

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



**CHỦ THỊ QUỲNH HOA**

**NGHIÊN CỨU MỐI QUAN HỆ GIỮA PHỤ THUỘC  
HÀM VÀ BẢNG QUYẾT ĐỊNH TRONG CHÂN  
ĐOÁN BỆNH**

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

**Thái Nguyên - 2015**

## LỜI CẢM ƠN

Em xin chân thành bày tỏ lòng biết ơn sâu sắc đến TS. Lê Văn Phùng, người thầy đã hết lòng giúp đỡ và tạo điều kiện tốt nhất để em hoàn thành luận văn này.

Em xin chân thành cảm ơn toàn thể các Thầy, cô công tác tại Viện CNTT và Trường Đại học Công Nghệ Thông Tin và Truyền Thông - Đại Học Thái Nguyên đã luôn tận tình chỉ bảo, giúp đỡ, tạo điều kiện và truyền đạt kiến thức cho em trong suốt quá trình học tập và thực hiện luận văn.

Xin chân thành cảm ơn Ban Giám Hiệu Trường THPT Định Hóa, gia đình, bạn bè và đồng nghiệp đã không ngừng quan tâm, động viên, giúp đỡ và tạo mọi điều kiện tốt nhất cho tôi trong suốt thời gian học tập và thực hiện luận văn.

Mặc dù đã cố gắng rất nhiều, nhưng do thời gian có hạn và bản thân còn những hạn chế nhất định nên luận văn không tránh khỏi thiếu sót. Em rất mong nhận được các ý kiến phê bình, góp ý của Hội đồng bảo vệ luận văn, các thầy cô giáo và đồng nghiệp để luận văn được hoàn chỉnh hơn.

Thái Nguyên, ngày 09 tháng 05 năm 2015

Học viên

Chữ Thị Quỳnh Hoa

## LỜI CAM ĐOAN

Tôi xin cam đoan luận văn này là sản phẩm do tôi tổng hợp và nghiên cứu. Trong luận văn có sử dụng một số định nghĩa, bổ đề, hệ quả, thuật toán lấy từ nguồn tài liệu có trích dẫn tên tài liệu và tên tác giả rõ ràng như đã nêu trong phần tài liệu tham khảo. Tôi xin chịu trách nhiệm về sản phẩm nghiên cứu của mình.

Học viên

Chữ Thị Quỳnh Hoa

# MỤC LỤC

Lời cam đoan	
MỤC LỤC .....	i
Danh mục các thuật ngữ .....	iv
Danh mục các ký hiệu, các từ viết tắt .....	v
Danh sách bảng .....	vii
Danh sách hình vẽ .....	viii
MỞ ĐẦU .....	1
<b>Chương 1. TỔNG QUAN VỀ KHAI PHÁ DỮ LIỆU VÀ BÀI TOÁN CHẨN ĐOÁN BỆNH.....</b>	<b>3</b>
<b>1.1. Các khái niệm cơ bản về hệ thống tin và tập thô .....</b>	<b>3</b>
1.1.1. Hệ thống tin.....	3
1.1.2. Tập thô.....	5
1.1.3. Bảng quyết định .....	8
1.1.4. Cơ sở dữ liệu quan hệ .....	10
1.1.5. Tập rút gọn và lõi .....	12
1.1.6. Luật quyết định .....	14
<b>1.2. Khai phá dữ liệu .....</b>	<b>15</b>
1.2.1. Khai phá dữ liệu và phát hiện tri thức .....	15
1.2.2. Các chức năng khai phá dữ liệu .....	17
1.2.3. Các hệ thống khai phá dữ liệu và kiến trúc tổng quát của chúng..	18
1.2.4. Các loại dữ liệu có thể khai phá được.....	22
1.2.5. Khai phá dữ liệu theo hướng tiếp cận tập thô .....	22
<b>1.3. Bài toán chẩn đoán bệnh .....</b>	<b>23</b>
1.3.1. Vai trò của phương pháp chẩn đoán lâm sàng trong y học .....	23
1.3.2. Giá trị thực tiễn của việc xác định quan hệ giữa phụ thuộc hàm và bảng quyết định trong chẩn đoán bệnh .....	24

<b>1.4. Kết luận chương 1 .....</b>	<b>26</b>
<b>Chương 2. MỐI QUAN HỆ GIỮA PHỤ THUỘC HÀM VÀ BẢNG QUYẾT ĐỊNH THEO HƯỚNG TIẾP CẬN TẬP THÔ .....</b>	<b>27</b>
<b>2.1. Xác định các phụ thuộc hàm từ bảng quyết định .....</b>	<b>27</b>
2.1.1. Thuộc tính rút gọn và các thuật toán tìm chúng .....	27
2.1.2. Thuật toán tìm họ tất cả các tập rút gọn trong bảng quyết định ..	36
2.1.3. Tập lõi trong bảng quyết định .....	38
2.1.4. Thuật toán xác định các phụ thuộc hàm từ bảng quyết định ..	42
<b>2.2. Xây dựng bảng quyết định từ tập phụ thuộc hàm .....</b>	<b>44</b>
2.2.1. Sự phụ thuộc giữa các thuộc tính trong bảng quyết định .....	44
2.2.2. Một số thuật toán cơ bản liên quan đến phụ thuộc hàm .....	45
2.2.3. Thuật toán xây dựng bảng quyết định từ tập phụ thuộc hàm ... ..	50
<b>2.3. Kết luận chương 2 .....</b>	<b>54</b>
<b>Chương 3. CHƯƠNG TRÌNH THỬ NGHIỆM .....</b>	<b>55</b>
<b>3.1. Bài toán chẩn đoán bệnh cúm.....</b>	<b>55</b>
3.1.1. Yêu cầu nghiệp vụ .....	55
3.1.2. Yêu cầu công nghệ .....	57
<b>3.2. Thiết kế chương trình .....</b>	<b>57</b>
3.2.1. Thiết kế cơ sở dữ liệu .....	57
3.2.2. Thiết kế xử lý .....	59
3.2.3. Thiết kế các mô-đun và lược đồ chương trình .....	60
<b>3.3. Cài đặt và thực hiện chương trình .....</b>	<b>63</b>
3.3.1. Các chức năng chương trình .....	63
3.3.2. Hệ thống dữ liệu đầu vào .....	64
3.3.3. Hệ thống giao diện .....	65
3.3.4. Kết quả thử nghiệm chương trình và đánh giá .....	67
<b>3.4. Kết luận chương 3 .....</b>	<b>71</b>

KẾT LUẬN .....	72
Tài liệu tham khảo .....	73
PHỤ LỤC .....	75

## Danh mục các thuật ngữ

Thuật ngữ tiếng Việt	Thuật ngữ tiếng Anh
<i>Tập thô</i>	<i>Rough Set</i>
<i>Hệ thống tin</i>	<i>Information System</i>
<i>Bảng quyết định</i>	<i>Decision Table</i>
<i>Quan hệ không phân biệt được</i>	<i>Indiscernibility Relation</i>
<i>Quan hệ dung sai</i>	<i>Tolerance Relation</i>
<i>Xấp xỉ dưới</i>	<i>Lower Approximation</i>
<i>Xấp xỉ trên</i>	<i>Upper Approximation</i>
<i>Rút gọn thuộc tính</i>	<i>Attribute Reduction</i>
<i>Tập rút gọn</i>	<i>Reduct</i>
<i>Tập lõi</i>	<i>Core</i>
<i>Ma trận phân biệt</i>	<i>Indiscernibility Matrix</i>
<i>Hàm phân biệt</i>	<i>Indiscernibility Function</i>
<i>Luật quyết định</i>	<i>Decision Rule</i>
<i>Quan hệ</i>	<i>Relation</i>
<i>Sơ đồ quan hệ</i>	<i>Relation Schema</i>
<i>Phụ thuộc hàm</i>	<i>Functional Dependency</i>
<i>Khóa, phản khóa</i>	<i>Key, Antikey</i>
<i>Tập tối thiểu của thuộc tính <math>a</math></i>	<i>Minimal set of the attribute <math>a</math></i>
<i>Họ các tập tối thiểu của thuộc tính <math>a</math></i>	<i>Family of all minimal sets of attribute <math>a</math></i>

## Danh mục các ký hiệu, các từ viết tắt

Ký hiệu, từ viết tắt	Diễn giải
$S = (U, A)$	Hệ thông tin
$T = (U, C \cup D)$	Bảng quyết định
$ U $	Số đối tượng
$ C $	Số thuộc tính điều kiện trong bảng quyết định
$ A $	Số thuộc tính trong hệ thông tin
$u(a)$	Giá trị của đối tượng $u$ tại thuộc tính $a$
$IND(B)$	Quan hệ $B$ – không phân biệt
$[u]_B$	Lớp tương đương chứa $u$ của quan hệ $IND(B)$
$U/B$	Phân hoạch của $U$ sinh bởi tập thuộc tính $B$ .
$\partial_B(u)$	Hàm quyết định suy rộng của đối tượng $u$ đối với $B$ .
$\underline{BX}$	$B$ – xấp xỉ dưới của $X$
$\overline{BX}$	$B$ – xấp xỉ trên của $X$
$BN_B(X)$	$B$ - miền biên của $X$
$POS_B(D)$	$B$ – miền dương của $D$
$PRED(C)$	Họ tất cả các tập rút gọn Pawlak
$PCORE(C)$	Tập lõi dựa trên miền dương
$HCORE(C)$	Tập lõi dựa trên entropy Shannon có điều kiện
$SCORE(C)$	Tập lõi dựa trên ma trận phân biệt
$ECORE(C)$	Tập lõi dựa trên entropy Liang có điều kiện
$MCORE(C)$	Tập lõi dựa trên metric



CSDL	Cơ sở dữ liệu
<i>SDQH</i>	Sơ đồ quan hệ
<i>PTH</i>	Phụ thuộc hàm

## Danh sách bảng

<b>Bảng 1.1.</b> Bảng dữ liệu về bệnh cúm .....	4
<b>Bảng 1.2.</b> Bảng thông tin về bệnh cúm .....	7
<b>Bảng 1.3.</b> Bảng quyết định .....	9
<b>Bảng 1.4.</b> Bảng quyết định về bệnh cúm .....	14
<b>Bảng 2.1.</b> Bảng quyết định ở Ví dụ 2.1 .....	30
<b>Bảng 2.2.</b> Bảng quyết định về bệnh cúm ở Ví dụ 2.2 .....	32
<b>Bảng 2.3.</b> Bảng rút gọn thứ nhất của hệ thống bệnh cúm $R_1$ .....	33
<b>Bảng 2.4.</b> Bảng rút gọn thứ hai của hệ thống bệnh cúm $R_2$ .....	33
<b>Bảng 2.5.</b> Bảng quyết định ở Ví dụ 2.3 .....	37
<b>Bảng 2.6.</b> Bảng quyết định minh họa Ví dụ 2.4 .....	41
<b>Bảng 2.7.</b> Bảng quyết định được xây dựng từ Thuật toán 2.16 .....	54