**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

## LUẬN ÁN TIẾN SĨ HỆ THỐNG THÔNG TIN

**Hà Nội - 2023**

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

## LUẬN ÁN TIẾN SĨ HỆ THỐNG THÔNG TIN

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

**Xác nhận của Học viện Khoa học và Công nghệ**

**Người hướng dẫn**
*(Ký, ghi rõ họ tên)*

**PGS. TSKH. Nguyễn Xuân Huy**

**Hà Nội - 2023**

# LỜI CAM ĐOAN

Tôi xin cam đoan đây là công trình nghiên cứu của tôi và được hướng dẫn bởi Thầy PGS. TSKH Nguyễn Xuân Huy. Những kết quả trong luận án có nghiên cứu chung với các đồng tác giả đều được sự đồng ý của các tác giả trước khi được sử dụng trong luận án.

Những kết quả được trình bày trong công trình đều trung thực và không được sao chép từ các công trình được công bố khác. Nếu xảy ra việc gian lận trong luận án này tôi sẽ chịu hoàn toàn mọi trách nhiệm.

Tôi xin cam đoan mọi giúp đỡ trong quá trình thực hiện luận án đã được NCS cảm ơn, mọi thông tin về nguồn trích dẫn và chú thích rõ ràng trong luận án đều và đã được NCS nêu rõ nguồn gốc.

*Hà nội, Ngày 26 tháng 9 năm 2023*

**Tác giả**

**Nguyễn Thị Vân**

## LỜI CẢM ƠN

Trong thời gian thực hiện nghiên cứu và hoàn thiện luận án, NCS nhận được nhiều sự giúp đỡ, tạo điều kiện từ các Thầy, Cô, các nhà nghiên cứu và đồng nghiệp. NCS xin được bày tỏ lời cảm ơn tự đáy lòng đến những người đã trợ giúp và chia sẻ khó khăn trong suốt thời gian qua.

Trước hết, NCS xin được bày tỏ lòng biết ơn nhất đến Thầy PGS. TSKH Nguyễn Xuân Huy, người luôn đồng hành và tận tình giúp đỡ NCS trong suốt chặng đường nghiên cứu, định hướng cho NCS.

NCS xin trân trọng cảm ơn các nhà khoa học các Thầy và Cô trong Viện công nghệ Thông tin – Học viện Khoa học và Công nghệ đã khuyến khích, tạo điều kiện thuận lợi và có nhiều ý kiến đóng góp quý báu trong chặng đường NCS nghiên cứu và thực hiện để hoàn thành được luận án.

NCS xin trân trọng cảm ơn Lãnh đạo Viện Công nghệ Thông tin, Học viện Khoa học và Công nghệ đã khuyến khích và tạo những điều kiện tốt nhất để NCS có được môi trường nghiên cứu tốt nhất. NCS xin chân thành gửi lời cảm ơn tới các Phòng ban của Học viện Khoa học và Công nghệ đã nhiệt tình giúp đỡ, tạo những điều kiện tốt nhất cho NCS trong suốt quá nghiên cứu và thực hiện luận án.

NCS xin chân thành cảm ơn thầy giáo PGS.TS Đặng Văn Đức, PGS.TS Nguyễn Long Giang đã tận tình chỉ bảo và là tấm gương về nghiên cứu khoa học cho NCS trên bước đường học tập. GS. TS Nguyễn Thanh Thuỷ, PGS.TS Ngô Quốc Tạo, GS.TS Từ Minh Phương, PGS.TS Trịnh Đình Thắng, GS.TS Lê Hoài Bắc, PGS.TS Đoàn Văn Ban, TS Nguyễn Duy Phương…là những người Thầy mà NCS đã học hỏi được rất nhiều, và đã có nhiều ý kiến quý báu giúp NCS hoàn thiện luận án của mình. NCS cũng xin bày tỏ lời cảm ơn đến anh chị, bạn bè và TS Trương Thị Thu Hà đã đồng hành cùng NCS trong quá trình học tập.

Cuối cùng, NCS biết ơn những người thân yêu trong gia đình đã luôn bên cạnh ủng hộ và dành những lời động viên cho NCS trong suốt quá trình NCS hoàn thành luận án.

*Hà nội, Ngày 26 tháng 9 năm 2023*

**Tác giả**

**Nguyễn Thị Vân**

# MỤC LỤC

## DANH MỤC THUẬT NGỮ, CÁC CHỮ VIẾT TẮT

| **Ký hiệu** | **Mô tả** |
|---|---|
| CSDL | Cơ sở dữ liệu |
| CSDLQH | Cơ sở dữ liệu quan hệ |
| CTB | Công thức Boole |
| CTBD | Công thức Boole dương |
| Đpcm | Điều phải chứng minh |
| CNF | Công thức logic dạng chuẩn hội |
| GT | Giả thiết |
| KL | Kết luận |
| LClosure | Thuật toán bao đóng trong phụ thuộc logic |
| LĐQH | Lược đồ quan hệ |
| LĐXX | Lược đồ xấp xỉ |
| NPC | Lớp thuật toán NP- đầy đủ |
| PTBD | Phụ thuộc Boole dương |
| PTBDTQ | Phụ thuộc Boole dương tổng quát |
| PTH | Phụ thuộc hàm |
| PTHXX | Phụ thuộc hàm xấp xỉ |
| PTHM | Phụ thuộc hàm mạnh |
| PTMTQ | Phụ thuộc mạnh tổng quát |
| PTY | Phụ thuộc yếu |
| PTYTQ | Phụ thuộc yếu tổng quát |
| Unif | Thuật toán hợp giải |
| PTNL | Phụ thuộc nới lỏng |
| PTHNL | Phụ thuộc hàm nới lỏng |
| PTBDXX | Phụ thuộc Boole dương xấp xỉ |
| PTBDXXTQ | Phụ thuộc Boole dương xấp xỉ tổng quát |
| PTYXX | Phụ thuộc yếu xấp xỉ |
| HSK | Hàm sai khác |
| PTHSK | Phụ thuộc hàm sai khác |

## DANH MỤC CÁC KÝ HIỆU

| **Ký hiệu** | **Diễn giải** |
|---|---|
| $\lambda$ | Tân từ |
| $\rho_a$ | Độ đo của thuộc tính *a* |
| $\delta_a$ | Độ sai khác của thuộc tính *a* |
| $\alpha_a$ | Phép sánh trị tổng quát trên thuộc tính *a* |
| $\neg a$, $a'$ | Phủ định của *a* |
| $(U, F)$ | Lược đồ quan hệ trên tập thuộc tính *U* và tập ràng buộc *F* |
| $\phi_X$ | Hàm sai khác $\phi$ trên tập thuộc tính X |
| $\|\|X\|\|$, $\|X\|$ | Lực lượng của *X* |
| $\vdash$ | Phép suy dẫn theo quan hệ |
| $\models$, $\rightarrow$ | Phép suy dẫn theo logic |
| $\vdash_2$ | suy dẫn theo quan hệ có không quá 2 bộ |
| $a$, $a_i$ | Thuộc tính |
| $d_a$ | Miền trị của thuộc tính *a* |
| $f$, $g$ | Phụ thuộc logic |
| $F$, $G$ | Tập các phụ thuộc logic $F = \{f_1, f_2, \ldots, f_k\}$ |
| $F^+$ | Bao đóng của tập phụ thuộc *F* được suy dẫn theo logic |
| $F^*$ | Bao đóng của tập phụ thuộc *F* được suy dẫn theo quan hệ |
| $L(U)$ | Tập các công thức Boole xây dựng trên *U* |
| $P(U)$ | Tập các công thức Boole dương trên *U* |
| $r$ | Quan hệ *r* |
| $r(f)$ | Quan hệ *r* thỏa phụ thuộc *f* |

| **Ký hiệu** | **Diễn giải** |
|---|---|
| $r(F)$ | Quan hệ $r$ thỏa tập phụ thuộc $F$ |
| $\mathbb{R}^+$ | Tập số thực không âm |
| $REL(U)$ | Tập các quan hệ trên tập thuộc tính $U$ |
| $SAT(F)$ | Tập các quan hệ trên $U$ thỏa tập ràng buộc $F$ |
| $t, u, v,...$ | Bộ |
| $t.a$ | Giá trị của bộ $t$ trên thuộc tính $a$ |
| $t.X$ | Bộ $t$ trên tập thuộc tính $X$ |
| $T_f$ | Bảng chân lý của phụ thuộc hàm $f$ |
| $T_F$ | Bảng chân lý của tập phụ thuộc hàm $F$ |
| $Tr, Vr$ | Bảng trị của quan hệ $r$ |
| $U$ | Tập thuộc tính |
| $X \cup Y$ | Phép hợp hai tập $X$ và $Y$ |
| $X \cap Y$ | Phép giao hai tập $X$ và $Y$ |
| $X \vee Y, X+Y$ | Tuyển (tổng) logic $X$ và $Y$ |
| $X \wedge Y, XY$ | Hội (tích) logic $X$ và $Y$ |
| $X \setminus Y, X - Y$ | Hiệu hai tập X và Y |
| $X, Y$ | Tập các thuộc tính |
| $X^+$ | Bao đóng của tập $X$ |
| $\Sigma=\{\gamma_{a1},\gamma_{a2},\ldots,\gamma_{an}\}$ | Tập các phụ thuộc sai khác |
| $\mathfrak{B}$ | Tập các giá trị Boole |
| $\mathcal{D}$ | Miền trị của tập các thuộc tính trong $U$ |
| $\mathcal{L}$ | Tập các phụ thuộc logic |
| $\mathbb{Z}^+$ | Tập số nguyên không âm |

# DANH MỤC HÌNH VÀ CÁC BẢNG

# MỞ ĐẦU

## 1. Tính cấp thiết của đề tài luận án

Với sự tiến bộ trong lĩnh vực khoa học và kỹ thuật, vai trò của các phụ thuộc dữ liệu trong quá trình thiết kế và sử dụng dữ liệu ngày càng trở nên quan trọng hơn. Do đó, việc phát triển các dạng phụ thuộc dữ liệu đa dạng đang là một chủ đề nghiên cứu đang thu hút sự quan tâm lớn và có ý nghĩa đối với thực tế. Trong lĩnh vực đảm bảo tính nhất quán và ngữ nghĩa, đây cũng được coi là một trong những mục tiêu quan trọng khi thực hiện khai thác tri thức từ các nguồn dữ liệu đa dạng.

*Tình hình nghiên cứu trên thế giới*

Phụ thuộc dữ liệu đã được nghiên cứu trong nhiều công trình. Điển hình là năm 1970, Codd [1] [2] giới thiệu khái niệm đầu tiên về CSDL quan hệ và khái niệm phụ thuộc hàm (PTH) để phản ánh ngữ nghĩa của dữ liệu tồn tại trong thế giới thực. Cùng với sự phát triển mạnh mẽ của lớp phụ thuộc hàm, một số phụ thuộc dữ liệu biến thể từ phụ thuộc hàm, cũng như hệ tiên đề cho lớp các phụ thuộc - tức là đặt nền móng cho cơ sở lý thuyết về phụ thuộc dữ liệu, cũng đã được giới thiệu bao gồm: phụ thuộc đối ngẫu, phụ thuộc yếu, phụ thuộc mạnh của nhóm nghiên cứu J. Demetrovics và O. Gyepesi đề xuất năm 1983 [3]. Từ năm 1977 đến năm 2003, R. Fagin và Zaniolo và một số nhóm nghiên cứu khác đã đề xuất phụ thuộc đa trị cùng một số ứng dụng của nó [4] [5] [6], phụ thuộc đa trị mở rộng tập trị không chỉ nhận hai giá trị {0, 1} mà bao gồm *n* giá trị thực nằm trong trong khoảng [1, 0].

Năm 1981 - 1985, Berman, Delobel và cộng sự [7] [8] đã phát triển khái niệm phụ thuộc hàm thành khái niệm phụ thuộc Boole dương (PTBD), bao gồm những ràng buộc dữ liệu được biểu diễn bằng các công thức Boole dương (CTBD), nhưng vẫn giữ nguyên phép sánh trị đẳng thức. Sau đó, nhóm nghiên cứu Nguyễn Xuân Huy, Lê Thị Thanh [9] mở rộng phụ thuộc Boole dương thành phụ thuộc Boole dương tổng quát (PTBDTQ), phụ thuộc Boole dương theo nhóm bộ và phụ thuộc Boole dương đa trị bằng việc thay thế phép sánh trị đẳng thức thành phép sánh trị tổng quát $\alpha$ thỏa ba tính chất đối xứng, phản xạ, bộ phận và chứng minh định lý tương đương cho phép thay việc kiểm tra phép suy dẫn theo dữ liệu bằng việc kiểm tra theo logic hình thức.

Năm 1995 Jyrki Kivinen và các đồng nghiệp đề xuất khái niệm phụ thuộc hàm xấp xỉ [10]. Phụ thuộc hàm xấp xỉ được phát triển từ khái niệm phụ thuộc hàm được phát biểu như sau: ta nói rằng $X \rightarrow Y$ đúng trong quan hệ $r$ nếu mọi cặp bộ $u$, $v \in R$ thỏa $u.X = v.X$ thì $u.Y = v.Y$ cũng thoả. Nếu $r$ không thỏa mãn phụ thuộc hàm $X \rightarrow Y$ nhưng sau khi xóa đi $n$ bộ trong $r$ ta thu được $r'$ thỏa mãn phụ thuộc hàm $X \rightarrow Y$ thì quan hệ $r$ thỏa mãn phụ thuộc hàm $X \rightarrow Y$ với độ xấp xỉ $\frac{n}{\#r}$, trong đó $\#r$ là tổng số bộ của quan hệ *r lúc đầu*. Tiếp theo, các nhóm nghiên cứu Hultala *Y* và các đồng nghiệp [11], Ronald S. K. và Janes J. L. [12] đã phát triển thêm một số thuật toán cho loại phụ thuộc hàm xấp xỉ này.

Gần đây, xuất phát từ các phụ thuộc dữ liệu truyền thống có nhiều công bố nghiên cứu về các phụ thuộc dữ liệu mở rộng cho các loại dữ liệu khác nhau được thực hiện bởi nhiều tác giả. Cụ thể, năm 2004, Ilyas [12] và đồng nghiệp đã phát triển loại phụ thuộc hàm mới được gọi là phụ thuộc hàm mềm, đây là phụ thuộc hàm mà trị của tập thuộc tính trên $X$ xác định trị của tập thuộc tính trên $Y$ với độ không chắc chắn cho trước. Phụ thuộc hàm có điều kiện được Bohannon [13] [14] [15], cùng các đồng nghiệp đề xuất năm 2007 để làm sạch dữ liệu. Với dữ liệu mờ, phụ thuộc hàm đối sánh [16] [17] [18] đã lần lượt được nghiên cứu. Phụ thuộc hàm độ đo [19] được đề xuất năm 2009. Phụ thuộc tuần tự được Golab và đồng nghiệp giới thiệu để khái quát dữ liệu theo thứ tự và thể hiện được các mối quan hệ giữa các thuộc tính có thứ tự [20] .

Đầu năm 2011, nhóm nghiên cứu Song S. và Chen L. đề xuất phụ thuộc sai khác (PTSK) để giải quyết các vấn đề như đảm bảo tính toàn vẹn, tối ưu truy vấn [20] Bản chất của phụ thuộc sai khác là mở rộng phép đối sánh trên các thuộc tính trong phụ thuộc hàm. Nếu $X \rightarrow Y$ là một phụ thuộc hàm và $\gamma_X$, $\gamma_Y$ là hai hàm sai khác trên $X$ và $Y$ tương ứng thì  quan hệ $r$ thỏa phụ thuộc sai khác $\gamma_X \rightarrow \gamma_Y$ khi và chỉ khi với hai dòng bất kì $u, v \in r$, $u.X$ và $v.X$ thỏa hàm sai khác $\gamma_X$ kéo theo $u.Y$ và $v.Y$ thỏa hàm sai khác $\gamma_Y$. Ở đây ta hiểu hàm sai khác mô tả mức độ khác nhau của những giá trị nằm trong miền trị của các thuộc tính.

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 của nhóm nghiên cứu Song S., Chen L., Yu P.S - 2013) [21]; hay việc phân tích các ràng buộc theo cấu trúc mẫu của Baixeries J., Kaytoue M., Napoli A. - 2015 [22] mở rộng phụ thuộc hàm xấp xỉ với dữ liệu lớn phân tán của Li W., Chen Q., Li Z., Yin Z. – 2016 [23] cũng được các nhóm đề xuất và ứng dụng vào thực tế. Gần đây nhất, [24] các tác giả Vincenzo Deufemia, Loredana Caruccio, Giuseppe Polese (2016) đã tổng kết 35 loại phụ thuộc mở rộng của các nhóm nghiên cứu trên thế giới. Tuy nhiên, ngoài các phụ thuộc mà các tác giả gọi là *nhóm các phụ thuộc hàm nới lỏng* vẫn còn một số phụ thuộc cơ bản chưa được đề cập tới như các phụ thuộc logic và phụ thuộc theo nhóm bộ, phụ thuộc đa trị.

*Tình hình nghiên cứu trong nước*

Tại Việt Nam, song song với các nhóm nghiên cứu trên thế giới, 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 và tạo điều kiện đơn giản hoá hơn trong quá trình truy vấn và tìm kiếm dữ liệu. Ví dụ như các tác giả Đàm Gia Mạnh [25] (25); Vũ Ngọc Loãn [26], Bùi Đức Minh [27], Nguyễn Hoàng Sơn [28], Lương Nguyễn Hoàng Hoa [29] [30] Lê Xuân Vinh [31]; Trương Thị Thu Hà [32]… đã khảo sát phụ thuộc yếu cùng với phụ thuộc sai khác và phụ thuộc đối ngẫu… và các lớp Boole dương tổng quát dưới nhiều góc nhìn khác nhau.

Trên cơ sở khảo sát và phân tích các công trình về phụ thuộc dữ liệu, những công trình nghiên cứu liên quan đã được các tác giả công bố trong nước và ngoài nước, tác giả nhận thấy một số đặc điểm như sau:

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

(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 các loại 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. Tuy vậy, bên cạnh đó, một số nghiên cứu hiện đại, sâu hơn xuất hiện khá nhiều trong thời gian qua của các nhà nghiên cứu về lý thuyết cơ sở dữ liệu theo hướng tổ hợp như tập đóng cũng như khóa và 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.

Ngoài ra, một số nghiên cứu tập trung vào các biến thể dựa trên phụ thuộc hàm. Các biến thể này được xây dựng và phát triển dựa trên cơ sở thay phép đối sánh đẳng thức (=) giữa hai giá trị của một miền dữ liệu bằng phép xấp xỉ với ngưỡng cho trước. Nếu ta thay phép sánh trị đẳng thức bằng một tân từ đủ để bao quát các phép xấp xỉ thì ta sẽ thu được một loại phụ thuộc mới có thể coi là một lớp biến thể của phụ thuộc hàm. Nếu ta thay tiếp công thức dạng suy dẫn $X \rightarrow Y$ bằng một công thức logic tổng quát thì ta thu được một phụ thuộc logic tổng quát.

Nối tiếp mạch nghiên cứu của các nghiên cứu trong nước và nước ngoài với mục đích cung cấp cho bên khai thác và quản trị dữ liệu một công cụ biểu diễn dữ liệu sát với thế giới thực theo một cấu trúc nhất định, tạo một khung nhìn tổng quát và thống nhất giữa các loại ràng buộc được sử dụng trong việc khai thác và tổ chức cơ sở dữ liệu, cập nhật và truy vấn các lượng lớn về dữ liệu đảm bảo tính toàn vẹn dữ liệu nhưng dễ dàng duy trì, cập nhật, an toàn và hiệu quả.

*Luận án tiếp tục nghiên cứu và mở rộng phụ thuộc Boole dương tổng quát để thu được các lớp 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 xây dựng một mô hình tổng quát nhất cho các lớp phụ thuộc nói trên bằng cách chứng minh mối tương quan giữa các loại phụ thuộc và cố gắng đề xuất một lớp phụ thuộc thống nhất bao gồm các lớp phụ thuộc dữ liệu hiện đang được các nhà khoa học quan tâm trong nghiên cứu và triển khai.

Đây cũng chính là nội dung nghiên cứu của luận án.

Vì vậy, để có thể giải quyết được các vấn nghiên cứu được nêu ở trên, trong luận án NCS mở rộng ba vấn đề quan trọng sau đây:

Thứ nhất: Tổng quát công thức logic $X \rightarrow Y$ bằng một công thức Boole dương thể hiện các ràng buộc dữ liệu trên các thuộc tính trong lược đồ quan hệ.

Thứ hai: Phát triển các phép sánh trị truyền thống (=, ≠, >, ≥, < , ≤) thành phép sánh trị α tổng quát [33] .

Thứ ba: Xây dựng hàm *lambda* định lượng cho các thuộc tính trong LĐQH để lượng hóa tập các thuộc tính số và hỗn hợp.

Việc mở rộng phụ thuộc dữ liệu theo ba hướng trên cho phép mang lại các lợi ích sau đây:

Cung cấp cho người thiết kế, người khai thác, người quản trị cơ sở dữ liệu một công cụ biểu diễn dữ liệu sát với thế giới thực.

Tạo một khung nhìn thống nhất giữa các loại ràng buộc trong cơ sở dữ liệu.

Tổ chức dữ liệu cũng như việc truy vấn và cập nhật khối lượng lớn dữ liệu an toàn, thuận lợi và hiệu quả hơn. Đáp ứng được các nhu cầu cần thiết về quản lý dữ liệu trong cơ quan, tổ chức và xí nghiệp.

## 2. Mục tiêu nghiên cứu

Dựa trên kết quả phân tích các vấn đề còn tồn tại của các công trình có liên quan, ***mục tiêu nghiên cứu chính của luận án*** là:

Phát triển các lớp 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 xây dựng một mô hình tổng quát nhất cho các lớp phụ thuộc về các loại phụ thuộc mới và mở rộng các khái niệm liên quan đến phụ thuộc trong lĩnh vực nghiên cứu tương ứng. Các kết quả sẽ cung cấp cơ sở lý thuyết và phương pháp mới để phân tích và xử lý dữ liệu trong cơ sở dữ liệu.

Luận án tập trung vào các mục tiêu cụ thể sau đây:

(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 và phụ thuộc yếu thành phụ thuộc yếu xấp xỉ tổng quát. Sự mở rộng này sẽ làm tiền đề để tìm ra các mối tương quan giữa các loại phụ thuộc này.

(2) Xây dựng các lớp phụ thuộc dữ liệu trong cơ sở dữ liệu dựa trên hai đặc trưng là phép sánh trị và công thức suy dẫn. Phát hiện các dạng khác nhau của phụ thuộc Boole dương tổng quát. Chỉ ra các đặc tả chung cho phụ thuộc nới lỏng tổng quát và chứng minh rằng phụ thuộc nới lỏng và phụ thuộc hàm nới lỏng là các trường hợp riêng của phụ thuộc Boole dương tổng quát.

(3) Phát triển lớp phụ thuộc mới đó là phụ thuộc logic rộng hơn các phụ thuộc đã biết. Với lớp phụ thuộc logic này, luận án bước đầu đã gặt hái được các kết quả như sau:

- 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 hằng đúng (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

- Đề xuất thuật toán tìm bao đóng cho lớp phụ thuộc logic. Bao đóng được vận dụng để giải các bài toán thành viên và trong các bài toán tìm khóa.
- Xây dựng thuật toán tìm khóa cho lớp 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. Tìm khoá khóa được vận dụng chủ yếu trong các thuật toán truy vấn cơ sở dữ liệu.

Từ kết quả thu được, người khai thác và người thiết kế cơ sở dữ liệu có thể ứng dụng các tính chất, định lý, mệnh đề, hệ quả, thuật toán,…hoặc lựa chọn một loại phụ thuộc logic phù hợp nhất cho từng trường hợp ràng buộc dữ liệu cụ thể (nới lỏng hay làm chặt các điều kiện ràng buộc). Xây dựng lớp các phụ thuộc logic là tạo ra nền tảng trong việc mở rộng các loại phụ thuộc bậc cao về lâu dài và trước mắt, tạo thuận lợi cho các nhà thiết kế, vận hành cơ sở dữ liệu.

## 3. Đối tượng và phạm vi nghiên cứu

- *Đối tượng nghiên cứu:*

Nghiên cứu các khái niệm, các đặc tính của các phụ thuộc logic như: Phụ thuộc Boole dương, phụ thuộc Boole dương xấp xỉ, phụ thuộc Boole dương tổng quát, phụ thuộc yếu, phụ thuộc Boole dương xấp xỉ tổng quát… Mối tương quan giữa các phụ thuộc logic. Các bài toán kinh điển trong lý thuyết các phụ thuộc. Bài toán thành viên, bài toán tìm phủ không dư, phương pháp hợp giải, phủ tối thiểu, bài toán tìm bao đóng, tìm khóa trong lớp các phụ thuộc logic.

- *Phạm vi nghiên cứu:*

Nghiên cứu 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 các phụ thuộc logic.

## 4. Phương pháp nghiên cứu

Phát biểu, chứng minh các mệnh, định lý đề về các lớp phụ thuộc. Trên cơ sở 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. Tổng quát hóa các loại phụ thuộc logic trong CSDL, được xây dựng trên các mối quan hệ giữa chúng và một hệ suy dẫn với những tính chất cơ bản nhất về các loại dữ liệu. Khái quát để khai thác và mở rộng một số loại phụ thuộc; phân tích, đối sánh, lập luận logic và hình thức hóa để chứng minh tính tương đương giữa các loại phụ thuộc logic. Mô hình hóa để hình thành một định nghĩa thống nhất; đồng thời tổng hợp các đặc trưng, bài toán chung nhất giữa các loại phụ thuộc, phương pháp quy nạp để xây dựng nên lớp các phụ thuộc logic dựa trên các loại phụ thuộc đã được nghiên cứu độc lập trước đó.

## 5. Nội dung nghiên cứu

Trên cơ sở khảo sát và phân tích các công trình về phụ thuộc dữ liệu, các kết quả của các nghiên cứu liên quan đã được công bố trong nước, ngoài nước, tác giả nhận thấy cần và có thể phát triển một số đặc trưng cơ bản và tập trung nghiên cứu vào một số vấn đề sau đây:

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

(2) Nghiên cứu các phép sánh trị giữa các bộ dữ liệu 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, phục vụ như nền tảng toán học để đảm bảo cho phụ thuộc Boole dương tổng quát nhằm thu được một số dạng phụ thuộc mới là phụ thuộc yếu xấp xỉ, phụ thuộc Boole dương xấp xỉ và Boole dương xấp xỉ tổng quát.

(3) Đề xuất khái niệm của phụ thuộc Boole dương xấp xỉ, phụ thuộc Boole dương xấp xỉ tổng quát trong mô hình hệ thống dữ liệu quan hệ, phát biểu và chứng minh các tính chất, các định lý của Phụ thuộc Boole dương xấp xỉ và phụ thuộc Boole dương xấp xỉ tổng quát, khẳng mối tương quan giữa phụ thuộc Bool dương xấp xỉ và phụ thuộc Boole dương tổng quát, giữa phụ thuộc Boole dương xấp xỉ tổng quát và phụ thuộc Boole dương tổng quát. Luận án tiếp tục đề xuất điều kiện cần và đủ của một thể hiện chặt của tập phụ thuộc Boole dương xấp xỉ... Từ các kết quả của phụ thuộc Boole dương xấp xỉ.

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

(5) Phân tích các thuật toán tìm khoá, thuật toán tìm bao đóng với phụ thuộc Boole dương tổng quát.

**6. Ý nghĩa khoa học và thực tiễn**

*Ý nghĩa về mặt khoa học*

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

*Ý nghĩa thực tiễn*

Kết quả của luận án được dùng để thiết lập một mô hình có tính tổng quát về các loại lớp phụ thuộc áp dụng trong việc khai thác sử dụng các công cụ học sâu và trí tuệ nhân tạo theo khung nhìn tổng quan về mối tương quan giữa các phụ thuộc logic trong CSDL. Từ kết quả này, người quản trị dữ liệu có thể lựa chọn lý thuyết của các phụ thuộc phù hợp hơn cho việc thiết kế và triển khai cơ sở dữ liệu cụ thể đáp ứng được các nhu cầu biến động từng ngày của thực tiễn, có khả năng hỗ trợ cho các ứng dụng đa phương tiện, đáp ứng được một số nhu cầu trong việc thu thập, tổ chức, quản trị CSDL của mỗi cá nhân và tổ chức hay trong các xí nghiệp.

## 7. Bố cục của luận án

Dựa vào một số nội dung liên quan về nghiên cứu đã trình bày ở trên, luận án bao gồm bốn phần: phần mở đầu, ba chương nội dung, phần kết luận và tài liệu tham khảo.

Phần mở đầu NCS trình bày một số vấn đề về tính cấp thiết của đề tài, các nghiên cứu liên quan đến các phụ thuộc trong cơ sở dữ liệu, chỉ ra rằng lĩnh vực nghiên cứu về phụ thuộc xấp xỉ còn nhiều bài toán mở và có ý nghĩa về mặt lý thuyết và thực tiễn, từ đó đưa ra vấn đề cần đi sâu về loại phụ thuộc này.

Chương 1 trình bày cơ sở lý thuyết về các phụ thuộc dữ liệu trong mô hình quan hệ.

Chương 2 đề xuất và mở rộng khái niệm về phụ thuộc xấp xỉ và các biến thể của phụ thuộc xấp xỉ, các kết quả quan trọng về cơ sở toán học của phụ thuộc xấp xỉ.

Chương 3 đề xuất ba thuật toán liên quan đến bài toán suy dẫn, bao đóng và khoá của lược đồ quan hệ theo phụ thuộc Boole dương tổng quát.

Kết luận Trình bày các kết quả của luận án, đưa ra những vấn đề còn tồn tại và chỉ ra một số hướng nghiên cứu tiếp theo của luận án.

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

## 1.1. Mở đầu

Trong chương 1, luận án trình bày các khái niệm cơ bản liên quan lý thuyết cơ sở dữ liệu quan hệ. Phần đầu tiên của chương là các khái niệm về cơ sở dữ liệu, phụ thuộc cơ sở dữ liệu, bộ, quan hệ , thuộc tính…và các khái niệm cơ sở về phụ thuộc hàm như là một công cụ giúp cho việc biêu diễn ngữ nghĩa của dữ liệu, đặc biệt đảm bảo sự nhất quán và độ tin cậy trong dữ liệu. Do nhu cầu phát triển, để quản lý các dữ liệu lớn và phức tạp (theo nghĩa có nhiều ràng buộc giữa các loại thuộc tính và các bộ), người ta đã phát triển nhiều lớp các phụ thuộc khác nhau. Ví dụ như phụ thuộc hàm có điều kiện, phụ thuộc độ đo, phụ thuộc đối sánh, phụ thuộc sai khác, phụ thuộc hàm mềm, phụ thuộc hàm xấp xỉ, phụ thuộc Boole dương, phụ thuộc sai khác, phụ thuộc sai khác tổng quát,…

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 Boole dương tổng quát, phụ thuộc hàm, phụ thuộc hàm xấp xỉ, phụ thuộc hàm nới lỏng. Các nội dung này được trình bày cụ thể trong các mục 1.2, 1.3 và 1.4 của chương 1.

(2) Sử dụng hai đặc trưng phép sánh trị và công thức suy dẫn để xây dựng các lớp phụ thuộc dữ liệu khác nhau trong CSDL. Nội dung chi tiết của mục này được trình bày cụ thể trong các mục 1.4, 1.5 và 1.6 của chương 1.

(3) Khảo sát, đặc tả và phân tích các biến thể của phụ thuộc Boole dương tổng quát nhằm phát hiện ra một loại đặc tả mới cho phụ thuộc nới lỏng tổng quát, từ đó, chứng minh rằng phụ thuộc nới lỏng, phụ thuộc hàm nới lỏng cũng là trường hợp riêng của phụ thuộc Boole dương tổng quát. Dựa vào các loại phụ thuộc hàm nới lỏng chúng ta có thể tổng quát hóa các loại phụ thuộc này thành phụ thuộc Boole dương tổng quát và ngược lại, từ cơ sở lý thuyết này có thể hiện thực hóa về các dạng của phụ thuộc hàm nới lỏng phụ thuộc vào các

yêu cầu của ứng dụng trong thực tế. Kết quả của phần này được trình bày cụ thể trong các mục 1.6 của chương 1.

Các kết quả của chương 1 được công bố trong CT5 – *Danh mục các công trình đã công bố* của NCS.

**1.2. Một số khái niệm và quy ước**

Phần này sẽ trình bày các khái niệm và một số quy ước được sử dụng trong luận án.

Trong luận án, khi nói đến phụ thuộc dữ liệu chính là đề cập đến sự ràng buộc giữa các thuộc tính và các bộ trong cơ sở dữ liệu quan hệ.

*Định nghĩa thuộc tính, miền trị, bộ và quan hệ :*

Cho tập hữu hạn $U = \{a_1, a_2, \ldots, a_n\}$ với $n \geq 1$, $U$ được gọi là *tập các thuộc tính*. Với mỗi phần tử $a \in U$ là tập *miền trị* $V_a$. Ký hiệu $\mathcal{D}$ là *hợp các miền trị* $V_a$ của các thuộc tính trong $U$, $\mathcal{D} = \bigcup_{a \in U} V_a$.

Quy định mỗi miền trị $V_a$ có ít nhất hai giá trị.

*Quan hệ r* với tập thuộc tính $U$, ký hiệu $r$ và $t$ là tập các ánh xạ $U \rightarrow \mathcal{D}$ sao cho mỗi thuộc tính $a \in U$ ta có $t.a \in V_a$, 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 $r$ [34]

Trong luận án sử dụng một số ký hiệu và quy ước sau:

Phép gán trị $e$ cho biến $x$

$$x := e$$

Kí hiệu $x := (e)\ ?\ a : b$, cho biết nếu điều kiện $e$ thỏa thì $x$ nhận giá trị $a$, ngược lại $x$ nhận giá trị $b$.

Các thuộc tính được đặt tên bằng các chữ cái có thể kèm chỉ số, ví dụ, *A, B, a, b,* $a_i$, $b_j$,… Kí hiệu $V_a$ là miền trị của mỗi thuộc tính $a$. Tập các thuộc tính được biểu diễn qua các chữ cái viết hoa nằm ở cuối của bảng chữ, ví dụ, *X, Y, Z,…*

Các bộ trong quan hệ $r$ được ký hiệu là $t, u, v,$...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 $t.X$, cụ thể là $t.X = \{t.a \mid a \in X\}$.

Các phần tử trong một tập thường không có các ký hiệu biểu diễn tập và nó được liệt kê như một chuỗi ký tự, Ví dụ; viết $X = cde$ để thay cho cách viết $X = \{c, d, e\}$.

Hợp của $X$ và $Y$ được kí hiệu là $X \cup Y$; giao của $X$ và $Y$ được kí hiệu là $X \cap Y$; $X$-$Y$ là phép trừ của hai tập thuộc tính $X$ và $Y$.

Gọi X và Y là các công thức login. Khi đó, kí hiệu $XY$ hoặc $X{\cdot}Y$ hoặc $X \wedge Y$ là phép hội và được gọi là *tích*; ký hiệu $X \vee Y$ là phép tuyển hoặc được viết $X+Y$ gọi là *tổng* logic, trong khi đó phép phủ định có được kí hiệu là dấu ' hoặc $\neg$. Ví dụ, ta có biểu thức $de + f + eg$' là một biểu thức giống với công thức logic sau đây: $d \wedge e \vee f \vee e \wedge (\neg g)$.

Mỗi biến hoặc phủ định của biến xuất hiện trong một công thức logic được gọi là một *kí biến*, trong đó biến kèm dấu phủ định, $a$', $b$'… được gọi là kí biến âm; biến không kèm dấu phủ định, $a$, $b$,… được gọi là kí biến dương.

Cho $U$ là tập $n$ biến logic. Với mỗi tập con các biến $V = \{v_1, \ldots, v_k\}$ trong $U$ ta qui ước sử dụng các kí hiệu sau:

$V = v_1 \ldots v_k$ là hội (tích) các biến, $V^* = v_1 + \ldots + v_k$ là tuyển (tổng) các biến, $V^*$ được gọi là *tuyển sơ cấp* khi trong $V^*$ chứa đủ $n$ kí biến khác nhau. Nếu $k = 0$, tức $V$ là tập rỗng, ta qui ước $V = 1$ (true) và $V^* = 0$ (false). Trường hợp $a$ là một biến thì $a^* = a$.

Phân hoạch của tập $M$ (thành các tập con rời nhau) $X_1, X_2, \ldots, X_k$ được ký hiệu $M = X_1 \mid X_2 \mid \ldots \mid X_k$ với ý nghĩa $M = X_1 \cup X_2 \cup \ldots \cup X_k$ và $X_i \cap X_j = \varnothing$, $1 \leq i, j \leq k$, $i \neq j$.

### 1.2. Phụ thuộc hàm

Phụ thuộc hàm là một loại phụ thuộc logic kinh điển được sinh ra sớm nhất và thường được dùng như một công cụ để thiết kế dữ liệu và để chuẩn hóa cơ sở dữ liệu . Các khái niệm, đặc điểm, tính chất, các chuẩn và thuật toán trong cơ sở dữ liệu được xây dựng một cách rất chặt chẽ thông qua phụ thuộc hàm, thuật toán tìm bao đóng, khóa, bài toán thành viên, bài toán tìm phủ không dư, bài toán tìm phủ tối thiểu [37]

Trong phần này, luận án trình bày một số khái niệm về phụ thuộc hàm [34]:

Gọi $U = \{a_1, a_2, \ldots, a_n\}$, $n \geq 1$ là tập thuộc tính. Giả sử ta có $X$, $Y \subseteq U$ ta định nghĩa một phụ thuộc hàm trên $U$ là biểu thức có dạng như sau: $X \rightarrow Y$.

Cho quan hệ *r*, với *U* là tập thuộc tính và phụ thuộc hàm *f* có dạng $X \rightarrow Y$ trên tập *U*. Khi đó ta phát biểu rằng quan hệ *r thỏa phụ thuộc hàm f* và được ký hiệu là *r*(*f*), Điều này có nghĩa là nếu hai bộ bất kỳ trong quan hệ *r* giống nhau trên tập thuộc tính của *X* thì hai bộ đó cũng phải giống nhau trên tập thuộc tính của *Y*, cụ thể là:

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

Nếu công thức $f$: $X \rightarrow Y$ là một phụ thuộc hàm trên *U* thì có nghĩa là tập thuộc tính *X* ở vế trái xác định được tập thuộc tính *Y* ở vế phải của phụ thuộc hàm *f*.

Cho quan hệ *r,* tập phụ thuộc hàm *F* trên tập *U*. Ta phát biểu rằng *r thỏa tập phụ thuộc hàm F* và được ký hiệu là *r*(*F*), nếu *r* thỏa mọi phụ thuộc hàm trong *F*, cụ thể là:

$$R(F) \overset{def}{\Leftrightarrow} \forall f \in F: R(f)$$

*Ví dụ 1.1*

Cho *r*(*A, B, C, D*) sau đây:

| **r(*c*** | ***d*** | ***e*** | ***f*)** |
|---|---|---|---|
| **x** | 1 | x | 2 |
| x | 1 | y | 2 |
| y | 2 | x | 1 |
| y | 2 | y | 1 |

và các phụ thuộc hàm

$f_1$: $A \rightarrow A$, $f_2$: $A \rightarrow B$, $f_3$: $AC \rightarrow C$, $f_4$: $A \rightarrow D$, $f_5$: $D \rightarrow A$, $f_6$: $A \rightarrow C$.

Ta thấy các phụ thuộc hàm $f_1$ - $f_5$ *"đúng"* trong quan hệ *r và* quan hệ *r* *không thỏa* phụ thuộc hàm $f_6$.

***Bảng trị***

Mỗi cặp bộ *u* và *v* trong quan hệ *r*, ta xây dựng vector 0/1: $\boldsymbol{t}(\boldsymbol{u}, \boldsymbol{v}) = (\boldsymbol{t_1}, \dots, \boldsymbol{t_n})$ như sau:

$t_i = 1$ khi và chỉ khi $u.x_i = v.x_i$, $1 \leq i \leq n$

Với mỗi quan hệ *r* ta xây dựng *bảng trị* $T_r$ của quan hệ *r* như sau:

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

*Ví dụ 1.2*

Giả sử ta có quan hệ *r* có các bộ như sau:

Bảng 1.1. Bảng trị của quan hệ *r*

*r*

| bộ | A | B | C | D |
|---|---|---|---|---|
| 1 | a | 1 | x | 2 |
| 2 | a | 1 | y | 2 |
| 3 | b | 2 | x | 1 |
| 4 | b | 2 | y | 1 |

$T_r$

| cặp bộ | A | B | C | D |
|---|---|---|---|---|
| (1,1) | 1 | 1 | 1 | 1 |
| (1,2) | 1 | 1 | 0 | 1 |
| (1,3) | 0 | 0 | 1 | 0 |
| (1,4) | 0 | 0 | 0 | 0 |

| Chú thích |
|---|
| (2,3) = (1,4) |
| (3,4) = (1,2) |
| (2,4) = (1,3) |

Ta xét phụ thuộc hàm *f*: $X \rightarrow Y$ như một công thức suy dẫn trong đại số Boole, trong đó vế trái *X* và vế phải *Y* là những hội (tích) logic của các biến Boole trong *U*. Ký hiệu $T_{X \rightarrow Y}$ là bảng chân lý của công thức $X \rightarrow Y$, tức là những giá trị Boole của X và Y cho công thức nhận giá trị 1. Nếu xét các phép sánh trị (=) trên các miền trị của các thuộc tính trong *U* thì ta thấy quan hệ *r* thỏa *phụ thuộc hàm* $X \rightarrow Y$ khi và chỉ khi $T_{r[XY]} \subseteq T_{X \rightarrow Y}$ trong đó $T_{r[XY]}$ là những dòng của bảng $T_r$ ứng với các cột $X \cup Y$.

$f: T_{A \rightarrow B}$

| A | B | f |
|---|---|---|
| 0 | 0 | 1 |
| 0 | 1 | 1 |
| 1 | 1 | 1 |

$f: T_{AC \rightarrow C}$

| A | C | AC | f |
|---|---|---|---|
| 0 | 0 | 0 | 1 |
| 0 | 1 | 0 | 1 |
| 1 | 0 | 0 | 1 |
| 1 | 1 | 1 | 1 |

$f: T_{A \rightarrow C}$

| A | C | f |
|---|---|---|
| 0 | 0 | 1 |
| 0 | 1 | 1 |
| 1 | 1 | 1 |

**Lược đồ quan hệ**

Một cặp gồm ($U$, $F$) gọi là *lược đồ quan hệ (LĐQH)* trong đó $U$ là tập hữu hạn các thuộc tính, $F$ là tập các ràng buộc. Khi chưa đề cập đến các ràng buộc thì tập thuộc tính $U$ được gọi là LĐQH.

**Bao đóng của tập thuộc tính**

Cho tập $U$, tập phụ thuộc hàm $F$ trên $U$ và một tập con các thuộc tính $X$ trong $U$, ký hiệu $F^+$ là tập nhỏ nhất các phụ thuộc hàm trên $U$ chứa $F$ . Ký hiệu $X^+$ là bao đóng của tập thuộc tính $X$:

$$X^+ = \{A \in U \mid X \to A \in F^+\}$$

*Một số tính chất của bao đóng*

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

(1) *Tính phản xạ*: $X \subseteq X^+$

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

(3) *Tính lũy đẳng*: $(X^+)^+ = X^+$

Ngoài ra, sử dụng ba tính chất (1), (2), (3) ở trên ta có thể chứng minh được các tính chất (4)-(8) sau:

(4) $(XY)^+ \supseteq X^+Y^+$

(5) $(X^+Y)^+ = (XY^+)^+ = (XY)^+$

(6) ($X \to Y$ *khi và chỉ khi* $Y^+ \subseteq X^+$

(7) $X \to X^+$ *và* $X^+ \to X$

(8) $X^+ = Y^+$ *khi và chỉ khi* $X \to Y$ và $Y \to X$

**Khóa của lược đồ quan hệ**

Cho lược đồ quan hệ $q$ = ($U$, $F$). $T$ập con T $\subseteq U$ được gọi là *khoá* của $a$ nếu:

(i) $K^+ = U$

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

Hai điều kiện trên tương đương với

(i') $K \to U$

(ii') $\forall A \in K$: $K - \{A\} \nrightarrow U$

Nếu khoá $T$ thoả các điều kiện (i) (hoặc (i')) khi đó $T$ sẽ được gọi là *siêu* khoá của LĐQH.

*Bài toán*: Cho LĐQH ($U$, $F$) và một phụ thuộc hàm $f$ trên $U$, xác định rằng phụ thuộc hàm $f$ có được suy dẫn từ tập phụ thuộc hàm $F$ hay không?

Có 3 tiếp cận để thực hiện bài toán suy dẫn trên như sau:

*Tiếp cận 1: Suy dẫn theo quan hệ*

Cho $F$ là tập phụ thuộc hàm và một phụ thuộc hàm $f$ trên tập thuộc tính $U$. Ký hiệu $SAT(F)$ là tập toàn bộ các quan hệ trên tập $U$ thoả tập phụ thuộc hàm $F$. Ta nói tập phụ thuộc hàm $F$ *suy dẫn theo quan hệ* phụ thuộc hàm $f$ và kí hiệu là $F \vdash f$, $r$ thoả phụ thuộc hàm $f$ nếu với mọi quan hệ trong $r$ thoả phụ thuộc hàm $F$

$$F \vdash f \Leftrightarrow SAT(F) \subseteq SAT(f)$$

Với tập thuộc tính $U$ và tập phụ thuộc hàm $F$ trên $U$, ta kí hiệu $F^*$ là tập các phụ thuộc hàm $f$ trên $U$ *được suy dẫn theo quan hệ* từ tập phụ thuộc hàm $F$.

$$F^* = \{ f: X \to Y \mid X, Y \subseteq U,\ F \vdash f \}$$

Tính đủ và tính chính xác trong hệ tiên đề Armstrong [3]

$$F^+ = F^*$$

Còn cách gọi khác là suy dẫn suy dẫn theo logic và suy dẫn theo quan hệ được gọi là một, điều này có nghĩa là:

$$F \models f \Leftrightarrow F \vdash f$$

*Tiếp cận 2: Suy dẫn theo quan hệ có không quá p bộ*

Cho tập thuộc tính $U$ và tập phụ thuộc hàm $F$ trên $U$, một phụ thuộc hàm $f$ trên $U$, ký hiệu $SAT\_p$(U), $p \geq 1$ là tập toàn bộ các quan hệ có số bộ trên tập $U$ không lớn hơn $p$ bộ và thoả tập phụ thuộc hàm $F$, tập các quan hệ trên tập $U$ kí hiệu là $REL\_p(U)$. Khi đó phụ thuộc hàm $f$ *được suy dẫn theo quan hệ có* *không quá p bộ* từ tập phụ thuộc hàm $F$, nó được kí hiệu là $F \vdash_p f$, $r$ cũng thoả $f$ nếu với mọi quan hệ trong $r$ thuộc $REL\_p(U)$ thoả $F$ [38].

$$F \vdash_p f \Leftrightarrow SAT\_p(F) \subseteq SAT\_p(f)$$

Cho tập phụ thuộc hàm $F$ *trên* $U$ ( $U$ là tập thuộc tính), đặt $F'$ là tập các phụ thuộc hàm $f$ trên $U$ được suy dẫn theo quan hệ có không quá hai bộ từ tập phụ thuộc hàm $F$

$$F' = \{ f: X \to Y \mid X, Y \subseteq U,\ F \vdash_2 f \}$$

Định lý tương đương [38]

$$F^+ = F^* = F'$$

*Tiếp cận 3:Suy dẫn theo tiên đề (hay suy dẫn logic)*

Phụ thuộc hàm $f$ được suy dẫn từ tập phụ thuộc hàm $F$ và được ký hiệu là $F \models f$, nếu và chỉ nếu $F \to f$ là một công thức logic hằng đúng. Cũng có thể phát biểu rằng, phụ thuộc hàm $f$ được suy dẫn logic từ tập phụ thuộc hàm $F$ nếu xuất phát từ $F$, thì ta áp dụng các tính chất của logic mệnh đề sau hữu hạn lần ta sẽ thu được phụ thuộc hàm $f$.

$$F \models f \Leftrightarrow f \in F^+$$

Có thể nói rằng phụ thuộc hàm $f$ được suy dẫn theo logic từ tập phụ thuộc hàm $F$ nếu bắt đầu từ $F$, chúng ta áp dụng các luật suy dẫn của tiên đề Armstrong $A^o$ sau hữu hạn lần ta sẽ có được phụ thuộc hàm $f$.

**Định lý 1.1:** *Định lý tương đương cho phụ thuộc hàm* [26]

Cho tập phụ thuộc hàm $F$ và một phụ thuộc hàm $f$ 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 lược đồ quan hệ ($U$, $F$), trong đó $F$ là tập phụ thuộc hàm trên $U$ (trong đó $U$ là tập thuộc tính).

Ký hiệu $F^+$ là *Bao đóng của tập* phụ thuộc hàm , bao gồm tập tất cả các phụ thuộc hàm trên *U* được suy dẫn logic từ tập phụ thuộc hàm *F* [36]

$$F^+ = \{f \mid F \models f\}$$

Ký hiệu là $F^*$ là tập các phụ thuộc hàm *f* trên *U* được suy dẫn theo quan hệ từ tập phụ thuộc hàm *F*.

Với lý do trên, trong chương này và các chương tiếp theo, khi nói về bao đóng của tập phụ thuộc hàm, luôn luôn dùng một ký hiệu $F^+$ thay cho ba ký hiệu nói trên. Ngoài ra, dùng thuật ngữ "*suy dẫn*" thay cho ba loại suy dẫn tương đương nói trên.

Tuy nhiên, trong thực tiễn việc khai thác cơ sở dữ liệu cho thấy rằng phụ thuộc hàm là quá chặt chẽ, bởi trên thực tế, không phải dữ liệu nào cũng có thể sử dụng giới hạn của việc so sánh hai bộ theo đẳng thức (=) (điều này có nghĩa là từng cặp giá trị tương ứng của hai bộ trong quan hệ được so sánhss để đưa ra kết luận chúng có giống nhau hay không) và đơn thuần chỉ dùng phép duy dẫn vế trái sang vế phải như phụ thuộc hàm.

*Ví dụ 1.2*, (tình huống thứ nhất) Trong một quan hệ có số lượng lớn các bộ. Tuy nhiên, trong số đó tồn tại vài bộ vi phạm ràng buộc do một số dữ liệu bị lỗi hoặc có một vài sai sót ngoại lệ.

Tình huống thứ hai khiến cho việc vận dụng phụ thuộc hàm bị hạn chế là phép sánh trị đẳng thức. Giả sử chúng ta qui định về số đo để may đồng phục học sinh với chiều cao (H) và cân nặng (W) quyết định cỡ quần áo (*S*), thì phụ thuộc $H \wedge W \rightarrow S$ không thể biểu đạt chính xác ngữ nghĩa : "nếu hai học sinh chênh lệch nhau về chiều cao dưới 3cm và cân nặng dưới 1kg thì hai học sinh đó mặc cùng cỡ đồng phục. Do đó, nhu cầu mở rộng phụ thuộc hàm là tất yếu và tự nhiên.

Một trong các giải pháp để thực hiện việc này là mở rộng khái niệm so sánh các giá trị của các thuộc tính của các bộ, tìm tập các phụ thuộc dữ liệu thu gọn tương đương với tập phụ thuộc dữ liệu ban đầu... Đây cũng chính là mục đích và ý nghĩa của việc mở rộng khái niệm so sánh trong lớp các phụ thuộc dữ liệu như phụ thuộc mạnh, phụ thuộc yếu, phụ thuộc sai khác, phụ thuộc đối ngẫu, phụ hàm xấp xỉ hay các phụ thuộc Boole dương tổng quát

### 1.3. Phụ thuộc hàm nới lỏng

Sau khi Codd công bố năm 1970 về phụ thuộc hàm trong cơ sở dữ liệu, các nhóm nghiên cứu đã liên tục phát triển và đề xuất nhiều loại phụ thuộc dữ liệu, nhằm cải thiện và mở rộng khả năng mô tả dữ liệu trong thế giới thực. Một trong những hướng nghiên cứu được chú trọng nghiên cứu là các phụ thuộc hàm *nới lỏng* X→Y, trong đó X và Y là các tập thuộc tính, ký hiệu → biểu thị mối quan hệ phụ thuộc giữa tập thuộc tính Y vào tập thuộc tính X với các điều kiện *nới lỏng* khác nhau [5]

Năm 2016 nhóm nghiên cứu của Giuseppe Polese tập hợp các công trình về PTH và biến thể của PTH trong một báo cáo tổng quan về các loại phụ thuộc này và đặt tên là *phụ thuộc hàm nới lỏng* (PTHNL).

Phụ thuộc hàm nới lỏng được thể hiện dưới dạng chung [24]:

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

Với $\lambda$ và $\gamma$ là các điều kiện nới lỏng như sau:

Ta nói quan hệ $r$ thoả phụ thuộc hàm nới lỏng $X(\lambda) \rightarrow Y(\gamma)$ nếu với mọi cặp $u$, $v$ là bộ huộc $r$ và $u.X$ thoả điều kiện nới lỏng $\lambda$ thì $v.Y$ thoả điều kiện nới lỏng $\gamma$

Ta xét một vài điều kiện nới lỏng sau đây:

*Nới lỏng phép so sánh:* Quan hệ $r$ được gọi là thỏa phụ thuộc hàm nới lỏng nếu khi và chỉ khi hai bộ $u$, $v$ bất kỳ trong quan hệ $r$ chênh lệch nhau không vượt ngưỡng $\lambda$ trên tập thuộc tính $X$ thì chúng cũng chênh lệch nhau không vượt ngưỡng $\gamma$ trên tập thuộc tính $Y$. Như vậy các tân từ $\lambda$ và $\gamma$ là các phép sánh trị.

*Nới lỏng theo lực lượng:*

Quan hệ $r$ được gọi là thỏa phụ thuộc hàm nới lỏng theo lực lượng $X(\delta)$

$\rightarrow Y$ khi các bộ của quan hệ $r$ nhỏ hơn ngưỡng $\delta$ đồng thời số lượng các bộ còn lại thì thỏa phụ thuộc hàm truyền thống $X \rightarrow Y$. Điều này có nghĩa là ta có thể loại khỏi $r$ một vài bộ với với ngưỡng $\delta$ để quan hệ còn lại thỏa PTH $X \rightarrow Y$.

*Phụ thuộc nới lỏng tổng quát (PTNLTQ)* [24]

Cho tập thuộc tính $X$ và $Y$ là tập thuộc tính trong $U$. Phụ thuộc nới lỏng tổng quát được biểu diễn như sau:

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

Như vậy, phụ thuộc nới lỏng tổng quát $f$ thỏa trong phụ thuộc hàm trong $r$ nếu:

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

Với mọi cặp bộ $(u, v)$ với $u, v \in r$ thoả điều kiện $\gamma$ trên $X$ thì cặp bộ này cũng thoả điều kiện $\vartheta$ trên $Y$.

Việc mở rộng phụ thuộc hàm nới lỏng tổng quát nhằm mục đích làm các điều kiện kèm theo của phụ thuộc hàm truyền thống được giảm nhẹ. Sử dụng các tân từ $\vartheta$ và $\gamma$ để phát biểu các điều kiện giảm nhẹ trong phụ thuộc nới lỏng tổng quát.

*Ví dụ 1.3*

Bảng 1.2. Quan hệ tính tiền Grab

| **Mã chuyến (MC)** | **Điểm đầu** | **Điểm cuối** | **Khoảng cách (KC)** | **Thành tiền (TT)** |
|---|---|---|---|---|
| C01 | 54 Lê Trọng Tấn | 19 Nguyễn Chí Thanh | 11 km | 100.000 |
| C02 | 52 Lê Trọng Tấn | 19 Nguyễn Chí Thanh | 10.9 km | 100.000 |
| C03 | 60 Lê Văn Lương | 24 Thái Thịnh | 3.6 km | 36.000 |
| C04 | 62 Lê Văn Lương | 22 Thái Thịnh | 3.65 km | 36.000 |

Quan hệ trên có các phụ thuộc hàm chặt sau đây:

$$MC \rightarrow TT$$

$$KC \rightarrow TT$$

Ta thấy, hai khoảng cách có độ chênh lệnh $\|u.KC - v.KC\| \leq 0.1$ thì thành tiền như nhau.

Chúng ta có phụ thuộc hàm nới lỏng như sau :

$$KC \rightarrow TT$$

Như vậy, phụ thuộc hàm nới lỏng làm cho việc quản lý dữ liệu mịn hơn, nới lỏng thay phép so sánh đẳng thức bằng một phép so sánh khác.

## 1.4. Phụ thuộc Boole dương

Nếu ta mở rộng các ràng buộc dữ liệu sang dạng các công thức Boole dương và vẫn giữ nguyên các phép so sánh đẳng thức trên các bộ của quan hệ thì ta sẽ nhận được phụ thuộc Boole dương [12].

### *1.4.1. Công thức Boole*

***Định nghĩa 1.1***

Cho $U = \{a_1,...,a_n\}$ là tập các *biến Boole*, kí hiệu $\mathcal{B}$ là tập giá trị *trị Boole* chứa 2 giá trị 0 và 1 và được viết là $\mathcal{B} = \{0,1\}$. Lúc này các *công thức Boole* (CTB) còn được gọi là các *công thức logic* (CTLG)

***Định nghĩa 1.2***

Mỗi vector hằng số 0/1 và $v = (v_1,...,v_n)$ trong $\mathcal{B}^n = \mathcal{B}\times\mathcal{B}\times...\times\mathcal{B}$ được gọi là một phép gán trị.

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

1) Thay toàn bộ trị của mỗi biến $x_i$ trong $f$ bằng trị $v_i$ tương ứng, $i$=1,2,…,$n$ để nhận 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* của các biến nằm trong tập $X$ là dãy ký hiệu của $X$. Như vậy, ký hiệu $X$ cũng là biểu diễn cho các đối tượng dưới 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.*

Ngoài ra, nếu $X = \{B_1,...,B_m\} \subseteq U$ ta ký hiệu:

- $\wedge X = B_1 \wedge B_2 \wedge ... \wedge B_m$, và gọi là *dạng hội*.
- $\vee X = B_1 \vee B_2 \vee ... \vee B_m$ và gọi là *dạng tuyển*.

Lưu ý rằng, theo quy định trên ta có, $\wedge X$ và $X$ biểu diễn cùng một hội suy dẫn của 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

$$f\text{: } \wedge Z \rightarrow \wedge V$$

- *công thức suy dẫn mạnh* nếu $Z$ có dạng tuyển và $V$ có dạng hội, tức là

$$f\text{: } \vee Z \rightarrow \wedge V$$

- *công thức suy dẫn yếu* nếu $Z$ có dạng hội và $V$ có dạng tuyển, tức là

$$f\text{: } \wedge Z \rightarrow \vee V$$

- *công thức suy dẫn đối ngẫu* nếu $Z$ và $V$ đều có dạng tuyển, tức là

$$f\text{: } \vee Z \rightarrow \vee V$$

Ta quan tâm hai phép gán trị đặc biệt là *phép gán trị, e* = (1,1,...,1) và *phép gán trị*, $z$ = (0,0,...,0).

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

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

### *1.4.2. Bảng chân lý và bảng trị*

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

Với mỗi công thức $f$ trên tập thuộc tính $U$, ta ký hiệu *bảng trị của f* là $V_f$ trong đó chứa $n$+1 cột, trong đó $n$ cột đầu tiên bao gồm các trị của các biến$U$, cột cuối cùng và cột thứ $n$+1 thì bao gồm trị của $f$ tương ứng với mỗi phép gán trị của mỗi dòng tương ứng. Ta thấy, bảng trị chứa $n$+1 cột và $2^n$ dòng, trong đó $n$ là lực lượng của tập $U$. Ký hiệu $T_f$ *là bảng chân lý*, chính là tập các phép gán trị $v$ để cho $f(v)$ nhận giá trị *true* hay còn gọi là nhận giá trị 1,

$$T_f = \{v \in \mathcal{B}^n \mid f(v) = 1\}$$

Khi đó *giao* của các bảng chân lý của mỗi công thức thành viên trong $F$ chính là *bảng chân lý* $T_F$ của tập các công thức $F$ trên $U$ và được viết như sau

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

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

*Chú ý*

Với mỗi công thức Boole $f$, bảng trị $V_f$ của $f$ chứa trị của *mọi phép gán trị* cho $f$, trong khi bảng chân lý $T_f$ chỉ là *một phần của bảng trị* ứng với các giá trị 1 của $f$.

*Ví dụ 1.4*

*Cho công thức Boole* $F = \{f\text{: } A \rightarrow B,\ g\text{: } C \wedge A\}$ trên $U = ABC$. Ta có bảng trị $V_f$, $V_g$ và các bảng chân lý $T_f$, $T_g$ và $T_F$ như sau:

Bảng 1.3. Bảng trị $V_f$, $V_g$ và các bảng chân lý $T_f$ , Tg và $T_F$

$T_f$

| A | B | C |
|---|---|---|
| 0 | 0 | 0 |
| 0 | 0 | 1 |
| 0 | 1 | 0 |
| 0 | 1 | 1 |
| 1 | 1 | 0 |
| 1 | 1 | 1 |

$T_g$

| A | B | C |
|---|---|---|
| 1 | 0 | 1 |
| 1 | 1 | 1 |

$T_F$

| A | B | C |
|---|---|---|
| 1 | 1 | 1 |

$V_f$, $V_g$

| A | B | C | f | g |
|---|---|---|---|---|
| 0 | 0 | 0 | 1 | 0 |
| 0 | 0 | 1 | 1 | 0 |
| 0 | 1 | 0 | 1 | 0 |
| 0 | 1 | 1 | 1 | 0 |
| 1 | 0 | 0 | 0 | 0 |
| 1 | 0 | 1 | 0 | 1 |
| 1 | 1 | 0 | 1 | 0 |
| 1 | 1 | 1 | 1 | 1 |

### *1.4.3. Công thức Boole dương*

Công thức $h \in L(U)$ được gọi là công thức Bool dương nếu $h(e)$=1, trong đó $e$ gồm các giá trị 1 là phép gán của đơn vị, được viết là $e = (1,\ldots, 1)$. Ví dụ, nếu $U = bcd$ thì $b \wedge c$, $b \vee (c \wedge d)$, $b \rightarrow c$ là các công thức Bool dương, còn công thức $f(c,d) = c \rightarrow d'$ không phải là Bool dương vì $f(e) = f(1,1) = 1 \rightarrow 1' = 1 \rightarrow 0 = 0$.

Ta ký hiệu $P(U)$ là tập các công thức Boole dương trên $U$. Tác giả Post E.L [39] đã chứng minh trong nghiên cứu của mình rằng mọi công thức Boole dương đều có thể được tạo ra từ các phép toán logic gồm: $\wedge$, $\vee$, $\rightarrow$ và hằng 1.

Mỗi công thức $g$ trên $U$, $n$ là lực lượng của tập $U$, *bảng trị* của $g$, ký hiệu là $V_g$, là một bảng hai chiều gồm $n + 1$ cột và $2^n$ dòng, trong đó $n$ cột đầu tiên chứa tổ hợp các trị 0/1 của các biến trong tập $U$, trong đó cột thứ $n + 1$ gồm các giá trị của $g$ là mỗi phép gán trị của dòng đó.

Mỗi công thức $g$ trên $U$, ta kí hiệu là $T_g$ là *bảng chân lý* của $g$, là tập các phép gán trị $v$ để $g(v)$ nhận giá trị 1 (true):

$$T_g = \{v \in \mathfrak{B}^n | \; g(v) = 1\}$$

Khi đó giao của các bảng chân lý của mỗi công thức trong $S$ là bảng chân lý $T_S$ của tập hữu hạn các công thức $S$ trên $U$.

$$T_S = \bigcap_{g \in S} T_g$$

Như vậy, $v \in T_S$ khi $\forall g \in S$: $g(v) = 1$.

Ta có nhận xét sau:

Tích các công thức Boole dương là một công thức Boole dương . Thật vậy, nếu $g$ và $q$ là hai công thức Boole dương và $e$ là phép gán đơn vị thì $g(e)=q(e) = 1$, do đó $g(e) \wedge q(e)= 1 \wedge 1 = 1$.

Tổng các công thức Boole dương là một công thức Boole dương . Vì $g(e) \vee q(e) = 1 \vee 1 = 1$.

Giả sử $g$ là một công thức Boole dương và $q$ không phải là công thức Boole dương thì $(g \rightarrow q)$'. Vì $e \in T_g$ và $e \notin T_q$ nên $T_g \not\subset T_q$ , và do đó, theo định nghĩa, $(g \rightarrow q)$'.

### 1.5. Phụ thuộc Boole dương tổng quát

Phụ thuộc Boole dương đã mở rộng ràng buộc dữ liệu từ phép suy dẫn $X \rightarrow Y$ thành tập các công thức Boole dương. Điều này giúp cho người sử dụng linh hoạt hơn khi thiết kế các ràng buộc dữ liệu trên các yêu cầu thực tế của bài toán. Tuy nhiên phụ thuộc Boole dương vẫn giới hạn bởi phép sánh trị trong miền trị của thuộc tính vẫn là phép sánh đẳng thức [15].

Do đó, phụ thuộc Boole dương tổng quát (PTBDTQ) như trình bày dưới đây sẽ là một sự mở rộng của phụ thuộc Boole dương bằng cách thay phép sánh trị đẳng thức bởi phép sánh trị $\alpha$ thỏa ba tính chất phản xạ, đối xứng và bộ phận.

**Phép sánh trị tổng quát**

Cho $U = \{x_1,...,x_n\}$ là tập các *biến Boole*, kí hiệu $\mathcal{B}$ là tập giá trị *trị Boole* chứa 2 giá trị 0 và 1 và được viết là $\mathcal{B} = \{0,1\}$. Ta quy ước $d_x$ là miền trị của $x$ có ít nhất 2 phần tử. Xét ánh xạ $\alpha_x$: $d_x^2 \rightarrow \mathfrak{B}$ trên mỗi miền trị $d_x$ thoả các tiên đề sau đây: [7] [40].

$\forall a, b \in V_x$

$A1$) 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 $\exists c \in d_x$: $\alpha_x(a, c) = 0$.

Các ánh xạ $\alpha_x$ là quan hệ bộ phận thoả mãn tính chất đối xứng, phản xạ trên miền giá trị $d_x$. Khi xác định $\alpha_x$ có nghĩa là ta tạo ra một *phép so sánh* trên miền trị $d_x$ cho thuộc tính $x$. Phép so sánh đẳng thức trong phụ thuộc Boole dương là trường hợp riêng của quan hệ nói trên.

**Phép gán trị**

Giả sử các ánh xạ $\alpha_a$ đã được xác định trên mỗi miền trị $d_x$ của các thuộc tính $a$ trong tập $U = \{a_1, a_2,\ldots, a_n\}$, $r$ là một quan hệ trên $U$. Với hai bộ $u$ và $v$ tùy ý trong quan hệ $r$ ta định nghĩa $\alpha(u,v)$ là một *phép gán trị* và được tính như sau:

$$\alpha(u,v) = (\alpha_1(u.a_1, v.a_1), \alpha_2(u.a_2, v.a_2), \ldots, \alpha_n(u.a_n, v.a_n))$$

Ta thấy, với mỗi cặp bộ $u, v \in r$, $\alpha(u, v)$ là một véc tơ 0/1 trong $\mathfrak{B}^n$.

Với mỗi quan hệ $r \in REL(U)$ ta gọi *bảng trị của quan hệ r* là tập

$$T_r = \{\alpha(u, v) |\ u, v \in r\}$$

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

*Phụ thuộc Boole dương tổng quát* (PTBDTQ) là sự mở rộng của phụ thuộc Boole dương khi thay phép sánh trị đẳng thức trong phụ thuộc Boole dương bởi phép sánh trị $\alpha$ thỏa ba tính chất phản xạ, đối xứng và bộ phận [12]

Ta nói quan hệ $r$ trên tập thuộc tính $U$ thỏa phụ thuộc Boole dương tổng quát $g$ và ký hiệu $r(g)$, nếu $T_r \subseteq T_g$.

Cho $S$ là tập các phụ thuộc Boole dương tổng quát trên $U$ và $r$ là một quan hệ trên $U$. Ta nói *quan hệ R thỏa tập phụ thuộc Boole dương tổng quát S* và ký hiệu $r(S)$ nếu $r$ thỏa mọi phụ thuộc Boole dương tổng quát trong $S$:

$$r(S) \overset{def}{\Leftrightarrow} \forall g \in S: r(g) \Leftrightarrow T_r \subseteq T_S$$

Nếu $r(g)$ ta cũng nói phụ thuộc Boole dương tổng quát $g$ đúng trong quan hệ $r$.

**Các suy dẫn trong phụ thuộc Boole dương tổng quát**

*Suy dẫn logic*

Cho $g$ và $q$ là hai công thức Boole. Ta nói công thức $g$ dẫn ra được công thức $q$ và ký hiệu $g \models q$, nếu $T_g \subseteq T_q$. Ta nói $g$ và $q$ là hai công thức tương đương và ký hiệu $g \equiv q$, nếu $T_g = T_q$.

Cho hai tập các công thức Boole dương $S$ và $Q$ trong $L(U)$. Ta nói *S dẫn ra được Q* và ký hiệu $S \models Q$, nếu $T_S \subseteq T_Q$. $S$ và $Q$ là *tương đương* và ký hiệu $S \equiv Q$, nếu $T_S = T_Q$.

*Suy dẫn theo quan hệ*

Cho tập phụ thuộc Boole dương tổng quát $S$ và một phụ thuộc Boole dương tổng quát $g$. Ta nói *S dẫn ra được g theo quan hệ*, ký hiệu $S \vdash g$ nếu

$$\forall R \in REL(U): R(S) \Rightarrow R(g)$$

*Suy dẫn theo quan hệ có không quá hai bộ*

Cho tập phụ thuộc Boole dương tổng quát $S$ và một phụ thuộc Boole dương tổng quát. Ta nói *S dẫn ra được g theo quan hệ có không quá hai bộ*, ký hiệu $S \vdash_2 g$ nếu

$$\forall R \in REL_2(U): R(S) \Rightarrow R(g)$$

**Định lý 1.2** (Định lý tương đương trong phụ thuộc Boole dương tổng quát ) [41]

*Cho tập PTBDTQ F và một PTBDTQ f. Ba mệnh đề sau đây là tương đương*:

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

Định lý tương đương đã được chứng minh trong tài liệu [41]

Ngoài các loại phụ thuộc logic được liệt kê ở trên, các nhóm nghiên cứu lần lượt đề xuất những loại phụ thuộc khác nhau trong cơ sở dữ liệu như: phụ thuộc hàm đối sánh, phụ thuộc hàm có điều kiện, phụ thuộc tuần tự, phụ thuộc hàm mềm, phụ thuộc hàm độ đo ,...Các loại phụ thuộc này đều được phát triển từ phụ thuộc hàm theo hướng mở rộng phép sánh trị và mỗi một loại phụ thuộc được nghiên cứu với một mục đích cụ thể khác nhau:

*Ví dụ 1.5*

Ví dụ, cho $U = ABC$ với $d_A$ là tập các số nguyên dương, $d_B$ là tập các số thực và giá trị không xác định $\sim$, $d_C$ là tập các từ hữu hạn trên bảng chữ cái không rỗng cho trước. Khi đó các quan hệ $\alpha_A$, $\alpha_B$ và $\alpha_C$ dưới đây thỏa ba tính chất $A1$, $A2$ và A3.

$\alpha_A(a,b) = 1$ khi và chỉ khi hai số $a$ và $b$ cùng tính chẵn lẻ.

$\alpha_B(a,b) = 1$ khi và chỉ khi $a$ và $b$ đồng thời là hai số thực hoặc hai giá trị không xác định ~.

$\alpha_C(a,b) = 1$ khi và chỉ khi hai từ $a$ và $b$ có cùng chiều dài.

Cho tập phụ thuộc Boole dương tổng quát $F$ và một phụ thuộc Boole dương tổng quát $f$. 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 xác định được f theo quan hệ có không quá hai bộ*,

ký hiệu $F \vdash_2 f$ nếu $\forall r \in REL\_2(U): r(F) \Rightarrow r(f)$

## 1.6. Phân lớp các loại phụ thuộc [CT5]

Trong mục này NCS nghiên cứu để phân tích và mô tả các dạng của phụ thuộc Boole dương tổng quát nhằm phát hiện được các dạng khác nhau của phụ thuộc này. Từ đó, NCS chỉ ra các đặc tả chung cho phụ thuộc nới lỏng tổng quát và chứng minh rằng phụ thuộc hàm nới lỏng và phụ thuộc nới lỏng các trường hợp riêng của phụ thuộc Boole dương tổng quát. Đây cũng là kết quả của NCS đã được công bố trong (CT5).

### *1.6.1. Lớp IE*

Lớp IE (*Implication Fomula & Equal Comparison*) là lớp các lược đồ phụ thuộc hàm truyền thống, lớp phụ thuộc này được xây dựng dựa trên hai đặc trưng cơ bản là phép sánh trị đẳng thức vàphép toán suy dẫn.

Cho tập các thuộc tính $U = \{x_1, \ldots, x_n\}$, với $n \geq 1$. Giả sử $X, Y \subseteq U$. Một phụ thuộc nằm trong lớp IE là các biểu thức dạng $f: X \rightarrow Y$.

Ta có thể phân loại được các dạng phụ thuộc hàm sau dựa trên biểu thức logic $X$ và $Y$.

**IE-1**: Biểu thức logic có dạng $X \rightarrow Y$, được gọi là *phụ thuộc hàm truyền thống*: Cho lược đồ quan hệ $(U, F)$ và một PTH $f: X \rightarrow Y$ trên tập $U$. Ta phát biểu rằng quan hệ *r* thỏa PTH *f* và được viết *r(f)*, điều này có nghĩa là nếu hai bộ tùy ý trong *r* giống nhau trên tập thuộc tính *X* thì cũng giống nhau trên tập thuộc tính *Y*.

**IE-2:** Biểu thức logic có dạng $\vee X \rightarrow Y$ được gọi là *phụ thuộc hàm mạnh* (PTHM): Cho quan hệ *r* trên tập *U*, ta nói rằng quan hệ *r* thỏa PTHM $\vee X \rightarrow Y$ trên *U* nếu hai bộ bất kỳ trong *r* và bằng nhau tại một thuộc tính nào đó trên *X* thì hai bộ đó cũng bằng nhau trên *Y*.

**IE-3:** Biểu thức logic có dạng $X \rightarrow \vee Y$, gọi là *phụ thuộc hàm yếu* (PTHY): Cho quan hệ *r* trên tập *U*, ta nói rằng quan hệ *r* thỏa PTHY $X \rightarrow \vee Y$ trên *U* nếu hai bộ bất kỳ trong *r* và bằng nhau trên *X* thì hai bộ đó cũng bằng nhau tại một thuộc tính nào đó trên *Y*.

**IE-4:** Biểu thức logic có dạng $\vee X \rightarrow \vee Y$ , gọi là *phụ thuộc hàm đối ngẫu* (PTHĐN): Cho quan hệ *r* trên tập *U*, quan hệ *r* thỏa PTHĐN $\vee X \rightarrow \vee Y$ trên *U* nếu hai bộ bất kỳ trong *r* và 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

Bảng 1.4. Bảng đặc tả các lớp con IE1-4

| Dạng phụ thuộc | Tên gọi | Đặc điểm |
|---|---|---|
| $X \rightarrow Y$ | phụ thuộc hàm | ràng buộc chặt |
| $\vee X \rightarrow Y$ | phụ thuộc hàm mạnh | nới lỏng vế trái |
| $X \rightarrow \vee Y$ | phụ thuộc hàm yếu | nới lỏng vế phải |
| $\vee X \rightarrow \vee Y$ | 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.

### *1.6.2. Lớp LA*

Dựa theo điều kiện nới lỏng, ta chia lớp LA (*Logic Fomula & Alpha Comparison*) 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 $U = \{x_1, x_2, \ldots, x_n\}$, $n \geq 1$. 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 $\alpha$.

Ta nói quan hệ $r$ là PTH thuộc loại LA-1: $\alpha_X \rightarrow \alpha_Y$ và được viết $r(X(\alpha) \rightarrow Y(\alpha))$, nếu với hai bộ bất kỳ trong $r$, thỏa các ràng buộc thông qua $\alpha_X$ trên tập $X$, thì hai bộ đó cũng thỏa các ràng buộc được đặc tả bởi hàm sai khác $\alpha_Y$ trên tập $Y$:

$$r(X(\alpha) \rightarrow Y(\alpha)) \overset{def}{\Leftrightarrow} \forall u,v \in r\text{: } \alpha(u.X, v.X)$$

$$\Rightarrow \alpha(u.Y, v.Y)$$

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 [42], với $\alpha_X$ và $\alpha_Y$ là các hàm sai khác được định nghĩa :

Cho quan hệ $r$ trên tập thuộc tính $U$, trong đó $a \in U$ với độ sai khác $m_a$. HSK $\alpha_a$ trên thuộc tính $a$ thể hiện ràng buộc của lệch (còn gọi là độ sai khác) khác $m_a$: Với hai trị $x, y \in d_a$, ta có $\alpha_a(x, y) = 1$ chỉ khi $m_a(x, y)$ thỏa điều kiện cho $\alpha_a$ dưới dạng các biểu thức so sánh và các phép sánh trị $=$, $\neq$, $<$, $\leq$, $>$, và $\geq$.

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

Phụ thuộc hàm sai khác là trường hợp riêng của phụ thuộc hàm nới lỏng [24]. 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 1.6*, 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 bốn 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 1.7*

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:** Là lớp các phụ thuộc được xây dựng dựa trên cơ sở phép sánh trị đẳng thức và phép toán logic Bool. PTBD là một loại phụ thuộc đại diện cho lớp LA-2 [2], [21].

**Lớp LA-3:** Là lớp các phụ thuộc được xây dựng trên cơ sở phép sánh trị alpha và phép toán logic Bool. PTBDTQ là một loại phụ thuộc đại diện cho lớp LA-3, PTBDTQ cũng là lớp phụ thuộc bao gồm tất cả các PTLG trong CSDL đã và đang được quan tâm bởi các nhà nghiên cứu trong nước 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 nới 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 Bool dương tổng quát. Tìm ra một đặc tả cho phụ thuộc nới lỏng tổng quát và chỉ ra rằng phụ thuộc nới lỏng nói chung và phụ thuộc hàm nới lỏng nói riêng chỉ là những trường hợp riêng của phụ thuộc Bool dương tổng quát.

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

## 2.1. Mở đầu

Hiện nay, các nhà nghiên cứu dữ liệu lớn đặc biệt quan tâm đến việc phân lớp và đề xuất một mô hình chung cho các loại phụ thuộc dữ liệu. Để thực hiện việc này, ta cần xác định các điều kiện để nhận biết các đặc trưng của một lớp phụ thuộc, từ đó phân loại và xây dựng mối tương quan giữa các lớp phụ thuộc. Việc tìm ra một mô hình chung sẽ giúp cho việc phân loại và xử lý các dữ liệu trở nên dễ dàng và hiệu quả hơn.

Ngoài ra, trong thực tế, việc so sánh hai đối tượng không chỉ giới hạn trong khuôn khổ thuộc tính số. Ví dụ, có hai chiếc ô tô, vậy làm thế nào để biết được hai chiếc xe là xấp xỉ nhau, muốn đánh giá được chiếc xe nào đẹp hơn hoặc tốt hơn thì phải thông qua một hàm biến đổi nào đó. Hoặc trong việc kiểm soát gian lận trong giao dịch rút tiền ngân hàng: Nếu có hai giao dịch của cùng một thẻ nhưng khoảng cách giữa hai vị trí giao dịch là ≥ 60 km, thì khoảng thời gian giữa hai giao dịch đó phải trên 30 phút. Nếu điều kiện này bị vi phạm có nghĩa là một trong hai giao dịch trên có sự gian lận. Vì vậy, nhu cầu khảo sát các loại phụ thuộc logic với nhiều thuộc tính trong các CSDL như vậy là rất cần thiết trong thực tế.

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 sành 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*

Khi định nghĩa phụ thuộc hàm nới lỏng trên các tập con thuộc tính tùy ý ta đã vận dụng độ đo (khoảng cách) giữa hai bộ. Với các thuộc tính phi số ta chưa thể vận dụng được độ đo Euclid mà chỉ có thể vận dụng độ đo Hamming. Muốn vận dụng độ đo Euclid cho các thuộc tính phi số như trang phục, kỹ thuật biểu diễn, sự hài lòng, ... ta cần lượng hóa các thuộc tính phi số thông qua một ánh xạ gọi là ánh xạ $\lambda$ cho mỗi miền trị của tập thuộc tính như sau.

***Định nghĩa 2.1***

***Cho U là tập thuộc tính. Với mỗi thuộc tính A trong U ta một ánh xạ $\lambda_A$ được định nghĩa từ miền trị của các thuộc tính sang không gian số thực như sau:***

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

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

Ta gọi $\lambda_A$ là *hàm định lượng cho thuộc tính A.*

*Ví dụ 2.1*

- Cho thuộc tính *A* thuộc kiểu dữ liệu string, ta định nghĩa $\forall s \in V_A$: $\lambda_A(s)$ = len($s$). Ta có, $\lambda_A$("Internet") = 8; $\lambda_A$("e-mail") = 6; $\lambda_A$("") = 0 (định lượng của string rỗng bằng 0).

• Cho thuộc tính $B$ là bậc lương ta định nghĩa

$\forall a \in V_B$: $\lambda_B(a) = a$. Khi đó

$$\lambda_B(2.3) = 2.3;\ \lambda_B(8.0) = 8.0$$

• Cho thuộc tính $C$ là mức độ hài lòng (của khách hàng đối với hãng hàng không chẳng hạn) với các giá trị:

$V_C$ = {rất hài lòng, hài lòng, tạm được, không hài lòng}

ta định nghĩa

$\lambda_C$(rất hài lòng) = 4; $\lambda_C$(hài lòng) = 3;

$\lambda_C$(tạm được) = 2; $\lambda_C$(không hài lòng) = 1.

***Định lý 2.1***

*Biết $\lambda_A$, khi đó hàm $\alpha_A$ được định nghĩa qua hệ thức* (2.1) *dưới đây*

$$\forall a, b \in V_A: \alpha_A(a,b) \overset{def}{\Leftrightarrow} (\lambda_A(a) = \lambda_A(b)) \qquad (2.1)$$

*thỏa ba tiên đề A1, A2 và A3 trong định nghĩa ánh xạ $\alpha$.*

*Chứng minh*

Giả sử $a \in V_A$. Khi đó $\lambda_A(a) = \lambda_A(a)$ nên $\alpha_A(a,a) = 1$.

Tính chất phản xạ $A1$ đcm.

Giả sử cho $a, b \in V_A$.

Ta có $\alpha_A(b,a) = 1 \overset{def}{\Leftrightarrow} \lambda_A(a) = \lambda_A(b) \overset{def}{\Leftrightarrow} \alpha_A(b,a) = 1$. Tính chất đối xứng $A2$ được chứng minh.

Theo tính chất khác biệt của hàm Lambda, $\exists x, y \in V_A$: $\lambda_A(x) \neq \lambda_A(y)$.

Từ đây suy ra $\alpha_A(x, y) = 0$.

Tính chất bộ phận $A3$ được chứng minh.

Định lý được chứng minh.

### 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 các số thực lớn hơn hoặc bằng 0 , 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: $d(a,b) + d(b,c) \geq d(a,c)$.

*Ví dụ 2.2*

Trên đường thẳng (không gian một chiều), độ đo giữa hai điểm $x$ và $y$ biểu thị khoảng cách giữa hai điểm, được tính theo công thức, $d(x,y) = ||x-y||$. Trong không gian $n$ chiều độ đo giữa hai điểm $\boldsymbol{x} = (\boldsymbol{x_1}, \boldsymbol{x_2}, \dots, \boldsymbol{x_n})$ và $\boldsymbol{y} = (\boldsymbol{y_1}, \boldsymbol{y_2}, \dots, \boldsymbol{y_n})$ được định nghĩa qua *khoảng cách Euclid* như sau:

$$e(x, y) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \cdots + (x_n - y_n)^2} \quad (2.2)$$

Cho LĐQH $p = (U,F)$ với các hàm Lambda được định nghĩa cho mỗi thuộc tính.

Với mỗi bộ $\boldsymbol{u} = (\boldsymbol{u_1}, \boldsymbol{u_2}, \dots, \boldsymbol{u_n}) \in r$, ta định nghĩa $\lambda(u)$ là vector

$$\boldsymbol{\lambda(u) = (\lambda_1(u_1), \lambda_2(u_2), \dots, \lambda_n(u_n))}$$

***Định lý 2.2***

*Cho quan hệ r trên tập thuộc tính U và các hàm Lambda trên mỗi thuộc tính 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:*

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

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

*Chứng minh*

Ta chứng minh rằng hàm $d^*$ thỏa ba tính chất trong định nghĩa độ đo. Thật vậy, do $d$ thỏa tính không âm nên

$d^*(u,v) = d(\lambda(u), \lambda(v)) \geq 0$.

*N*ếu $u = v$ thì $\lambda(u) = \lambda(v)$, do đó $d^*(u, v) = d(\lambda(u), \lambda(v)) = 0$.

Tính chất đối xứng của $d^*$ được suy trực tiếp từ tính chất đối xứng của độ đo $d$.

Ta chứng minh tính chất tam giác. Giả sử $u$, $v$ và $w$ là ba bộ trong $r$. Do $d$ là độ đo nên $d$ thỏa tính chất tam giác, ta có,

$$d^*(u,v) + d^*(v,w) = d(\lambda(u), \lambda(v)) + d(\lambda(v), \lambda(w)) \geq d(\lambda(u), \lambda(w)) = d^*(u,w).$$

Định lý được chứng minh.

Với hai bộ trong quan hệ $r$ $\boldsymbol{u} = (\boldsymbol{u_1}, \boldsymbol{u_2}, \dots, \boldsymbol{u_n})$ và $\boldsymbol{v} = (\boldsymbol{v_1}, \boldsymbol{v_2}, \dots, \boldsymbol{v_n})$ trên cùng một tập thuộc tính $X$, ta định nghĩa khoảng cách giữa hai bộ $u$ và $v$ trong quan hệ $r$ và gọi là *khoảng cách Hamming* như sau:

$$\boldsymbol{h(u,v) = \#\{u_i \neq v_i : 1 \leq i \leq n\}}$$

trong đó $\#S$ là lượng lượng (số phần tử) trong tập $S$.

### 2.3. Đề xuất phụ thuộc hàm xấp xỉ tổng quát

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 tổng quát 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 $r$ thỏa phụ thuộc hàm xấp xỉ tổng quát $X\ \varepsilon_1\!\rightarrow \varepsilon_2\ Y$ nếu với mọi cặp bộ $u$ và $v$ trong quan hệ $r$, hai bộ $u$ và $v$ có độ đo lệch nhau không quá $\varepsilon_1$ trên tập thuộc tính $X$ thì hai bộ này cũng lệch nhau không quá $\varepsilon_2$ trên tập thuộc tính $Y$ trong đó $d$ là hàm độ đo trên các bộ của quan hệ.

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

*Ví dụ 2.3*

Sau khi tổng kết cuộc thi đồng diễn của các đội trong trường, ban giám hiệu quyết định trao giải cho các đội với tiêu chí hai đội có độ đo Euclid lệch nhau không quá 1 sẽ được xếp cùng giải.

*Bảng 2.1. Tiêu chí xếp giải cuộc thi*

| | Trang phục | Âm nhạc | Kỹ thuật | Giải |
|---|---|---|---|---|
| đội 1 | 5 | 4 | 4 | 2 |
| đội 2 | 5 | 5 | 5 | 1 |
| đội 3 | 5 | 4 | 5 | 2 |

Như vậy quan hệ trên thỏa phụ thuộc hàm xấp xỉ tổng quát

$$X \,_{1}{\rightarrow}_{0}\, Y$$

trong đó:

$X$ = {Trang phục, Âm nhạc, Kỹ thuật}

$Y$ = {Giải}

$\varepsilon_1 = 1$ và $\varepsilon_2 = 0$.

Điều đó có nghĩa: *hai đội có thành tích tính theo độ đo Euclid lệch nhau không quá 1 sẽ được xếp cùng hạng giải.*

Để ý rằng quan hệ trên cũng thỏa phụ thuộc hàm xấp xỉ tổng quát

$$X\,(1)\rightarrow(0)\;Y$$

nếu ta đánh giá theo độ đo *Hamming*.

## 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 $\varepsilon$ được gọi là *ngưỡng không tầm thường*; ngược lại, nếu với mọi cặp trị $a$, $b \in D$, ta đều có $d(a,b) \leq \varepsilon$ thì $\varepsilon$ được gọi là *ngưỡng tầm thường*.

Trong thực tế, nếu một ngưỡng là tầm thường thì nó mang giá trị lớn nhất trong mọi "*khoảng cách*" của mọi cặp phần tử, do đó sẽ không có trường hợp vượt ngưỡng và như vậy sẽ làm hạn chế ứng dụng.

*Ví dụ 2.4,* nếu hiểu độ đo $d$ giữa hai vị trí địa lí (điểm địa lí) là khoảng cách địa lí và ngưỡng trong ngày sương mù là $\varepsilon = 2$ km thì có những cặp điểm $A$, $B$ dưới ngưỡng, $d(A, B) \leq 2$ km và những cặp điểm $C$, $D$ trên ngưỡng $d(C, D) > 2$ km là việc chọn ngưỡng hợp lí. Với lí do trên ta qui định trong luận án, mọi ngưỡng đều là không tầm thường.

### 2.5. Phụ thuộc yếu

Cho tập thuộc tính $U$. Một công thức suy dẫn yếu gọi là một phụ thuộc yếu [7]. 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 $r$ 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 $U$. Ta nói quan hệ $r$ thoả tập phụ thuộc yếu $F$, và được viết $r(F)$, nếu $r$ thoả 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 $T_r \subseteq T_F$.

*Ví dụ 2.4*

Xét quan hệ $r$ mô tả các cá thể với các thuộc tính: huyết thống $H$, ADN bố $B$ và ADN mẹ $M$. Ta thấy: hai cá thể cùng huyết thống thì có ADN bố hoặc ADN mẹ giống nhau. Ta có $H \rightarrow B+M$ là một phụ thuộc yếu. Với quan hệ $r(H,$

*B*, *M*) chứa thông tin về huyết thống ta có thể xây dựng bảng trị $T_r$ như sau:

Bảng 2.1. Quan hệ *r* với các thuộc tính:Huyết thống H, ADN mẹ: M, ADN bố: B

| **Bộ** | ***H*** | ***B*** | ***M*** |
|---|---|---|---|
| 1 | h1 | b0 | m2 |
| 2 | h2 | b1 | m1 |
| 3 | h1 | b0 | m3 |
| 4 | h3 | b2 | m4 |
| 5 | h2 | b3 | m1 |

Bảng 2.2. Bảng trị của quan hệ *r*

| **Cặp bộ** | ***H*** | ***B*** | ***M*** |
|---|---|---|---|
| 1 & 1 | 1 | 1 | 1 |
| 1 & 2 | 0 | 0 | 0 |
| 1 & 3 | 1 | 1 | 0 |
| 2 & 5 | 1 | 0 | 1 |

Bảng 2.3. Bảng trị của hàm H → B+M

| **H** | **B** | **M** |
|---|---|---|
| 1 | 1 | 1 |
| 1 | 0 | 1 |
| 1 | 1 | 0 |
| 0 | 0 | 0 |
| 0 | 0 | 1 |
| 0 | 1 | 0 |
| 0 | 1 | 1 |

Các bảng trên cho ta $T_r \subseteq T_{H \rightarrow B+M}$. Điều đó chứng tỏ quan hệ *r* thỏa phụ thuộc yếu $H \rightarrow B+M$.

***Mệnh đề 2.1***

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

Để chuyển công thức Boole sang tập các công thức suy dẫn yêu ta thực hiện hai bước sau :

*Bước 1.* Đưa công thức $g$ về dạng chuẩn hội (CNF), ta thu được $g \equiv u_1u_2...u_k$, trong đó mỗi $u_i$, $1 \le i \le k$ là một tuyển. Trong đó CNF là công thức dạng chuẩn hội được biểu diễn dưới dạng tích của các tuyển.

*Bước 2.* Chuyển mỗi $u_i$ thành dạng suy dẫn yếu tương đương: $w_i \equiv X \to Y$, $1 \le i \le k$ như sau [43]

$X = \wedge \{x \mid$ phủ định $x'$ xuất hiện trong $w_i\}$

$Y = \vee \{y \mid y$ xuất hiện trong $w_i\}$

Nếu $X$ là tập rỗng thì thay $X$ bằng 1, nếu $Y$ là tập rỗng thì thay $Y$ bằng 0.

Ta thu được tập các công thức suy dẫn yếu $W = \{w_1, w_2, \ldots, w_k\}$.

*Ví dụ 2.5*

Cho $g = (a \to b)(b + c)$. Ta có, CNF($g$) $\equiv (a \to b)(b + c) \equiv (a'+b)(b+c)$.

Vậy $W \equiv \{a \to b, 1 \to b+c\}$.

*Ví dụ* (Tiếp )

Cho $g = (a + b) \to (c \to d)$

Ta có

CNF($g$) = $(a + b' + c' + d)(a' + b + c' + d)\ (a' + b' + c' + d)$

Vậy $W \equiv \{bc \to a + d, ac \to b + d, abc \to d\}$.

Do mỗi tập CTB $P = \{g_1, g_2, \ldots, g_k\}$ chính là một tích các công thức thành phần $P \equiv p_1p_2\ldots p_k$, nên ta có ngay hệ quả sau:

***Hệ quả 2.1***

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

***Hệ quả 2.2***

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

### 2.6. Đề xuất phụ thuộc yếu xấp xỉ

Các khái niệm về phụ thuộc hàm xấp xỉ đã được Jyrky Kivinen et al đề xuất vào năm 1995 [16], trong đó thuật ngữ *xấp xỉ* được hiểu là nếu bỏ đi ít nhất ε bộ khỏi quan hệ *r* thì phần còn lại sẽ thỏa phụ thuộc hàm. Trong luận án đề xuất một 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 sành đẳng thức.

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

$$X\ \varepsilon_1\!\rightarrow\ \varepsilon_2 \vee Y$$

Ta nói quan hệ *r* 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 $\varepsilon_1$ trên tập thuộc tính *X* thì hai bộ này cũng lệch nhau không quá $\varepsilon_2$ 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.

*Ví dụ 2.6*

Trong ngày khuyến mại, một siêu thị bán mặt hàng *H* có quy định giảm giá cho mỗi khách hàng 10 đơn vị tiền tệ giả định là *T* vào thời điểm từ 1 giờ chiều trở đi. Thông tin tại quầy thu ngân được cho dưới dạng quan hệ *r* như sau:

Bảng 2.4. Quan hệ đơn hàng

| Giao dịch | Giờ | Số lượng | Đơn giá | Thành tiền |
|---|---|---|---|---|
| 1 | 10 | 5 | 10 | 50 |
| 2 | 11 | 8 | 10 | 80 |
| 3 | 12 | 4 | 10 | 40 |
| 4 | 13 | 5 | 10 | 40 |
| 5 | 14 | 8 | 10 | 70 |
| 6 | 15 | 4 | 10 | 30 |

Ta khảo sát các phụ thuộc sau đây:

- Phụ thuộc hàm:

$$h\text{: } \textit{Số lượng}, \textit{Đơn giá} \ \rightarrow\ \textit{Thành tiền}$$

- Phụ thuộc yếu:

*w: Số lượng, Đơn giá* $\rightarrow$ *Thành tiền*

- Phụ thuộc yếu xấp xỉ:

*f: Số lượng, Đơn giá* (0) $\rightarrow$ (10) *Thành tiền*

Ta thấy do hai giao dịch 1 và 4 cùng mua một số lượng nhưng vào hai thời điểm khác nhau nên mục thành tiền mang hai giá trị khác nhau, do đó quan hệ *r* không thỏa phụ thuộc hàm *h*.

Tương tự, quan hệ *r* cũng không thỏa phụ thuộc yếu *w* với cùng lý do như trên.

Tuy nhiên quan hệ *r* thỏa phụ thuộc yếu xấp xỉ với ý nghĩa, hai khách hàng mua cùng một lượng thì trả tiền chênh nhau tối đa là 10 *T*.

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 $u = (u_1, u_2, \ldots, u_n)$, $v = (v_1, v_2, \ldots, v_n)$ trong quan hệ *r*, ta định nghĩa:

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

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

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

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

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, $\varepsilon_1$, $\varepsilon_2$, λ, 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***

(*i*) Giả sử quan hệ *r* thỏa phụ thuộc yếu xấp xỉ *f*. Ta cần chứßng minh $T_r \subseteq T_f$. Nếu *r* là quan hệ rỗng thì do $T_r = \varnothing \subseteq T_f$ nên $T_r \subseteq T_f$. Ngoài ra, ta biết ε ≥ 0. Theo tính chất của các hàm λ, độ đo d và cách xây dựng vector $t(u,v)$, ta có $t(u,u) = e$ (*e* là vector đơn vị). Từ đó suy ra $t(u,u) = e \in T_r$. Vì mọi công thức yếu dạng $f$: $X \rightarrow Y$, $Y \neq \varnothing$ đều là công thức Boole dương nên $e \in T_f$. Giả sử $t \in T_r$ theo định nghĩa của $T_r$, trong quan hệ *r* tồn tại hai bộ *u* và *v* để $t(u,v) = t$. Do quan hệ *r* thỏa f nên ta có $f(t)$, Điều này chứng tỏ $t \in T_f$.

Đảo lại, giả sử ta có $T_r \subseteq T_f$

Ta cần chứng minh quan hệ *r* thỏa phụ thuộc yếu xấp xỉ *f*. Với mọi cặp bộ *u*, *v* trong *r*, ta có $t(u,v) \in T_r$. Do $T_r \subseteq T_f$ nên $t(u,v) \in T_f$, tức là $f(t(u,v))$, điều này có nghĩa là *r* thỏa phụ thuộc yếu xấp xỉ *f*.

(*ii*) Theo định nghĩa của $T_F$ ta có

$$T_F = \cap\{T_f : f \in F\}$$

Theo phần chứng minh (*i*) quan hệ *r* thỏa tập phụ thuộc yếu xấp xỉ *F* khi và chỉ khi với mọi *f* trong *F* ta có $T_r \subseteq T_f$. Do $T_F$ là giao của các $T_f$ nên $T_r \subseteq T_F$. Đảo lại, nếu $T_r \subseteq T_F$ thì $T_r \subseteq T_f$ với mọi *f* trong *F*. Theo phàn chứng minh (i) thì *r* thỏa mọi phụ thuộc yếu xấp xỉ *f*. Từ đó suy ra quan hệ *r* thỏa tập phụ thuộc yếu xấp xỉ *F*. Định lý được 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. Như trong hình sự và an ninh cần tìm các đối tượng tóc hơi nhạt, da ngăm đen, cao khoảng 1m60. Trong kinh doanh cần cung cấp cho khách hàng các loại xe hơi công suất trên 3.0, màu mận chín,…

Các hướng nghiên cứu về cơ sở dữ liệu mờ và thô cũng cho phép chúng ta tìm kiếm các truy vấn thuộc loại trên. Tuy nhiên, vấn đề vẫn chưa được giải quyết hoàn toàn khi sự phụ thuộc giữa các thuộc tính được biểu diễn dưới dạng logic tổng quát kèm theo độ đo phản ánh tính xấp xỉ. Đây vẫn là một vấn đề mở đáng được quan tâm và nghiên cứu.

Cùng với sự phát triển ngày càng mạnh mẽ của khoa học và kỹ thuật, các loại phụ thuộc như đã nêu trên đã và đang xuất hiện rộng rãi trong nhiều lĩnh vực như di truyền học, vật lý học, sinh học phân tử, công nghệ vật liệu... Sự cần thiết và hợp lý của một công cụ toán học để mô tả và biểu diễn các loại phụ thuộc này trong các cơ sở dữ liệu là điều hiển nhiên và tất yếu

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

Cho tập *U* gồm *n* thuộc tính, một công thức Bool *f* dương trên *U*, và các hàm $\lambda_i$ trên mỗi thuộc tính *i* trong *U*. Cho các ngưỡng $\varepsilon_i$ 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 Bool dương *f* thỏa quan hệ *r* 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:

$$t(u, v) = (t_1, t_2, \dots, t_n)$$
$$t_i = (|\lambda_i(u_i) - \lambda_i(v_i)| \le \varepsilon_i)?\ 1 : 0$$

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*.

*Ví dụ 2.7*

Cho phụ thuộc Boole dương $AB \rightarrow C+D$ thể hiện các thông tin thi biểu diễn nghệ thuật vòng 1 của học sinh, trong đó

- Thuộc tính *A* là trang phục gồm các trị {rất đẹp, đẹp, bình thường}
- Thuộc tính *B* là kỹ thuật biểu diễn gồm các trị {rất hay, hay, bình thường}
- Thuộc tính C là khả năng vào vòng 2 gồm các trị {1, 0}
- Thuộc tính D là khả năng nhận phần thưởng sau vòng 1 gồm các trị {1, 0}

Giả sử nhà trường quy định các hàm λ như sau:

Bảng 2.5. Bảng quy định các hàm $\lambda_A$

| $\lambda_A$ | rất đẹp | đẹp | bình thường |
|---|---|---|---|
| | 10 | 8 | 5 |
| $\lambda_B$ | rất hay | hay | bình thường |
| | 5 | 4 | 2 |

$\lambda_c(x) = x$; $\lambda_D(x) = x$

Với ngưỡng $\varepsilon = (A{:}\,2, B{:}\,1, C{:}\,0, D{:}\,0)$, thì hai đội lớp 3 có kết quả

Đội 3A1: *u* = (trang phục: rất đẹp, kỹ thuật: hay, vào vòng 2: 1, thưởng: 1)

Đội 3A2: *v* = (trang phục: đẹp, kỹ thuật: rất hay, vào vòng 2: 0, thưởng: 1)

sẽ cho các kết quả trung gian như sau:

$t_A = 1$, vì $|\lambda_A(rất\ đẹp) - \lambda_A(đẹp)| = 10 - 8 = 2 \leq \varepsilon_A$

$t_B = 1$, vì $|\lambda_B(hay) - \lambda_B(rất\ hay)| = |4 - 5| = 1 \leq \varepsilon_B$

$t_C = 0$, vì $|\lambda_C(1) - \lambda_C(0)| = 1 - 0 = 1 > \varepsilon_C$

$t_D = 1$, vì $|\lambda_D(1) - \lambda_D(1)| = 1 - 1 = 0 \leq \varepsilon_D$

sẽ có *t(u,v)* = (1, 1, 0, 1) và cho ta $1 \wedge 1 \rightarrow 0 \vee 1$, hay $1 \rightarrow 1 = 1$.

***Định nghĩa 2.2***

Ta gọi bộ bốn *p* = (*U*, λ, ε, *F*) là một *lược đồ quan hệ với phụ thuộc Bool dương xấp xỉ, trong đó*

- *U là tập* n thuộc tính,
- *F* là tập các công thức Bool 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 lươc đồ quan hệ với phụ thuộc Bool dương xấp xỉ $p = (U, \lambda, \varepsilon, F)$ và một phụ thuộc Bool 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 $T_r \subseteq T_F$ :

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

***Chứng minh***

(i) Theo định nghĩa của $T_F$ ta có

$$T_F = \cap\{T_f : f \in F\}$$

Giả sử quan hệ $r$ thỏa phụ thuộc Boole dương $f$. Ta cần chứng minh $T_r \subseteq T_f$

Nếu $r$ là quan hệ rỗng thì do $T_r = \emptyset \subseteq T_f$ nên $T_r \subseteq T_f$.

Ngoài ra, với mọi thuộc tính $i$, $1 \leq i \leq n$, ta biết $\varepsilon_i \geq 0$. Theo tính chất của các hàm định lượng $\lambda$ ta có

$$|\lambda_i(x) - \lambda_i(x)| = 0 \leq \varepsilon_1$$

Từ đó suy ra

$$t(u,u) = e \in T_r$$

trong đó $e$ là vector đơn vị.

Vì $f$ là công thức Boole dương nên $e \in T_f$.

Giả sử $t \in T_r$. Theo định nghĩa của $T_r$, trong r tồn tại hai bộ $u$ và $v$ để $t(u,v) = t$. Do quan hệ $r$ thỏa $f$ nên ta có $f(t)$, Điều này chứng tỏ $t \in T_f$ .

Đảo lại, giả sử ta có $T_r \subseteq T_f$

Ta cần chứng minh quan hệ $r$ thỏa phụ thuộc Boole dương xấp xỉ $f$. Với mọi cặp bộ $u$, $v$ trong $r$, ta có $t(u,v) \in T_r$. Do $T_r \subseteq T_f$ nên $t(u,v) \in T_f$, tức là $f(t(u,v))$, điều này có nghĩa là $r$ thỏa phụ thuộc Boole dương xấp xỉ $f$.

(*ii*) Theo định nghĩa của $T_F$ ta có

$$T_F = \cap\{T_f : f \in F\}$$

Theo phần chứng minh (*i*) quan hệ $r$ thỏa tập phụ thuộc Boole dương xấp xỉ $F$ khi và chỉ khi với mọi $f$ trong $F$ ta có $T_r \subseteq T_f$. Do $T_F$ là giao của các $T_f$ nên $T_r \subseteq T_F$. Đảo lại, nếu $T_r \subseteq T_F$ thì $T_r \subseteq T_f$ với mọi $f$ trong $F$. Theo phần chứng minh (*i*) thì $r$ thỏa mọi phụ thuộc Boole dương xấp xỉ $f$. Từ đó suy ra quan hệ $r$ thỏa tập phụ thuộc Boole dương xấp xỉ $F$.

Định lý được chứng minh.

***Xây dựng phụ thuộc Boole dương xấp xỉ thông qua các phụ thuộc yếu xấp xỉ***

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

*Ví dụ 2.8*

*Cho* $f = (a + b) \rightarrow (c \rightarrow d)$ là một phụ thuôc Boole dương trên $U = \{a,b,c,d\}$ và hai ngưỡng $\varepsilon_1$, $\varepsilon_2$. *Ta có các phụ thuộc xấp xỉ sau đây*

$\{bc\ \varepsilon_1 \rightarrow \varepsilon_2\ a + d,\ ac\ \varepsilon_1 \rightarrow \varepsilon_2\ b + d,\ abc\ \varepsilon_1 \rightarrow \varepsilon_2\ d\}$.

### 2.8. Đề xuất thuộc Boole dương xấp xỉ tổng quát

#### *2.8.1. Xây dựng phép sánh trị alpha dựa trên hàm lambda*

Hàm alpha là quan hệ bộ phận thoả mãn tính chất đối xứng, phản xạ được xác định từ miền trị của các thuộc tính trên không gian số thực dựa trên hàm

Lambda.

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 $\lambda_A$ . Ta định nghĩa

$$\forall a,\ b \in d_A:\ \alpha_A(a,b) \equiv (\lambda_A(a) = \lambda_A(b))\ (2.3)$$

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

***Định lý 2.5***

*Hàm $\alpha_A$ được định nghĩa qua hệ thức thỏa ba tính chất phản xạ, đối xứng và bộ phận trong định nghĩa ánh xạ $\alpha$.*

***Chứng minh***

Giả sử $a \in d_i$. Khi đó $\lambda_A(a) = \lambda_A(a)$ nên $\alpha_A(a,a) = 1$. Tính chất phản xạ A1 được chứng minh.

Giả sử $a,\ b \in d_i$ và $\alpha_A(a,b) = 1$. Khi đó $\lambda_A(a) = \lambda_A(b)$, do đó $\alpha_A(b,a) = 1$,

Đảo lại, nếu $\alpha_A\ (a,b) = 0$, thì $\lambda_A(a) \neq \lambda_A(b)$, do đó $\alpha_A(b,\ a) = 0$. Tính chất đối xứng *A2* được chứng minh.Tính chất đối xứng *A2* được chứng minh.

Theo tính chất khác biệt của hàm lambda, $\exists x, y \in d_i$: $\lambda_A(x) \neq \lambda_A(y)$. Từ đây suy ra $\alpha_A(x,\ y) = 0$.

Tính chất bộ phận *A*3 được chứng minh.

Định lý được chứng minh.

### *2.8.2. Phụ thuộc Boole dương xấp xỉ tổng quát*

***Định nghĩa 2.3***

Cho tập thuộc tính *U* và một công thức Bool dương *f* trên *U*. Cho hai ngưỡng $\varepsilon_1$, $\varepsilon_2$ là hai số thực không âm. *Phụ thuộc Bool 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* thỏa phụ thuộc Boole dương xấp xỉ tổng quát *f* nếu *r* thỏa mọi phụ thuộc yếu xấp xỉ 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 2.4:

***Hệ quả 2.3***

*Cho lược đồ quan hệ p =(U, $\varepsilon_1, \varepsilon_2$, $\lambda$, d, F) và một Phụ thuộc Boole dương xấp xỉ tổng quát (PTBDXXTQ) 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 Bool dương tổng quát, luận án đề xuất các phụ thuộc dữ liệu mới bao gồm: phụ thuộc Bool dương xấp xỉ, phụ thuộc Bool dương xấp xỉ tổng quát và phụ thuộc yếu xấp xỉ và chứng minh sự tương đương giữa phụ thuộc Bool dương và chỉ ra và chứng minh mối quan hệ giữa phụ thuộc Bool dương tổng quát với phụ thuộc Bool dương xấp xỉ và phụ thuộc Bool 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 LƯỢC ĐỒ QUAN HỆ

## 3.1. Mở đầu

Việc sử dụng công cụ logic để miêu tả các ràng buộc dữ liệu trong cơ sở dữ liệu và chỉ ra các phép so sánh giúp người quản trị dữ liệu hiểu rõ hơn về tính đa dạng của dữ liệu trong thực tế. Trong các phần dưới đây, NCS sẽ trình bày cách sử dụng công cụ logic để giải quyết các bài toán kinh điển trong cơ sở dữ liệu. Những khái niệm và thuật toán thu được từ đó trở thành các công cụ hữu ích trong việc tổ chức và tìm kiếm dữ liệu trong cơ sở dữ liệu. Vì vậy, trong chương 3 NCS tập trung trình bày các nội dung sau:

(1) Tổng hợp qui trình giải bài toán suy dẫn theo ba tiếp cận hình thức bao gồm: chứng minh trực tiếp theo thuật giải Vương Hạo, chứng minh trực tiếp theo chuẩn hội và chứng minh phản chứng theo hợp giải. Bên cạnh đó, chúng tôi cũng đưa ra các kết quả mới về việc áp dụng các phương pháp chứng minh hằng đúng để giải quyết bài toán về thành viên trong lớp các phụ thuộc logic và thu gọn các luật suy dẫn trong cơ sở tri thức. Các phương pháp này sẽ giúp tăng tính hiệu quả và giảm thiểu độ phức tạp trong việc xử lý dữ liệu.

(2) Xây dựng thuật toán tìm bao đóng của một tập thuộc tính cho lớp các phụ thuộc logic. Bao đóng của tập thuộc tính được vận dụng trong các bài toán tìm khóa và bài toán thành viên giới hạn trên phụ thuộc hàm.

(3) Xây dựng thuật toán tìm khóa cho lớp các phụ thuộc logic. 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, CT3, CT4 - *Danh mục các công trình công bố* của NCS.

## 3.2. Tổng hợp các phương pháp chuyển công thức logic về dạng chuẩn hội

*Công thức logic dạng chuẩn hội* (CNF) đóng vai trò quan trọng trong quá trình giải các bài toán như bài toán chứng minh một công thức Bool là hằng đúng, giải bài toán thành viên, phủ và phủ không dư, bài toán tìm bao đóng và khóa,…

Để có thể mở rộng khái niệm phụ thuộc Boole dương xấp xỉ, các phần lập luận dưới đây liên quan đến việc chuyển một công thức logic bất kỳ về công thức dạng chuẩn hội (CNF) [16] [18].

***Định nghĩa 3.1***

Cho tập các biến Boole $U = \{a_1, a_2, \ldots, a_n\}$. Công thức logic $\mathcal{L}$ trên $U$ có dạng chuẩn hội nếu $\mathcal{L}$ được biểu diễn dưới dạng tích của các tuyển [16]:

$$\mathcal{L} = g_1 g_2 \ldots g_m$$

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

Ví dụ, Cho $U = \{x, y, z, w\}$. Công thức $\mathcal{L}$ sau đây thuộc dạng chuẩn hội:

$\mathcal{L} = (x+y+x')(y+w)x'(y+x)$.

***Định lý 3.1***

*Mọi công thức logic đều tương đương với một công thức CNF* [15].

Theo tài liệu [14] [18] có hai phương pháp chuyển một công thức logic về dạng CNF như sau :

***3.2.1. Phương pháp biến đổi logic***

Chuyển công thức logic $\mathcal{L}$ 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 Bool $U$. 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 [43] [44] cho đến khi thu được công thức logic có dạng CNF.

*Ví dụ 3.1*

Chuyển đổi công thức $\mathcal{L} = (b + c) \rightarrow (d \rightarrow e)$ sang dạng CNF?

Ta có $\mathcal{L} = (b + c) \rightarrow (d \rightarrow e)$

$\equiv (b + c)' + (d \rightarrow e)$ // Áp dụng luật R7

$\equiv (b + c)' + (d' + e)$ // Áp dụng luật R7

$\equiv b'c' + (d' + e)$ // Áp dụng luật R9

$\equiv (b' + d + e')\,(c' + d' + e)$ // Áp dụng luật R10

Vậy, $\mathcal{L} \equiv (b' + d' + e)(c' + d' + e)$ có dạng CNF.

### *3.2.2. Phương pháp lập bảng*

Chuyển công thức logic $\mathcal{L}$ trên tập thuộc tính $U$ về dạng CNF theo phương pháp lập bảng với các bước sau đây:

1. Lập bảng trị $T_{\mathcal{L}}$ theo các biến trong $U$ và công thức logic $\mathcal{L}$

2. Từ mỗi dòng $t = (v_1,\ldots,v_n)$ thỏa $\mathcal{L}(t) = 0$ trong bảng trị $T_{\mathcal{L}}$ tạo ra một tuyển cơ sở $g_t = (z_1 + \ldots + z_n)$, với

$$z_i = \begin{cases} x_i \text{ nếu } v_i = 0 \\ x_i' \text{ nếu } v_i = 1 \end{cases} \qquad 1 \le i \le n$$

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

Kết quả thu được sau bước 3 chính là dạng chuẩn hội của công thức $\mathcal{L}$.

*Ví dụ 3.2*

Chuyển công thức $\mathcal{L} = (a + b) \rightarrow (\mathrm{c} \rightarrow d)$ về dạng CNF?

Ta có bảng 3.1 thể hiện các bước chuyển công thức logic $\mathcal{L}$ về dạng CNF:

Bảng 3.1. Chuyển $\mathcal{L}$ về dạng CNF

| ***a*** | ***b*** | ***c*** | ***d*** | $\boldsymbol{a+b}$ | $\mathbf{c}\rightarrow\boldsymbol{d}$ | $(\boldsymbol{a+b})\rightarrow(\mathbf{c}\rightarrow\boldsymbol{d})$ | ***M*** |
|---|---|---|---|---|---|---|---|
| (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| 0 | 0 | 0 | 0 | 0 | 1 | 1 | |
| 0 | 0 | 0 | 1 | 0 | 1 | 1 | |
| 0 | 0 | 1 | 0 | 0 | 0 | 1 | |
| 0 | 0 | 1 | 1 | 0 | 1 | 1 | |
| 0 | 1 | 0 | 0 | 1 | 1 | 1 | |
| 0 | 1 | 0 | 1 | 1 | 1 | 1 | |
| **0** | **1** | **1** | **0** | 1 | 0 | **0** | *a* + *b*’ + c’ + *d* |
| 0 | 1 | 1 | 1 | 1 | 1 | 1 | |
| 1 | 0 | 0 | 0 | 1 | 1 | 1 | |
| 1 | 0 | 0 | 1 | 1 | 1 | 1 | |
| **1** | **0** | **1** | **0** | 1 | 0 | **0** | *a*’ + *b* + *c*’ + *d* |
| 1 | 0 | 1 | 1 | 1 | 1 | 1 | |
| 1 | 1 | 0 | 0 | 1 | 1 | 1 | |
| 1 | 1 | 0 | 1 | 1 | 1 | 1 | |
| **1** | **1** | **1** | **0** | 1 | 0 | **0** | *a*’ + *b*’ + *c*’ + *d* |
| 1 | 1 | 1 | 1 | 1 | 1 | 1 | |

Vậy, CNF $\mathcal{L} \equiv M \equiv (a+b'+c'+d)(a'+b+c'+d)(a'+b'+c'+d)$.

Trong bảng 3.1, các cột (1) – (4) chứa các trị 0/1 cho các biến *a*, *b*, *c* và *d* ; các cột (5) – (7) là công thức logic trong $\mathcal{L}$, cột (8) là tuyển cơ sở của các dòng tại đó $\mathcal{L}$ nhận giá trị 0.

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 [CT4].

## 3.3. Xây dưng phương pháp chứng minh công thức hằng đúng

Hiện nay, bài toán hằng đúng (tautology)  được vận dụng với tần suất cao trong các lĩnh vực dữ liệu lớn, khai thác tri thức và trí thệ nhân tạo nói chung. Khi vận dụng, ngoài việc chứng minh các định lý toán học đòi hỏi các luật chính xác, trong thực tiễn, các luật thu được chỉ có một độ chính xác nhất định, do đó việc nghiên cứu các bài toán tựa-tautology, tức là bài toán xác định các biểu thức logic đa trị mang trị true với một độ chắc chắn nhất định có ý nghĩa thiết thực hơn. Tuy nhiên, lĩnh vực nghiên cứu mới này cũng được xây dựng trên cơ sở lý thuyết của bài toán tautology tryền thống.

***Định nghĩa 3.2***

Cho tập ký biến Bool *U*. Một công thức Bool *f* trên được gọi là *hằng đúng* (*tautology*) nếu $f(x) = 1$ với mọi phép gán trị *x* [18].

Như vậy chứng minh $f \rightarrow g$ tương đương với chứng minh công thức $(f \rightarrow g)$ là công thức hằng đúng [5].

*Ví dụ 3.3*

Các công thức sau đây là hằng đúng:

$$a \rightarrow a$$

$$a + a'$$

$$ab \rightarrow a$$

Công thức $f(a,b) = ab+b'$ không phải là hằng đúng vì $f(0,1) = 0\cdot1 + 1' = 0 + 0 = 0$.

Luận án xây dựng các phương pháp chứng minh CTB *f* là hằng đúng thực hiện theo một trong ba phương pháp sau [CT4]:

### ***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 [CT4]:

*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 $u_i = 1$, $1 \le i \le k$.

*Ví dụ 3.4*

Khảo sát hằng đúng sau:

$$f \equiv ((a \rightarrow b)(b \rightarrow c)) \rightarrow (a \rightarrow c)$$

*Pha 1*. CNF. Ta có

$$f \equiv ((a \to b)(b \to c)) \to (a \to c)$$
$$\equiv ((a'+b)(b'+c))' + (a'+c)$$
$$\equiv (a' + b)' + (b' + c)' + (a' + c)$$
$$\equiv a''b' + b''c' + a' + c$$
$$\equiv ab' + bc' + a' + c$$
$$\equiv ab' + bc' + a' + c$$
$$\equiv ab' + (b + a')(c' + a') + c$$
$$\equiv ab' + (b + a' + c)(c' + a' + c)$$
$$\equiv (ab' + b + a' + c)(ab' + c' + a' + c)$$
$$\equiv (a + b + a' + c)(b' + b + a' + c)(a + c' + a' + c)(b' + c' + a' + c)$$

*Pha 2*. Kết luận

$(a + b + a' + c) \equiv 1$, vì có $a + a' = 1$

$(b' + b + a' + c) \equiv 1$ , vì có $b + b' = 1$

$(a + c' + a' + c) \equiv 1$, vì có $c' + c = 1$

$(b' + c' + a' + c) \equiv 1$, vì có $c' + c = 1$

Vậy $f$ là công thức hằng đúng.

### ***3.3.2. Phương pháp Vương Hạo***

Phương pháp Vương Hạo dùng để chứng minh tautology $T \to P$ theo các bước sau [CT4]

*Bước 1* (CNF $\to$ DNF).

Đưa vế trai $T$ về dạng CNF.

Đưa vế phải $i$ về dạng DNF là công thức logic có *dạng chuẩn tuyển,* cụ thể là $P$ được biểu diễn dưới dạng tuyển của các hội;

Để đưa công thức $T \to P$ về dạng CNF $\to$ DNF ta có thể vận dụng các luật (R1) – (R11) để biến đổi hai vế $T$ và $P$ thành dạng chuẩn sau:

$$t_1\ t_2 \dots t_u \to p_1 + p_2 + \dots + p_v$$

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

$$t_1 \ldots (a+b) \ldots t_u \to p_1 + \cdots + p_v$$

Thì thay bằng 2 dòng:

$$\begin{cases} t_1 \ldots a \ldots t_u \to p_1 + \cdots + p_v \\ t_1 \ldots b \ldots t_u \to p_1 + \cdots + p_v \end{cases}$$

- Dạng 2

$$t_1 \ldots t_u \to p_1 + \ldots + a{\cdot}b + \ldots + p_v$$

Thì thay bằng 2 dòng:

$$\begin{cases} t_1 \ldots t_u \to p_1 + \ldots + a + \ldots + p_v \\ t_1 \ldots t_u \to p_1 + \ldots + b + \ldots + p_v \end{cases}$$

*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.

*Ví dụ 3.5*

Khảo sát tautology

$$f \equiv ((a \to b)(b \to c)) \to (a \to c)$$

Ta có:

*Vết trái* $T \equiv (a \to b)(b \to c) \equiv (a' + b)(b' + c)$ // CNF

*Vế phải* P $\equiv (a \to c) \equiv$ a' + c // DNF

$(a' + b)(b' + c) \to a' + c$

$(a' + b)(b' + c)a \to c$ *// chuyển a qua trái*

$$\begin{cases} a' \, (b' + c)a \to c \\ b \, (b' + c)a \to c \end{cases} \quad \text{// tách cụm } (a'+b)$$

*// Tách cum (b'+c)*

$$\begin{cases} a' \, b' \, a \to c \\ a' \, c \, a \to c \\ b \, b' \, a \to c \\ b \, c \, a \to c \end{cases}$$

// chuyển các $x'$ sang vế phải

$$\begin{cases} a \rightarrow \mathrm{c} + a + b \\ c\ a \rightarrow c + a \\ b\ a \rightarrow c + \mathrm{b} \\ b\ c\ a \rightarrow c \end{cases}$$

// Xét 2 vế có cùng ký biến

$$\begin{cases} a \rightarrow c + a + b \text{ (được chứng minh/đcm)} \\ c\ a \rightarrow c + a \quad \text{(đcm)} \\ b\ a \rightarrow c + b \quad \text{(đcm)} \\ b\ c\ a \rightarrow c \quad \text{(đcm)} \end{cases}$$

// chuyển các $x'$ sang vế phải

$$\begin{cases} a \rightarrow \mathrm{c} + a + b \\ c\ a \rightarrow c + a \\ b\ a \rightarrow c + \mathrm{b} \\ b\ c\ a \rightarrow c \end{cases}$$

Xét 2 vế có cùng ký biến

$$\begin{cases} a \rightarrow c + a + b \text{ (được chứng minh/đcm)} \\ c\ a \rightarrow c + a \quad \text{(đcm)} \\ b\ a \rightarrow c + b \quad \text{(đcm)} \\ b\ c\ a \rightarrow c \quad \text{(đcm)} \end{cases}$$

*Bước 4* (Kết luận): Vì 4 dòng đều được chứng minh nên bài toán được chứng minh, có nghĩa là $f$ là tautology.

### ***3.3.3. Phương pháp hợp giải***

Trong logic, phương pháp hợp giải [18] được sử dụng để giải bài toán suy dẫn $f \rightarrow g$. Để chứng minh $f \rightarrow g$ là tautology ta xét phủ định $Q = (f \rightarrow g)'$ $\equiv f\, g'$. Nếu ta chứng minh được $Q$ là sai, thì $f \rightarrow g$ là hằng đúng. Do đó, hợp giải còn được gọi là phương pháp chứng minh *bằng phản chứng* [CT4].

Cho công thức logic $Q$ 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 $B + C$, trong đó $x$ là một biến logic, $B$ 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 $Q = \varnothing$ thì ta nói là *hợp giải* *thành công*, tức là $Q \to false$, ngược lại ta nói hợp giải *không thành công*. Khi hợp giải thành công, do $Q \to false$ nên ta kết luận $Q$ là hằng sai.

**Thuật toán 3.1** (Thuật toán hợp giải)

```
Algorithm  Unif(H)
Input      CNF(H)
Output     unif(W)
Begin
  W := H;
  while there are terms (x + B) and
      (x' + C)  in W do
       delete (x + B) in W;
       delete (x' + C) in W;
       (B + C) to W
  endwhile
  return W;
End Unif
```

*Ví dụ 3.6*

Chứng minh $f \to g$, trong đó

$$f \equiv (a \to b)(b \to c)$$

Và $$g \equiv a \to c$$

Ta sẽ dùng kỹ thuật hợp giải.

Xét $fg' \equiv (a \to b)(b \to c)(a \to c)'$

$\equiv (a \to b)(b \to c)ac'$

$\equiv (a'+b)(b'+c)ac'$ (CNF)

Hợp giải

$$(\underline{a}'+b)(b'+c)\underline{a}c' \to \underline{b}(\underline{b}'+c)c' \to cc' \equiv \varnothing.$$

Vậy hợp giải thành công, nghĩa là $fg' \to$ false. Vậy $f \to g$.

***Mệnh đề 3.1***

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

*(i) Chứng minh theo phương pháp Vương Hạo.*

*(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. Ta quy ước rằng mỗi miền trị $d_i$ của thuộc tính i, $1 \le i \le n$, 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:

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

$$t_1 = (u_i = v_i)\ ?\ 1 : 0$$

Ta gọi *t*(*u*, *v*) là phép *gán trị đẳng thức* trên cặp bộ *u* và *v* trong quan hệ *r*.

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

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

và gọi $T_r$ *là bảng chân lý* của quan hệ *r*. Theo định nghĩa ta thấy:

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

*Ví dụ 3.7:* Xét quan hệ *r* dưới đây. Ta có bảng chân lý $T_r$ như sau

Bảng 3.2. Quan hệ r và $T_r$

***r***

| Bộ | *A* | *B* | *C* |
|---|---|---|---|
| U | 14 | 7.5 | Hội An |
| *V* | 6 | 7.5 | Hà Nội |
| *w* | 1 | 2.0 | Hội An |

***$T_r$***

| Cặp bộ | *A* | *B* | *C* |
|---|---|---|---|
| $t(u,u)$ | 1 | 1 | 1 |
| t(u,$v$) | 0 | 1 | 0 |
| t(u,$w$) | 0 | 0 | 1 |
| t(v,$w$) | 0 | 0 | 0 |

***Định nghĩa 3.4*** [12]

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

Cho quan hệ *r* trên *U*. Ta nói quan hệ *r* thỏa phụ thuộc Boole dương *f* và ký hiệu *r*(*f*) nếu với mọi cặp bộ *u*, *v* trong *r*:

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

Cho quan hệ *r* trên *U*. Ta nói quan hệ *r* thỏa tập phụ thuộc Boole dương *F* và ký hiệu *r*(*F*) nếu *r* thỏa mọi phụ thuộc trong *f*:

$$r(F) \Leftrightarrow \forall f \in F\text{: } r(f)$$

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

***Định lý 3.2***

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

(i) Quan hệ *r* thỏa phụ thuộc Boole dương *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 phụ thuộc Boole dương *F* khi và chỉ khi và chỉ khi $T_r \subseteq T_F$

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

***Chứng minh***

(i) Nếu *r* là quan hệ rỗng thì do $T_r = \varnothing \subseteq T_f$ nên $\mathrm{T_r} \subseteq T_f$.

Nếu *r* là quan hệ không rỗng, giả sử $u \in r$, thì do $t(u,u) = e \in T_r$ và do *f* là một công thức Boole dương nên $e \in T_f$. Giả sử hai bộ *u*, *v* trong *r* cho ta phép gán trị $t(u,v) \in T_r$. Theo định nghĩa của tính thỏa của phụ thuộc Boole dương, ta có $f(t(u,v)) = 1$. Điều này chứng tỏ $t(u,v) \in T_f$.

Đảo lại, giả sử quan hệ *r* có $T_r \subseteq T_f$. Ta cần chứng minh *r* thỏa phụ thuộc Boole dương *f*. Giả sử hai bộ *u, v* trong *r* có $t(u,v) \in T_r$. Vì $T_r \subseteq T_f$ nên $t(u,v) \in T_f$. Điều này chứng tỏ $t(u,v)$ thỏa công thưc Boole dương *f*.

(ii) Theo định nghĩa của $T_F$ ta có

$$T_F = \cap\{T_f : f \in F\}$$

Theo phần chứng minh ($i$) quan hệ $r$ thỏa tập phụ thuộc Boole dương $F$ khi và chỉ khi với mọi f trong F ta có $T_r \subseteq T_f$. Do $T_F$ là giao của các $T_f$ nên $T_r \subseteq T_F$. Đảo lại, nếu $T_r \subseteq T_F$ thì $T_r \subseteq T_f$ với mọi $f$ trong $F$. Theo phần chứng minh ($i$) thì $r$ thỏa mọi phụ thuộc Boole dương $f$. Từ đó suy ra quan hệ $r$ thỏa tập phụ thuộc Boole dương $F$.

Định lý được chứng minh.

Định lý tương đương [12]

*Cho tập PTBD F và một phụ thuộc Boole dương f. Ba tiên đề sau là tương đương,*

(i) $F \models f$ (*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*)

Đối với phụ thuộc hàm ta đã định nghĩa quan hệ $r$ thỏa PTH $f: X \to Y$, và ký hiệu là $r(f)$ nếu

$$\forall u,v \in r: u.X = v.X \Rightarrow u.Y = v.Y$$

Khi coi PTH như là một trường hợp riêng của CTBD ta đã chấp nhận định nghĩa quan hệ $r$ thỏa PTH $f: X \to Y$ nếu $T_r \subseteq T_f$.

- Quan hệ $r$ thỏa phụ thuộc hàm $f: X \to Y$ *trên U khi và chỉ khi* $T_r \subseteq T_f$.

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

- Quan hệ $r$ thỏa tập phụ thuộc hàm $F$ *trên U khi và chỉ khi* $T_r \subseteq T_F$:

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

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

Bảng 3.3. Các lớp phụ thuộc Boole dương

| 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 | $X \twoheadrightarrow Y$ | $\wedge X \rightarrow (\wedge Y) \vee (\wedge(U\backslash X\backslash Y))$ |

### 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 như sau. Như ta đã biết, 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 nới lỏng và các phụ thuộc mạnh, yếu và đỗi ngẩu đều được biểu diễn qua các công thức suy dẫn $X \rightarrow Y$ với $X$ và $Y$ là các tập con thuộc tính và được gán ngữ nghĩa như dạng hội logic. Vấn đề đặt ra là: các phụ thuộc dạng hàm $X \rightarrow Y$ có đủ để mô tả các loại ràng buộc trong thực tiễn không? Câu trả lời là phủ định. Tồn tại những loại ràng buộc (logic) không thể được mô tả thông qua các phụ thuộc dạng hàm.

#### *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 *Thành tiền = số lượng × đơn giá*

Các biểu thức cũng có thể là những mệnh đề, 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.

* Trong kỳ thi 4 môn chuyên $A$, $B$, $C$, $D$ của cụm trường, hai thí sinh của cùng một trường phải đăng kí hai môn thi khác nhau.

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 $\gamma$ và $\lambda$ là các tân từ xác định trên $f$ và cặp thuộc tính $u$, $v$.

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

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

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

$$r(\Sigma) \overset{def}{\Longleftrightarrow} \forall f \in \Sigma: r(f)$$

Phụ thuộc hàm truyền thống và các biến thể của phụ thuộc hàm truyền thống được đặc tả như sau

Bảng 3.4. Đặc tả các loại phụ thuộc phụ thuộc hàm truyền thống phụ thuộc mạnh, yếu và đối ngẫu

| | |
|---|---|
| PTH (truyền thống) | $\forall u, v \in R$: Eq$(u.X, v.X) \Rightarrow$ Eq$(u.Y, v.Y)$<br>Eq$(u.X, v.X) \overset{def}{\Longleftrightarrow} \forall A \in X: u.A = v.A$ |
| PT mạnh | $\forall u, v \in R$: Eq1$(u.X, v.X) \Rightarrow$ Eq$(u.Y, v.Y)$<br>Eq1$(u.X, v.X) \overset{def}{\Longleftrightarrow} \exists A \in X: u.A = v.A$ |
| 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$(u.Y, v.Y)$ |

### *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 $\Sigma$.

- Thao tác thêm: $r' = r \cup \{t\}$: thêm một bộ t vào quan hệ $r$.
- Thao tác xóa: $r' = r - \{t\}$: loại bỏ bộ t khỏi quan hệ $r$.
- Thao tác sửa: sửa bộ $t$ trong $r$ thành bộ $t'$, $r' = r - \{t\} \cup \{t'\}$

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

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

Nói chung, tính chất $P$ của một tập $S$ có thể không bất biến đối với thao tác *xóa*. Ví dụ, $S$ là một tập hợp số thoả tính chất $P$: $S$ có ít nhất $k$ số chẵn ($k$ là hằng số). Ta thấy $P$ là bất biến đối với thao tác thêm nhưng không bất biến đối với thao tác xóa.

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 $\Sigma$ 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 phụ thuộc hàm 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, thí dụ,

* Hãy cho biết $t$ xuất hiện tại vị trí nào trong $r$?

* Hãy cho biết $t$ và những bộ gần với $t$ xuất hiện tại những vị trí nào trong $r$? *Thuật ngữ "gần" có thể được hiểu là xấp xỉ theo một tiêu chí nào đó.*

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* [34] : Cho lược đồ quan hệ $p = (U,\Sigma)$ và quan hệ $r$ trên $p$, ta có $r(\Sigma)$, nghĩa là $r$ thỏa mọi phụ thuộc trong $\Sigma$. Ngoài $\Sigma$, $r$ 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 $\Sigma$, $r$ còn thỏa phụ thuộc $A\rightarrow C$.

Tổng quát, cho LĐQH $p = (U, \Sigma)$ và một phụ thuộc $g$ trên $U$. Ta nói phụ thuộc $g$ được *suy dẫn theo quan hệ* từ tập phụ thuộc $\Sigma$ và viết $\Sigma \vdash g$ nếu $\forall r\in REL(U)$: $r(\Sigma)\Rightarrow r(g)$. Kí hiệu, $\Sigma^+ = \{\ g: \Sigma \vdash g\ \}$ và gọi là *bao đóng* của $\Sigma$.

Bài toán thành viên khi đó được phát biểu như sau:

Giả thiết: Cho lược đồ quan hệ $p = (U, \Sigma)$ 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 [34] Ngoài ra bài toán thành viên giúp ta thu gọn tập phụ thuộc $\Sigma$, tức là thay thế $\Sigma$ bằng tập phụ thuộc tương đương với kích thước nhỏ gọn hơn, do đó sẽ giảm được độ phức tạp của các thuật toán. Muốn giải bài toán thành viên ta phải kiểm tra mọi quan hệ trên lược đồ quan hệ $p$. Độ phức tạp tính toán khi đó sẽ rất lớn, đặc biệt khi gặp các thuộc tính với miền trị vô hạn, như kinh phí, hoặc khoảng cách giữa các thiên thể. Từ đó phát sinh nhu cầu tất yếu sau đây: 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 $\Sigma \models g$ ?

Đầu tiên, vấn đề này được Armstrong giải quyết trọn vẹn với phụ thuộc hàm truyền thống [26] như sau

Cho lược đồ quan hệ $p = (U, \Sigma)$ 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 $\Sigma$ và viết $\Sigma \vdash_2 g$ nếu $\forall r \in REL_2(U)$: $r(\Sigma) \Rightarrow r(g)$.

*Định lý tương đương* [34] [40]

Cho lược đồ quan hệ $p = (U, \Sigma)$ 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 phụ thuộc hàm truyền thống.

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

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

Theo định lý tương đương, muốn chứng minh $\Sigma \vdash g$ thì vận dụng 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 $g$.

Tiếp theo, nhóm nghiên cứu Fagin et al. đề xuất phụ thuộc logic và chứng minh định lý tương đương cho loại phụ thuộc này [12] Sau đó nhóm Blok phát hiện ra rằng định lí tương đương không đúng cho các phụ thuộc logic nói chung mà chỉ đúng cho các công thức Boole dương.

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

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

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

Năm 1992 Nguyễn Xuân Huy cùng nhóm nghiên cứu của mình đã mở rộng khái niệm sánh trị và cung cấp mô hình phụ thuộc Boole dương 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*

Các *công thức suy dẫn* (*CTSD*) có dạng $X \to Y$, trong đó $X$ và $Y$ là các hội logic của hữu hạn biến được sử dụng khá rộng rãi trong tin học. Chúng đóng vai trò chủ yếu trong các motor suy dẫn của các hệ chuyên gia, trong việc thể hiện các ràng buộc dữ liệu của các cơ sở dữ liệu cũng như trong các thuật toán trích chọn luật từ các kho dữ liệu thuộc lĩnh vực khai thác tri thức[45].

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 [1].

Ta định nghĩa $\forall v \in \mathcal{B}^n$, $Set(v) = \{A \in U \mid v.A = 1\}$, khi đó giữa $\mathcal{B}^n$ 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 $f$: $X \to Y$ là một công thức suy dẫn 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 công thức suy dẫn trên tập biến $U$.

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

Cho $V$ là tập các phép gán trị trên $U$. Với hai phép gán trị $u,v \in V$ ta xét phép toán &, $u$ & $v$, như là phép nhân logic trên các thành phần tương ứng của $u$ và $v$. Ta quy ước tích của một tập rỗng các phần tử trong $V$ chính là phép gán trị đơn vị $e = (1,...1)$. Tập các phép gán trị $V$ được gọi là *đóng đối với phép nhân &* nếu $V$ chứa tích của mọi cặp phần tử trong $V$, tức là $\forall u,v \in V$: $u\&v \in V$. Dễ thấy, $Set(u\&v) = Set(u) \cap Set(v)$, tức là song ánh giữa $\mathcal{B}^n$ và $2^U$ trở thành đẳng cấu với hai phép toán tương ứng là & và $\cap$.

*Hội suy dẫn*

Ta cũng biết mọi công thức logic đều có thể biểu diễn dưới dạng chuẩn tuyển (hội). Nói cách khác, mỗi bảng $T \subseteq \mathcal{B}^n$ đều ứng với một công thức logic dạng chuẩn tuyển (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 [40]. 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* [40].

***Bổ đề 3.1***

(Bổ đề về tính đóng của phép & trong $T_f$ [40]. ) *Với mỗi công thức suy dẫn f trên U*, $T_f$ *chứa các phép gán trị đơn vị e, gán trị không z và đóng với phép &.*

***Chứng minh***

Cho CTSD $f: X \rightarrow Y$ và $e$ là phép gán trị đơn vị, $z$ là phép gán trị 0. Dễ thấy $f(e)=f(z)=1$, do đó $e, z \in T_f$. Giả sử $u,v \in T_f$ . Đặt $t = u\&v$, ta cần chứng minh $t \in T_f$ . Giả sử $Set(t) \supseteq X$. Vì $Set(t) = Set(u\&v) = Set(u) \cap Set(v)$ nên $Set(u) \supseteq X$ và $Set(v) \supseteq X$. Vì $f(u) = f(v) = 1$ nên ta phải có $Set(u) \supseteq Y$ và $Set(v) \supseteq Y$ và do đó $Set(t) = Set(u) \cap Set(v) \supseteq Y$. Vậy $f(t) = 1$, và do đó $t \in T_f$.

Với mỗi hội suy dẫn $F$ 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 3.2***

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

*Ví dụ 3.8*

Công thức logic $A+B$ không tương đương với bất kì công thức suy dẫn nào vì bảng chân lí của $A+B$ không chứa phép gán tri $z = (0,0)$.

Việc xác định giới hạn cho các khái niệm về phụ thuộc dữ liệu cho phép xây dựng các hệ tiên đề cho các lớp phụ thuộc. Về bản chất, một hệ tiên đề cho một lớp phụ thuộc là tập các quy tắc thực hiện các suy dẫn trên các phụ thuộc để nhận được các phụ thuộc mới thuộc cùng lớp. 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

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

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

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

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

*Thuật toán tìm bao đóng tập thuộc tính trong phụ thuộc Boole dương tổng quát*

*Tư 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 $X \models a \in \mathcal{L}^+$ thì bổ sung $a$ vào $X^+$. Với $\mathcal{L}$ đã được đưa về dạng chuẩn không dư thì điều kiện $X \models a \in \mathcal{L}^+$ sẽ cho CNF($\mathcal{L}$ $(X \models a)$') = $\mathcal{L}Xa$', do đó sẽ tương đương với điều kiện Unif($\mathcal{L}Xa$') = $\varnothing$.

**Thuật toán 3.2** (Thuật toán tìm bao đóng trong lớp các phụ thuộc logic)

```
Algorithm  LClosure(X, Ψ)
Input      LNorm(Ψ); X ⊆ U
Output   X+ = {a ∈ U | X⊨ a ∈  Ψ+}
Begin
   Y :=  X; V := U - X;
       for each attribute a in V do
            if  Unif(ΨXa') = ∅ then
                add a to Y;
             endif;
       endfor;
     return Y;
End  LClosure
```

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

*Ví dụ 3.9*

Cho tập thuộc tính $U = abcd$, tập phụ thuộc logic $\Psi = \{b'+c,\ a' + b\}$, $X = ab$, tìm bao đóng $X^+$?

Ta có $\Psi = (b' + c)(a' + b)$ đã ở dạng CNF.

Đặt $Y = X = ab$, $V = U - X = cd$. Xét các thuộc tính trong $V$:

Xét $c \in V$. Ta có:

$\text{Unif}(\Psi Xc') = (\underline{b}' + c)(a' + \underline{b})abc' \Rightarrow (c + \underline{a}')\underline{a}bc' \Rightarrow bcc' \Rightarrow \varnothing$.

Vậy $Y = ab \cup \{c\} = abc$.

Xét $d \in V$. Ta có:

$\text{Unif}(\Psi Xd') = (\underline{b}' + c)(a' + \underline{b})abd' \Rightarrow (c + \underline{a}')\underline{a}bd' \Rightarrow cbd' \neq \varnothing$.

Cuối cùng ta thu được $X^+ = (ab)^+ = abc$.

***Mệnh đề 3.1***

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

***Chứng minh***

Trong thuật toán giải bài toán tìm bao đóng trong lớp các phụ thuộc logic có sử dụng thuật toán hợp giải *Unif* để giải quyết bài toán thành viên. Mà theo mệnh đề trên. thì bài toán thành viên thuộc lớp NPC. Do đó bài toán tìm bao đóng cũng thuộc lớp NPC, đpcm.

## 3.6. Xây dựng thuật toán tìm khóa với phụ thuộc Booe 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*.

Ý nghĩa thực tế của khóa là để nhận diện một bộ trong một quan hệ, giúp cho việc tìm kiếm dữ liệu nhanh và chính xác hơn.

Thuật toán sau thực hiện các bước tìm một khóa trên tập thuộc tính $U$ và tập phụ thuộc logic $F$ 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 $K$.

Do phải bảo toàn bất biến $K^+ = U$ nên điều kiện $(K - a)^+ = U$ tương đương với điều kiện $(K - a) \models a$. Điều kiện này lại tương đương với điều kiện Unif$(F(K - a)a') = \emptyset$.

**Thuật toán 3.3** (Tìm khóa $K$ trong lớp các phụ thuộc logic)

```
Algorithm  LKey(U,  Lnorm(Ψ))
Input     U,  Lnorm(Ψ)
Output  K ⊆ U:
K+ = U
∀a ∈ K: (K – a)+ ≠ U
Begin
           K ≔ U;
           for each attribute a in U
do
             if Unif(Ψ(K – a)a’) = ∅
then
                  delete a from K;
              endif
          endfor
          return K;
End LKey
```

Các thuật toán được kế thừa thuật toán giải bài toán thành viên, thuật toán tìm phủ không dư, thuật toán tìm bao đóng và khóa trong phụ thuộc hàm [46]; nhưng các thuật toán được phát triển để viết các thuật toán giải các bài toán này trong lớp các phụ thuộc logic và được cải tiến như sau:

Tập phụ thuộc hàm $F$ được thay thế bởi tập phụ thuộc logic tổng quát $\mathcal{L}$.

Công thức suy dẫn được chuẩn hóa về dạng CNF trước khi giải bài toán.

Áp dụng phương pháp hợp giải để giải quyết bài toán.

*Ví dụ 3.10*

Cho tập các thuộc tính $U = \{b, c, d, e\}$, tập PTBD $F = \{c'+d, b'+c\}$, hãy cho biết một khóa $K$ trong quan hệ $r$?

Trước hết ta đưa $F$ về dạng CNF, $F = (c'+d)(b'+c)$.

Áp dụng thuật toán tìm khóa, đặt $K = U = bcde$. Lần lượt rút gọn khóa $K$ bằng cách xét từng thuộc tính $b$, $c$, $d$, $e$ trong $K$:

Xét $b$: $K\text{-}b = cde$. Ta có Unif($Fcdeb'$) = $(c'+d)(b'+c)bceb' = (b'+d)cdeb' \neq \emptyset$. Vậy $K$ không thay đổi, $K = bcde$.

Xét c: $K\text{-}c = bde$. Ta có Unif($Fbdec'$) = $(c'+d)(b'+c)bdec' = (c'+d)cdec' = \emptyset$. Vậy $K = bde$.

Xét $d$: $K\text{-}d = be$. Ta có Unif($Fbed'$) = $(c'+d)(c'+d)bed' = (c'+d)ced' = ded' = \emptyset$. Vậy $K = be$.

Xét $e$: $K\text{-}e = b$. Ta có Unif($Fbe'$) = $(c'+d)(b'+c)be' = (c'+d)de' = de' \neq \emptyset$. Vậy $K$ không thay đổi. Cuối cùng ta thu được khóa $k = be$.

***Mệnh đề 3.2***

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

***Chứng minh***

- Trong thuật toán giải bài toán tìm khóa trong lớp các phụ thuộc logic có sử dụng thuật toán hợp giải Unif để giải quyết bài toán thành viên. Mà theo mệnh đề trên thì bài toán thành viên thuộc lớp NPC. Do đó bài toán tìm khóa cũng thuộc lớp NPC, đpcm.

Trường hợp tập phụ thuộc logic đều là phụ thuộc hàm, thì thuật toán tìm một khóa có độ phức tạp đa thức. Tuy nhiên bài toán tìm tất cả các khóa của một lược đồ quan hệ trang bị PTH vẫn thuộc lớp NPC [46].

### 3.7. Kết luận chương 3

Trong chương 3 của luận án, NCS đã tổng kết ba phương pháp tiếp cận bài toán suy dẫn trong các phụ thuộc logic, đó là phương pháp Wang Hạo, phương pháp gián tiếp theo hợp giải và phương pháp trực tiếp. Tuy thuật toán của các phương pháp này vẫn có độ phức tạp là hàm mũ, nhưng đây là những phương pháp tiếp cận suy luận logic hình thức thay vì suy luận theo quan hệ. Do đó, thời gian thực hiện thuật toán không phụ thuộc vào kích thước của dữ liệu đầu vào (không phụ thuộc vào số lượng phần tử trong cơ sở dữ liệu). Điều này làm cho thời gian thực hiện một số bài toán được giảm đáng kể khi dữ liệu vào đủ lớn so với các thuật toán truyền thống.

Nghiên cứu các phụ thuộc logic theo tiếp cận của logic hình thức cho phép ta thiết kế và quản lý các cơ sở dữ liệu và tri thức với những phụ thuộc phức tạp và đa dạng một cách thống nhất. Các kết quả thu được trong chương này như 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 Bool dương tổng quát, thuật toán tìm khóa với phụ thuộc hàm và thuật toán tìm khoá với phụ thuộc Bool dương tổng quát được xây dựng là hướng đi mới giúp người dùng có thể được vận dụng trong lĩnh vực khai thác tri thức từ các tập dữ liệu lớn bằng cách dựa vào các công thức logic có sẵn thay vì phụ thuộc vào kích cỡ của dữ liệu.

# KẾT LUẬN VÀ HƯỚNG PHÁT TRIỂN

## I. Các kết quả đạt được của luận án

Việc 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 của các cá nhân, tổ chức, xí nghiệp…

Luận án đã phát triển và mở rộng lớp các phụ thuộc logic với các tính chất và đặc trưng chung nhất. Cụ thể, 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 nới 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, phương pháp trực tiếp và phương pháp hợp giải để giải bài toán suy dẫn. Đây là cách tiếp cận hình thức trên các biểu thức logic thay vì giải theo dữ liệu.

- 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 nới 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, Nguyễn Thị Vân, Trương Thị Thu Hà (2016), "Tương quan giữa phụ thuộc hàm xấp xỉ và phụ thuộc Bool 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*, 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, Nguyễn Xuân Huy, Trương Thị Thu Hà (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, ISBN: 978-604-67-1009-7, Hà Nội, tr.260-265.

[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*, 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 nới lỏng và phụ thuộc Bool 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.

# TÀI LIỆU THAM KHẢO


[1] Codd E. F., "A Relational Model of Data for Large Shared Data Banks", *CACM 13:6,* pp. 377-387, 1970.

[2] Codd E. F., "Further Normalization of the Database Relational Model", *Database Systems, Courant Comp. Sci., Symp.,* pp. 65-98, 1971.

[3] Armstrong W.W., "Dependency Structure of Data-base Relationship", *Information Processing 74, North Holland, Amsterdam,* pp. 580-583, 1974.

[4] Fagin, R., "Multivalued dependencies and a new normal form for relational databases", *ACM Tans. Database Syst,* p. 262 – 278, 1977.

[5] D. C. 19. Armstrong W.W., "Decomposition and Functional Dependencies in Relations", *ACM Tods 5, 4, Dec.,* pp. 404-430, 1980.

[6] Sagiv Y., Delobel C., Parker D.S., Fagin R., "An equivalence between Relational Database Dependencies and a Fragment of Propositional Logic", *J. ACM,* vol. 28, p. 435 – 453, 1981.

[7] Demetrovics J., and Gyepesi O., "Some generalized type functional dependencies formalized as equality set in matrices", *Discrete App,* p. 35 – 47, 1983.

[8] Maier D., "The Theory of Relational Databases," *Computer Science Press,* 1983.

[9] Beeric C., Dowd M., Fagin R., and Statman R., "On the Structure of Armstrong Relations for Functional Dependencies", *J.ACM,* Vols. 31, No.1, pp. 30-46, 1984.

[10] Dowling, William F., Gallier, Jean H., "Linear - time algorithms for testing the satisfiability of propositional Horn formula," *J. Logic Programing 1,* p. 267 – 284, 1984.

[11] Demetrovics J., Ho Thuan, Nguyen Xuan Huy, "Balanced Relation Schemes and Keys of Relation Schemes (in Russian)", *In book: Cybernetics and Computer Science, Nauka, Moscow,* pp. 296-316, 1987.

[12] Berman J., Blok W. J., "Positive Boolean dependencies", *Inf. Processing Letters,* vol. 27, p. 147 – 150, 1988.

[13] F. W. G. F. J. X. a. K. A. 27. Bohannon P., "Conditional functional dependecie for data cleaning In ICDE", p. p. 745 – 756, 2007.

[14] Demetrovics J., Nguyen Xuan Huy, "Closed Sets and Translations of Relation Schemes", *Computers Math. Applic.,* vol. 21, pp. 13-23, 1991.

[15] Huy Nguyen Xuan, Thanh Le Thi, "Generalized Positive Boolan Dependencies", *J. Inf. Process. Cybern. EIK,* vol. 28, p. 363 – 370, 1992.

[16] Jyrky Kivinen et al, "Approximate Inference of Funtional Dependencies from Relations", *Journal of Theoretical Computer Science,* vol. 149, no. 1, pp. 129-149, 1995.

[17] Hultala Y. et al, "Tane: An efficient algorothm for discovering functional and approximate dependencies", *The Computer Journal,* vol. 42, no. 2, pp. 100-111, 1999.

[18] Franz Baader, Wayne Snyder, "Handbooks of Automated Reasoning, Ed. Alan Robinson and Andrei Voronkov", *Elsevier Science Publishers B.V,* 2001.

[19] Vincent, M. & Liu, J., "Multivalued dependencies in XML", *in Proceedings of the 20th British National Conference on Databases, number 2712 in Lecture Notes in Computer Science, Springer,* pp. 4-8, 2003.

[20] Ronald S. K and Janes J.L, "Discovery of Functional and Approximate Functional Dependencies in Relational Databases", *Journal of Applied Mathematics and Decision Sciences,* vol. 7, pp. 49-59, 2003.

[21] Song S. and Chen L., "Discovering matching dependencies", *rXiv:0903.3317v2[cs.DB],* 2013.

[22] Baixeries J., Kaytoue M., and Napoli A., "Computing similarity dependencies with pattern structures", *In CLA,* pp. 33-44, 2013.

[23] Li W., Li Z., Chen Q., Yin Z., "Discovering Approximate Functional Dependencies from Distributed Big Data", *In Book: Web Technologies and Applications, DOI,* p. 289 – 301, 2016.

[24] Loredana Caruccio, Vincenzo Deufemia, Giuseppe Polese, "Relaxed Functional Dependencies - A Survey of Approaches", *IEEE Transactions on Knowledge & Data Engineering,* vol. 28, p. 147–165, 2016.

[25] Đoàn Văn Ban, Nguyễn Xuân Huy, Đàm Gia Mạnh, Nguyễn Thế Dũng, "Về mối liên hệ giữa suy diễn phụ thuộc hàm và suy diễn logic", *Tạp chí Tin học và điều khiển học,* Vols. T.17, S.4, p. tr.11 – 16, 2001.

[26] Vũ Ngọc Loãn, "Các lớp phụ thuộc logic tổng quát trong mô hình cơ sở dữ liệu quan hệ", *luận án tiến sĩ, LA95.0496.3 - Thư viện Quốc Gia,* 1995.

[27] Bùi Đức Minh, "Nghiên cứu hệ sinh ánh xạ đóng và ứng dụng trong thể hiện ngữ nghĩa dữ liệu", *Luận án tiến sĩ tại Viện Công nghệ thông tin, Viện Hàn lâm Khoa học Việt Nam,* 2014.

[28] Nguyễn Hoàng Sơn, "Một số vấn đề liên quan đến ràng buộc dữ liệu trong cơ sở dữ liệu quan hệ", *luận án tiến sĩ tại ,* Vols. Viện Công nghệ thông tin, Viện Hàn lâm Khoa học Việt Nam, 2006.

[29] Lương Nguyễn Hoàng Hoa , "Phát triển một số phụ thuộc logic trong cơ sở dữ liệu", *luận án tiến sĩ, LA13.0454.3 - Thư viện Quốc Gia.*

[30] Huy Nguyen Xuan, Minh Le Duc, Loan Vu Ngoc, "Some Result Concerning in the Class of Multivalued Positive Boolean Dependencies, in Book: The Mathematial Foundation of Informatics, Eds by Do Long van and M. Ito", *Proceeding of the Conference, World S,* 2005.

[31] Lê Xuân Vinh, "Về một cơ sở đại số và logic cho lập luận xấp xỉ và ứng dụng", *luận án tiến sĩ tại Viện Công nghệ thông tin, viện Hàn lâm Khoa học Việt Nam,* 2008.

[32] Trương Thị Thu Hà, "Tương quan giữa các lớp phụ thuộc logic trong cơ sở dữ liệu", *Luận án tiến sĩ, V-LA2/4042 – Học viện Kỹ thuật quân sự,* 2017.

[33] Nguyễn Thị Kim Anh, "Các phụ thuộc logic trong mô hình dữ liệu quan hệ", *Luận án tiến sĩ, LA93.0273.3 – Thư viện Quốc Gia,* 1993.

[34] Những người dịch: Hồ Thuần, Nguyễn Quang Vinh, Nguyễn Xuân Huy, "Date C. J., Nhập môn Cơ sở dữ liệu", *NXB Thống kê, Hà Nội,* Vols. Tập I, Tập II, 1985.

[35] Fan, W., Geerts, F., Jia, X., and Kementsietsidi, "Conditional functional dependencies for capturing data inconcistencies", *ACM Trans Database Syst,* vol. 33, no. 2, 2008.

[36] Fan, W., Geerts, F., Lakshmanan, L. V. S. and Xiong, M, "Discovering Conditional functional dependencies", *In ICDE,* p. 1231 – 1234, 2009.

[37] Fan W., Li J., Jia X., Ma S., "Reasoning about record matching rules", *PVLDB 2(1),* p. 407 – 418, 2009.

[38] Nguyễn Xuân Huy, "Các phụ thuộc logic trong cơ sở dữ liệu," *NXB Thống Kê,* 2006.

[39] Post E.L., "The two valued Interative Systems of Mathematical logic", *Annals of Math. Studies 5, Princeton University.,* 1991.

[40] Ho Thuan, "Contribution to The Theory of Relational Databases", *Tanulmanyok, Studies 1984/1986,* 1986.

[41] Nguyen Xuan Huy, Le Thi Thanh, "Generalized Positive Boolean Dependencies", *J. Inf. Process. Cybern. EIK,* Vols. 28, 6, 363-370, 1992.

[42] S. S. a. C. L., "Differential Dependencies: Reasoning and Discovery", *ACM Trans. Datab. Syst.,,* Vols. vol.9 , no 4, Article 39., 2011.

[43] Hoa Luong Nguyen Hoang, "Some results concerning Generalized Positive Boolean Dependencies in relational database", *Internatinal Journal of Computer Electrical Engineering (IJCEE),* Vols. 3, No. 6, p. 779 – 783, 2011.

[44] Jalal Atoum, "Mining Approximate Functional Dependencies from Databases", *European Journal of Scientific Research, ISSN 1450-216X,* vol. 33, no. 2, pp. 338-346, 2009.

[45] Song S. and Chen L., "Efficient discovery of similarity constraints for matching dependencies", *Data & Knowledge Engineering,* 2013.

[46] M. V. M. T. K. 61. Maier D., "Challenges for Dataset Search, Database System for Advanced Applications", *Lecture Notes in Computer Science, ,* Vols. Vol. 8421, p.1 – 15., 2014.

[47] Huy Nguyen Xuan, Thanh Le Thi, "Generalized Positive Boolan Dependencies", *J. Inf. Process. Cybern. EIK,* Vols. vol. 28, p. 363 – 370, 1992.

[48] Vũ Đức Thi, "Cơ sở dữ liệu: Kiến thức và thực hành", *NXB Thống kê, Hà Nội,* 1997.

[49] Ilyas, I.F., Markl, V., Haas, P.J., Brown, P., and Aboulnaga, A., "Cords: Automatic discovery of correlations and soft functional dependencies," *In Sigmod Conference,* p. 647 – 658, 2004.

[50] Link, S., "On the Logical Implication of Multivalued Dependencies with Null Values. In Proc. Twelfth Computing: The Australasian Theory Symposium Hobart", *Australia. CRPIT, 51. Gudmundsson, J. and Jay, B., Eds. ACS.,* p. 113 – 122, 2006.

[51] Bohannon P., Fan W., Geerts F., Jia, X. and K.ementsietsidis A., "Conditional functional dependencies for data cleaning", *In ICDE,* p. 746 – 755, 2007.

[52] Koudas, N., Saha, A., Srivastava, D., and Venkatasubramanian, S., "Metric Functional Dependencies", *In ICDE,* p. 1275 – 1278, 2009.

[53] Golab, L., Karloff, H., Korn, F., Saha, A., and Srivastava, D., "Sequential dependencies," *In PVLDB, 2(1),* p. 574 – 585, 2009.

[54] David Maier, V. M Megler, Kristin Tufte, "Challenges for Dataset Search, Database System for Advanced Applications", *Lecture Notes in Computer Science,* vol. 8421, pp. 1-15, 2014.

[55] Baixeries J., Kaytoue M., and Napoli A., "Characterizing functional dependencies in formal concept analysis with pattern structures", *Annals of Mathematics and Articial Intelligence,* pp. 1-21, 2015.

[56] Li J., Todorov V., Tauchen G., "Inference theory for volatility functional dependencies", *Journal of conometrics,* vol. 193, pp. 17-34, 2016.

[57] Panagiotis Mandros, Mario Boley, and Jilles Vreeken, "Discovering Reliable Approximate Functional Dependencies", *In Proceedings of KDD '17, Halifax, NS, Canada,* August 13-17, 2017.

[58] Loredana Caruccio, Vincenzo Deufemia, Giuseppe Polese, "Mining relaxed functional dependencies from data. Data Mining and Knowledge Discovery", *Giovanni Paolo II n.132, Fisciano, SA 84084, USA, Springer Science+Business Media, LLC, part of Springer Na,* 2019.

[59] Loredana Caruccio, Vincenzo Deufemia, Member, IEEE, Felix Naumann, and Giuseppe Polese, Member, IEEE, "Discovering Relaxed Functional Dependencies based on Multi-attribute Dominance", *IEEE Transactions on Knowledge and Data Engineering,* vol. 33, no. 9, 2021.

[60] Bernardo Breve, Loredana Caruccio, Vincenzo Deufemia, Giuseppe Polese, "A Missing Value Imputation Algorithm based on Relaxed Functional Dependencies", *International Conference on Extending Database Technology,* 2022.

[61] Armstrong W.W., "Dependency Structure of Data-base Relationship", *Information Processing 74, North Holland, Amsterdam,,* pp. p. 580-583, 1974..