

**ĐẠI HỌC THÁI NGUYÊN  
TRƯỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN VÀ TRUYỀN THÔNG**  
-----

**PHẠM XUÂN HÀ**

**NGHIÊN CỨU CÁC PHƯƠNG PHÁP TRÍCH CHỌN SỰ KIỆN VÀ ỨNG DỤNG  
VÀO BÀI TOÁN TRÍCH CHỌN SỰ KIỆN DỊCH BỆNH**

**LUẬN VĂN THẠC SĨ KHOA HỌC MÁY TÍNH**

**Thái Nguyên - 2016**

**ĐẠI HỌC THÁI NGUYÊN  
TRƯỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN VÀ TRUYỀN THÔNG**  
-----

**PHẠM XUÂN HÀ**

**NGHIÊN CỨU CÁC PHƯƠNG PHÁP TRÍCH CHỌN SỰ KIỆN VÀ ỨNG DỤNG  
VÀO BÀI TOÁN TRÍCH CHỌN SỰ KIỆN DỊCH BỆNH**

**Chuyên ngành: KHOA HỌC MÁY TÍNH**

**Mã số: 60 48 01 01**

**LUẬN VĂN THẠC SĨ KHOA HỌC MÁY TÍNH**

**NGƯỜI HƯỚNG DẪN KHOA HỌC**

**GS. TS VŨ ĐỨC THI**

**Thái Nguyên - 2016**

## **LỜI CAM ĐOAN**

Tôi xin cam đoan luận văn này là công trình nghiên cứu thực sự của cá nhân mình, thực hiện dưới sự hướng dẫn tận tình của thầy giáo TS. Lê Văn Phùng.

Các số liệu, kết quả do bản thân nghiên cứu và tìm hiểu được trình bày trong luận văn này trung thực và chưa từng được công bố dưới bất cứ hình thức nào.

Tôi xin chịu hoàn toàn trách nhiệm về nghiên cứu của mình.

**Học viên**

**Lê Long Giang**

## LỜI CẢM ƠN

Lời đầu tiên, tôi xin được gửi lời cảm ơn sâu sắc nhất tới thầy giáo TS. Lê Văn Phùng, người thầy đã trực tiếp dành nhiều thời gian tận tình hướng dẫn, cung cấp những thông tin, tài liệu quý báu giúp đỡ tôi hoàn thành bản luận văn này.

Tôi cũng xin gửi lời cảm ơn đến các giảng viên trường Đại Học Công nghệ Thông Tin và Truyền Thông - Đại học Thái Nguyên, các thầy Viện Công nghệ thông tin - Viện Hàn lâm Khoa học Công nghệ Việt Nam đã giảng dạy, truyền đạt những kiến thức và giúp đỡ tôi trong suốt quá trình học tập của mình.

Tôi cũng xin gửi lời cảm ơn tới Ban Giám đốc - Sở Khoa học và Công nghệ tỉnh Lào Cai, Phòng Quản lý Công nghệ và Thị trường công nghệ đã tạo mọi điều kiện thuận lợi cho em tham gia khóa học và trong suốt quá trình hoàn thành luận văn.

Cuối cùng, tôi xin cảm ơn những người thân, bạn bè và gia đình đã luôn cổ vũ động viên tôi hoàn thành luận văn tốt nghiệp này.

Mặc dù đã hết sức cố gắng hoàn thành luận văn với tất cả sự nỗ lực của bản thân, nhưng luận văn không tránh khỏi những thiếu sót. Kính mong nhận được những ý kiến đóng góp của quý thầy cô và bạn bè, đồng nghiệp.

Tôi xin chân thành cảm ơn!

*Thái Nguyên, ngày 20 tháng 04 năm 2016*

**Học viên**

**Lê Long Giang**

## MỤC LỤC

LỜI CAM ĐOAN .....	i
LỜI CẢM ƠN .....	iv
MỤC LỤC .....	v
CÁC KÍ HIỆU VIẾT TẮT DÙNG TRONG LUẬN VĂN.....	viii
MỘT SỐ QUI ƯỚC VỀ KÍ HIỆU THƯỜNG ĐƯỢC .....	ix
SỬ DỤNG TRONG LUẬN VĂN: .....	ix
BẢNG CÁC HÌNH VẼ .....	x
DANH MỤC CÁC ĐỊNH NGHĨA, ĐỊNH LÝ, BỔ ĐỀ, THUẬT TOÁN .....	x
MỞ ĐẦU .....	1
CHƯƠNG 1: PHỤ THUỘC HÀM MẠNH VÀ PHẦN TỬ NGOẠI LAI.....	3
1.1. Khái niệm về phụ thuộc mạnh.....	3
1.2. Phương pháp xác định phụ thuộc mạnh trong CSDL.....	5
1.3. Phần tử ngoại lai và mối quan hệ giữa chúng với khai phá dữ liệu .....	13
1.3.1. Khái niệm về phần tử ngoại lai .....	13
1.3.2. Các phương pháp xác định phần tử ngoại lai.....	14
1.3.3. Mối quan hệ giữa phần tử ngoại lai và khai phá dữ liệu.....	15
1.4. Mô hình phát hiện phần tử ngoại lai trong dữ liệu và trong cơ sở dữ liệu quan hệ.....	16
1.4.1. Định nghĩa mô tả.....	17
1.4.2. Phân loại các phần tử ngoại lai trong CSDL quan hệ .....	18
1.4.3. Mô hình phát hiện phần tử ngoại lai dựa theo luật đối với CSDL quan hệ.....	18
1.5. Ứng dụng của các phần tử ngoại lai .....	22
KẾT LUẬN CHƯƠNG 1.....	23
CHƯƠNG 2: PHÁT HIỆN PHẦN TỬ NGOẠI LAI ĐỐI VỚI PHỤ THUỘC HÀM MẠNH TRONG CƠ SỞ DỮ LIỆU QUAN HỆ.....	24
2.1. Phần tử ngoại lai đối với các dạng phụ thuộc hàm đặc biệt .....	24
2.1.1. Phần tử ngoại lai đối với phụ thuộc hàm dạng bằng nhau .....	24

2.1.2. Phần tử ngoại lai đối với phụ thuộc hàm dạng tỉ lệ.....	26
2.2. Phần tử ngoại lai đối với hệ ràng buộc dạng phụ thuộc hàm .....	28
2.3. Thuật toán phát hiện các phần tử ngoại lai đối với các dạng chuẩn.....	32
2.3.1. Thuật toán phát hiện phần tử ngoại lai đối với dạng chuẩn 2NF....	33
2.3.2. Thuật toán phát hiện phần tử ngoại lai đối với dạng chuẩn 3NF....	34
2.3.3. Thuật toán phát hiện phần tử ngoại lai đối với dạng chuẩn BCNF.	36
2.4. Phần tử ngoại lai đối với phụ thuộc mạnh.....	37
2.4.1. Thuật toán tìm các phụ thuộc hàm mạnh trong quan hệ và SDQH	38
2.4.2. Xác định phụ thuộc hàm mạnh cực đại đối với một tập thuộc tính	39
2.4.3. Thuật toán tìm phần tử ngoại lai đối với phụ thuộc mạnh trong CSDL quan hệ .....	42
KẾT LUẬN CHƯƠNG 2.....	43
CHƯƠNG 3: ỨNG DỤNG TÌM PHẦN TỬ NGOẠI LAI .....	45
3.1. Lựa chọn bài toán để cài đặt.....	45
3.2. Cài đặt chương trình .....	48
3.2.1. Yêu cầu hệ thống.....	48
3.2.2. Cấu trúc của chương trình .....	48
3.3. Chương trình minh họa:.....	48
3.3.1. Demo 01: .....	48
3.3.2. Demo 02: .....	52
3.3.3. Một số đoạn mã lệnh sử dụng trong chương trình .....	56
KẾT LUẬN VÀ HƯỚNG PHÁT TRIỂN .....	65
TÀI LIỆU THAM KHẢO.....	66

**DANH MỤC HÌNH ẢNH**

- Hình 1.3.1 Phần tử ngoại lai trong tập điểm có tọa độ  $(x,y)$  trên mặt phẳng có giá trị tung độ  $y$  nhỏ hơn hẳn các phần tử khác của tập hợp..... 13
- Hình 1.4.3 Sơ đồ phát hiện phần tử ngoại lai dựa theo luật trong CSDL quan hệ...21

**CÁC KÍ HIỆU VIẾT TẮT DÙNG TRONG LUẬN VĂN****Kí hiệu****Nghĩa**

CNTT

Công nghệ thông tin

CSDL

Cơ sở dữ liệu

PTH

Phụ thuộc hàm

PTM

Phụ thuộc mạnh

SDQH

Sơ đồ quan hệ

SDM

Sơ đồ mạnh



## MỘT SỐ QUI ƯỚC VỀ KÍ HIỆU THƯỜNG ĐƯỢC

### SỬ DỤNG TRONG LUẬN VĂN:

- Các thuộc tính được kí hiệu bằng các chữ Latin hoa đầu bảng chữ A, B, C,...
- Tập thuộc tính được ký hiệu bằng các chữ Latin hoa cuối bảng chữ X, Y, Z,...
- $XY$  hoặc  $X \cup Y$  biểu diễn hợp của hai tập  $X$  và  $Y$ . Phép trừ hai tập  $X$  và  $Y$  được ký hiệu là  $X \setminus Y$ , hoặc  $X - Y$ .
- Một phân hoạch của tập  $M$  (thành các tập con rời nhau và có hợp là  $M$ ),  $X_1, X_2, \dots, X_m$  được ký hiệu là  $M = X_1 | X_2 | \dots | X_m$   
 Với ý nghĩa  $M = X_1 \cup X_2 \cup \dots \cup X_m$  và  $X_i \cap X_j = \emptyset, 1 \leq i, j \leq m, i \neq j$ .
- Kí hiệu  $R, U \dots$  để chỉ tập toàn bộ các thuộc tính trong một sơ đồ quan hệ
- Các quan hệ (hoặc bảng dữ liệu) được kí hiệu bằng các chữ cái thường:  $r, p, q, \dots$
- Các bộ được biểu diễn bằng các chữ Latin thường có thể kèm chỉ số  $t, u, v, t_1, \dots$
- Với mỗi bộ  $t$  trong quan hệ  $r$  và mỗi tập con các thuộc tính  $X \subseteq R$  ta kí hiệu  $t[X]$  hoặc  $t.X$  là hạn chế của bộ (ánh xạ)  $t$  trên tập thuộc tính  $X$ .
- Kí hiệu  $|r|$  là lực lượng (số bộ) của quan hệ  $r$ .
- Kí hiệu  $X \rightarrow Y$  để chỉ phụ thuộc hàm giữa  $X$  và  $Y$
- Kí hiệu  $X \xrightarrow{s} Y$  để chỉ phụ thuộc mạnh giữa  $X$  và  $Y$ ; hoặc có thể sử dụng kí hiệu  $X \rightarrow Y$  để chỉ phụ thuộc mạnh với lời chú dẫn đi trước.
- Kí hiệu  $X \rightarrow \sigma Y$  để chỉ phụ thuộc hàm xấp xỉ mức  $\sigma$  giữa  $X$  và  $Y$
- Kí hiệu  $\Rightarrow$  để chỉ sự kéo theo trong mệnh đề logic

## BẢNG CÁC HÌNH VẼ

### DANH MỤC CÁC ĐỊNH NGHĨA, ĐỊNH LÝ, BỔ ĐỀ, THUẬT TOÁN

Định nghĩa 1.4.1 Định nghĩa mô tả.....	17
Định nghĩa 2.1.1.1 Phụ thuộc hàm dạng bằng nhau .....	24
Định nghĩa 2.1.1.2 Phần tử ngoại lai đối với phụ thuộc hàm dạng bằng nhau.....	24
Định nghĩa 2.1.2.1 Phụ thuộc hàm dạng tỉ lệ.....	26
Định nghĩa 2.1.2.2 Phần tử ngoại lai đối với phụ thuộc hàm dạng tỉ lệ .....	27
Định nghĩa 2.1.3.1 Phần tử ngoại lai đối với hệ ràng buộc dạng PTH.....	29
Định nghĩa 2.2.1 Phần tử ngoại lai đối với dạng chuẩn.....	32
Định nghĩa 2.2.3 Phụ thuộc mạnh cực đại .....	39
Định lý 1.2.1 Tính đúng và đầy đủ của hệ T1-T3.....	6
Định lý 1.2.2 Sự tồn tại họ phụ thuộc hàm sinh ra họ phụ thuộc mạnh.....	9
Định lý 1.2.3 Họ phụ thuộc mạnh.....	12
Bổ đề 1.2.1 Tính bắc cầu hỗn hợp .....	8
Bổ đề 1.2.2: .....	11
Bổ đề 2.1.3.1 .....	29
Mệnh đề 1.2.2 Phụ thuộc mạnh của các tập phụ thuộc hàm tương đương. ....	11
Mệnh đề 2.2.3 Sự tồn tại của phụ thuộc mạnh.....	39
Thuật toán 2.1.1 (phát hiện phần tử ngoại lai đối với PTH dạng bằng nhau) .....	25
Thuật toán 2.1.2 (phát hiện phần tử ngoại lai đối với phụ thuộc hàm dạng tỉ lệ).....	27
Thuật toán 2.1.3 (Thuật toán xác định phần tử ngoại lai đối với hệ ràng buộc dạng PTH) .	31
Thuật toán 2.2.1.1 (Thuật toán NL_ 2NF).....	33
Thuật toán 2.2.1.2 (Thuật toán NL_ 3NF).....	34
Thuật toán 2.2.1.3 (Thuật toán NL_ BCNF) .....	36
Thuật toán 2.2.2 (Tìm các PTM trong SDQH).....	38
Thuật toán 2.2.3.1 (Tìm phụ thuộc mạnh cực đại).....	40
Thuật toán 2.3.1.2 (Tìm các phụ thuộc mạnh trong quan hệ r) .....	40