)) = \Psi X a'$, do đó sẽ tương đương với điều kiện $Unif(\Psi X a') = \emptyset$.

Thuật toán 3.2 (Thuật toán tìm bao đóng trong lớp các PTLG)

```

1  Algorithm  LClosure(X,  $\Psi$ )
2  Input      LNorm( $\Psi$ );  $X \subseteq U$ 
3  Output     $X^+ = \{a \in U \mid X \vdash a \in \Psi^+\}$ 
4  Begin
5       $Y := X$ ;  $V := U - X$ ;
6      for each attribute  $a$  in  $V$  do
7          if  $\text{Unif}(\Psi X a') = \emptyset$  then
8              add  $a$  to  $Y$ ;
9          endif;
10     endfor;
11     return  $Y$ ;
12 End  LClosure

```

(Thuật toán được đề xuất trong CT2 - Danh mục công trình công bố của NCS).

Mệnh đề 3.1

Bài toán tìm bao đóng trong lớp các phụ thuộc logic thuộc lớp NPC.

3.6. Xây dựng thuật toán tìm khóa với phụ thuộc Boole dương tổng quát

Khóa là tập con các thuộc tính đủ nhỏ và xác định đơn trị một bộ trong quan hệ.

Cho F là tập phụ thuộc logic trên U . Tập $K \subseteq U$ được gọi là khóa nếu thỏa:

- $K^+ = U$,
- $\forall a \in K: (K - a)^+ \neq U$.

Nếu K chỉ thỏa điều kiện thứ nhất thì K được gọi là *siêu khóa*.

Thuật toán sau thực hiện các bước tìm một khóa trên tập thuộc tính U và tập PTLG F cho trước:

Thuật toán tìm khóa

Xuất phát từ một siêu khóa K tùy ý, lần lượt duyệt từng thuộc tính a trong K , nếu $(K - a)^+ = U$ thì loại a khỏi K .

Thuật toán 3.3 (Tìm khóa K trong lớp các PTLG)

```

1  Algorithm  LKey(U, Lnorm( $\Psi$ ))
2  Input      U, Lnorm( $\Psi$ )
3  Output     $K \subseteq U$ :
4   $K^+ = U$ 
5   $\forall a \in K: (K - a)^+ \neq U$ 
6  Begin
7       $K := U$ ;
8      for each attribute  $a$  in  $U$ 
9  do
10         if  $\text{Unif}(\Psi(K - a)a') = \emptyset$ 
11     then
12         delete  $a$  from  $K$ ;
13     endif
14     endfor
15     return  $K$ ;
16 End  LKey

```

Mệnh đề 3.2

Bài toán tìm khóa trong lớp các phụ thuộc logic thuộc lớp NPC.

3.7. Kết luận chương 3

Trong chương 3, luận án trình bày kết quả đề xuất:

- Tiếp cận bài toán suy dẫn trong lớp các phụ thuộc logic theo ba phương pháp: Phương pháp Vương Hạo, phương pháp gián tiếp theo hợp giải, phương pháp trực tiếp
- Thuật toán tìm bao đóng của tập thuộc tính trong lược đồ quan hệ với phụ thuộc hàm, thuật toán tìm bao đóng của tập thuộc tính trong lược đồ quan hệ với phụ thuộc Boole dương tổng quát, thuật toán tìm khóa với phụ thuộc hàm và thuật toán tìm khoá với phụ thuộc Boole dương tổng quát được xây dựng.

KẾT LUẬN VÀ HƯỚNG PHÁT TRIỂN

I. Các kết quả đạt được của luận án

Một số đóng góp mới của luận án tập trung vào ba nhóm kết quả:

(1) Xây dựng hàm biến đổi lambda, dựa vào hàm lambda được xây dựng, kết hợp giữa phụ thuộc hàm xấp xỉ và phụ thuộc Boole dương tổng quát, luận án đề xuất các phụ thuộc dữ liệu mới bao gồm: *phụ thuộc xấp xỉ tổng quát, phụ thuộc Boole dương xấp xỉ, phụ thuộc Boole dương xấp xỉ tổng quát và phụ thuộc yếu xấp xỉ.*

(2) Xây dựng mối quan hệ giữa các loại phụ thuộc logic: Chứng minh sự tương đương giữa phụ thuộc Boole dương tổng quát và phụ thuộc hàm nói lỏng, sự tương đương giữa phụ thuộc Boole dương tổng quát và phụ thuộc Boole dương xấp xỉ, sự tương đương giữa phụ thuộc Boole dương tổng quát và phụ thuộc Boole dương xấp xỉ tổng quát

(3) Xây dựng một số thuật toán giải các bài toán đặc trưng sau:

- Sử dụng phương pháp Vương Hạo, CNF và hợp giải để giải bài toán suy dẫn.
- Xây dựng và đánh giá độ phức tạp tính toán cho các thuật toán tìm bao đóng của một tập thuộc tính cho phụ thuộc hàm và phụ thuộc Boole dương tổng quát.
- Xây dựng và đánh giá độ phức tạp tính toán cho các thuật toán tìm khóa cho phụ thuộc hàm và phụ thuộc Boole dương tổng quát.

Kết quả của đề tài có thể cung cấp cho người thiết kế và quản trị cơ sở dữ liệu một khung nhìn tổng quan về mối quan hệ giữa các phụ thuộc logic, qua đó, có thể lựa chọn loại phụ thuộc logic phù hợp với nhu cầu thực tiễn.

II. Những đóng góp mới của luận án

(1) Đề xuất các dạng phụ thuộc mới là phụ thuộc xấp xỉ tổng quát, phụ thuộc Boole dương xấp xỉ, phụ thuộc boole dương xấp xỉ tổng quát và phụ thuộc yếu xấp xỉ.

(2) Xây dựng mối quan hệ giữa các loại phụ thuộc logic, giữa phụ thuộc Boole dương tổng quát và phụ thuộc hàm nói lỏng, phụ thuộc Boole dương tổng quát và phụ thuộc Boole dương xấp xỉ, phụ thuộc Boole dương tổng quát và phụ thuộc Boole dương xấp xỉ tổng quát

(3) Đề xuất qui trình giải bài toán suy dẫn theo ba tiếp cận hình thức. Xây dựng và đánh giá thuật toán tìm bao đóng, thuật toán tìm khóa của một tập thuộc tính cho lớp các phụ thuộc logic.

III. Hướng phát triển tiếp theo cho luận án:

Hiện nay, có khá nhiều loại phụ thuộc logic phục vụ cho việc thiết kế cơ sở dữ liệu đã được đề xuất. Để xây dựng nên một mô hình tổng quan cho các loại phụ thuộc logic, ta phải xác định được mối quan hệ giữa chúng. Với mục tiêu đó, luận án đã chứng minh sự tương đương giữa một số loại phụ thuộc logic như: phụ thuộc Boole dương tổng quát, phụ thuộc sai khác tổng quát, phụ thuộc yếu tổng quát, phụ thuộc xấp xỉ và xây dựng nên lớp các phụ thuộc logic. Nếu tiếp tục mở rộng lớp các phụ thuộc logic này bằng cách thu nạp thêm một số loại phụ thuộc logic khác nữa thì mức tổng quát của lớp các phụ thuộc logic càng có độ tin cậy cao hơn. Đây cũng là hướng phát triển tiếp theo của NCS khi định hướng tiếp tục theo đuổi đề tài này:

Tiếp tục tìm hiểu một số loại phụ thuộc mới được nghiên cứu như: phụ thuộc hàm mềm, phụ thuộc đối sánh, phụ thuộc có điều kiện, phụ thuộc theo mẫu, phụ thuộc tuần tự,... và một số phụ thuộc được đề xuất trong tương lai. Tìm ra mối quan hệ giữa các loại phụ thuộc mới này với các phụ thuộc đã nêu trong luận án.

Bổ sung các phụ thuộc mới được các nhóm nghiên cứu vào lớp các phụ thuộc logic mà luận án đã xây dựng.

Nghiên cứu thêm một số bài toán như bài toán kinh điển và xây dựng nên các chuẩn cho lớp các phụ thuộc logic.

Xây dựng phần mềm ứng dụng, giải quyết bài toán thực tế cho lớp các phụ thuộc logic và giải quyết bài toán trong lớp này theo phương pháp hợp giải.

DANH MỤC CÔNG TRÌNH ĐÃ CÔNG BỐ

- [CT1] Nguyễn Xuân Huy, Trương Thị Thu Hà, Nguyễn Thị Vân (2016), “Tương quan giữa phụ thuộc hàm xấp xỉ và phụ thuộc Boole dương tổng quát trong cơ sở dữ liệu quan hệ”, *Kỷ yếu Hội thảo quốc gia lần thứ XIX: Một số vấn đề chọn lọc của Công nghệ thông tin và Truyền thông*, NXB Khoa học và Kỹ thuật, ISBN: 978-604-67-0781-3, Hà Nội, tr.361-365.
- [CT2] Trương Thị Thu Hà, Nguyễn Thị Vân, Nguyễn Xuân Huy (2016), “Thuật toán xác định bao đóng và khóa theo tiếp cận hợp giải trong lớp các phụ thuộc logic”, *Tạp chí Thông tin và TT, Chuyên san Các công trình nghiên cứu, phát triển và ứng dụng Công nghệ thông tin và truyền thông*, ISSN 1859-3526, kỳ 3, tập V-2, số 16(36), tr.50-57.
- [CT3] Nguyễn Thị Vân, Trương Thị Thu Hà, Nguyễn Xuân Huy (2017), “Phụ thuộc trong cơ sở dữ liệu theo tiếp cận logic”, *Kỷ yếu Hội thảo quốc gia lần thứ XX: Một số vấn đề chọn lọc của Công nghệ thông tin và Truyền thông*, Qui Nhơn, 23-24/11/2017, NXB Khoa học và Kỹ thuật, ISBN: 978-604-67-1009-7, Hà Nội, tr.260-265. [41] [42] [43] [44] [45] [46] [47] [48, 49, 50]
- [CT4] Nguyễn Xuân Huy, Nguyễn Thị Vân (2018), “Bài toán suy dẫn logic và ứng dụng trong cơ sở tri thức”, *Kỷ yếu Hội thảo quốc gia lần thứ XIX: Một số vấn đề chọn lọc của Công nghệ thông tin và Truyền thông*, NXB Khoa học và Kỹ thuật, ISBN: 978-604-67-1104-9, Thanh Hóa, tr.27-31.
- [CT5] Nguyễn Xuân Huy, Nguyễn Thị Vân, Trương Thị Thu Hà (2020), “Quan hệ giữa phụ thuộc hàm nói lỏng và phụ thuộc Boole dương tổng quát”, *Tạp chí Chuyên san Công nghệ thông tin và truyền thông*, ISSN 1859-3526, số 1, tháng 6-tập 2020, tr.44-58.
- [CT6] Nguyen Xuan Huy, Nguyen Thi Van (2022), “Lambda functions and approximate generalized positive Boolean dependencies”, *Journal on Information Technologies & Communications*, ISSN 1859-3526, Vol 2022, No 2, page.112-118.