

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

**PHẠM VĂN DƯƠNG**

*Nghiên cứu một số phương pháp rút gọn  
thuộc tính trên bảng quyết định không  
đầy đủ và ứng dụng*

LUẬN VĂN THẠC SĨ KHOA HỌC MÁY TÍNH

Thái Nguyên - 2015

## LỜI CẢM ƠN

Em xin chân thành cảm ơn và biết ơn sâu sắc đến GS.TS Vũ Đức Thi, Viện Công nghệ thông tin – Đại học Quốc gia Hà Nội. Người đã tận tình hướng dẫn và giúp đỡ em hoàn thành luận văn này.

Em xin chân thành cảm ơn các Thầy ở Viện Công nghệ thông tin đã dạy bảo, giúp đỡ và truyền đạt kiến thức cho em trong suốt khóa học và quá trình em làm luận văn.

Em xin chân thành cảm ơn các Thầy, các Cô ở trường Đại học Công nghệ thông tin và truyền thông Thái Nguyên đã tận tình dạy bảo, động viên, giúp đỡ và tạo điều kiện cho em trong suốt thời gian học tập và nghiên cứu.

Cuối cùng xin chân thành cảm ơn bạn bè, người thân và gia đình luôn là người đồng hành, động viên, chia sẻ những khó khăn trong suốt thời gian hoàn thành luận văn.

**Học viên**

**Phạm Văn Dương**

## LỜI CAM ĐOAN

Tôi xin cam đoan đề tài "*Nghiên cứu một số phương pháp rút gọn thuộc tính trên bảng quyết định không đầy đủ và ứng dụng*" là công trình nghiên cứu được tôi thực hiện dưới sự hướng dẫn của giáo viên hướng dẫn khoa học.

Một số Định nghĩa, Định lý, Tính chất, Mệnh đề và Thuật toán tôi lấy từ nguồn tài liệu chính xác có trích dẫn tên tài liệu và tên tác giả rõ ràng. Tôi xin chịu trách nhiệm về luận văn của mình.

**Học viên**

**Phạm Văn Dương**

## MỤC LỤC

<i>LỜI CẢM ƠN</i> .....	<i>i</i>
<i>LỜI CAM ĐOAN</i> .....	<i>iii</i>
<i>MỤC LỤC</i> .....	<i>iv</i>
<i>DANH MỤC CÁC THUẬT NGỮ</i> .....	<i>vi</i>
<i>BẢNG CÁC KÝ HIỆU, TỪ VIẾT TẮT</i> .....	<i>vii</i>
<i>DANH SÁCH BẢNG</i> .....	<i>ix</i>
<i>DANH SÁCH HÌNH</i> .....	<i>x</i>
<i>MỞ ĐẦU</i> .....	<i>1</i>
<b><i>Chương 1: NHỮNG KHÁI NIỆM CƠ BẢN VỀ CÁC PHƯƠNG PHÁP RÚT GỌN TRONG BẢNG QUYẾT ĐỊNH</i></b> .....	<b><i>3</i></b>
<b>1.1. Hệ thống tin đầy đủ và mô hình tập thô truyền thống.</b> .....	<b>3</b>
1.1.1. Hệ thống tin đầy đủ.....	3
1.1.2. Mô hình tập thô truyền thống.....	4
1.1.3. Bảng quyết định đầy đủ. ....	7
1.1.4. Tập rút gọn và tập lõi.....	7
1.1.5. Ma trận phân biệt .....	9
<b>1.2. Hệ thống tin không đầy đủ và mô hình tập thô dung sai.</b> .....	<b>10</b>
1.2.1. Hệ thống tin không đầy đủ.....	10
1.2.2. Mô hình tập thô dung sai .....	14
1.2.3. Bảng quyết định không đầy đủ .....	17
<b>1.3. Tình hình nghiên cứu các phương pháp rút gọn thuộc tính</b> .....	<b>18</b>
1.3.1. Các phương pháp rút gọn đã được nghiên cứu .....	18
1.3.2. Mối liên hệ giữa các khái niệm tập rút gọn.....	19
1.3.3. Phân nhóm và mối liên hệ giữa các phương pháp .....	21
1.3.4. Luật quyết định của bảng quyết định không đầy đủ và các độ đo cổ điển .....	22
1.3.5. Các độ đo đánh giá hiệu năng tập luật và các tính chất .....	25

1.3.6. Sự thay đổi giá trị các độ đo đánh giá hiệu năng tập luật quyết định trên các tập rút gọn .....	27
<b>Chương 2. PHƯƠNG PHÁP RÚT GỌN TRONG BẢNG QUYẾT ĐỊNH KHÔNG ĐẦY ĐỦ .....</b>	<b>31</b>
<b>2.1. Mở đầu .....</b>	<b>31</b>
<b>2.2. Entropy Liang mở rộng trong hệ thông tin không đầy đủ và các tính chất</b>	<b>31</b>
2.2.1. Entropy Liang mở rộng của tập thuộc tính .....	32
2.2.2. Entropy Liang mở rộng có điều kiện .....	33
2.2.3. Một số tính chất của entropy Liang mở rộng .....	34
<b>2.3. Rút gọn thuộc tính trong hệ quyết định không đầy đủ sử dụng ma trận phân biệt và hàm phân biệt mở rộng .....</b>	<b>37</b>
<b>2.4. Thuật toán rút gọn thuộc tính .....</b>	<b>42</b>
<b>Chương 3: XÂY DỰNG CHƯƠNG TRÌNH THỰC NGHIỆM.....</b>	<b>45</b>
<b>3.1 Cấu trúc lớp chương trình.....</b>	<b>45</b>
<b>3.2. Thiết kế phần mềm thực nghiệm .....</b>	<b>47</b>
3.2.1. Yêu cầu hệ thống .....	47
3.2.2. Dữ liệu thử nghiệm .....	47
3.2.3. Chuẩn bị dữ liệu.....	47
3.2.4. Một số giao diện chương trình .....	49
3.2.5. Kết quả thực nghiệm.....	51
<b>KẾT LUẬN.....</b>	<b>52</b>
<b>TÀI LIỆU THAM KHẢO.....</b>	<b>53</b>
<b>PHỤ LỤC.....</b>	<i>Error! Bookmark not defined.</i>

## 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>Hệ thống tin đầy đủ</i>	<i>Complete Information System</i>
<i>Hệ thống tin không đầy đủ</i>	<i>Incomplete Information System</i>
<i>Bảng quyết định</i>	<i>Decision Table</i>
<i>Bảng quyết định đầy đủ</i>	<i>Complete Decision Table</i>
<i>Bảng quyết định không đầy đủ</i>	<i>Incomplete Decision Table</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>Quan hệ</i>	<i>Relation</i>
<i>Tập tối thiểu của thuộc tính a</i>	<i>Minimal set of the attribute a</i>
<i>Họ các tập tối thiểu của thuộc tính a</i>	<i>Family of all minimal sets of attribute a</i>

## BẢNG CÁC KÝ HIỆU, TỪ VIẾT TẮT

<b>Ký hiệu, từ viết tắt</b>	<b>Diễn giải</b>
$IS = U, A, V, f$	Hệ thông tin, hệ thông tin đầy đủ
$IIS = U, A, V, f$	Hệ thông tin không đầy đủ
$DS = U, C \cup D, V, f$	Bảng quyết định, bảng quyết định đầy đủ
$IDS = U, C \cup D, V, f$	Bảng quyết định không đầy đủ
$ 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
$SIM \ B$	Quan hệ dung sai trên tập thuộc tính $B$
$u \ B$	Lớp tương đương chứa $u$ của quan hệ $IND \ B$
$S_B \ u$	Lớp dung sai của đối tượng $u$ trên quan hệ $SIM \ B$
$U / B$	Phân hoạch của $U$ sinh bởi tập thuộc tính $B$ .
$U / SIM \ B$	Phủ của $U$ sinh bởi tập thuộc tính $B$ .
$COVER \ U$	Họ tất cả các phủ của $U$ .
$\partial_B(u)$	Hàm quyết định suy rộng của đối tượng $u$ đối với $B$ .
$\underline{B}X$	$B$ – xấp xỉ dưới của $X$
$\overline{B}X$	$B$ – 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
$SRED \ C$	Họ tất cả các tập rút gọn sử dụng ma trận phân biệt
$ERED \ C$	Họ tất cả các tập rút gọn Entropy Liang
$MRED \ C$	Họ tất cả các tập rút gọn dựa trên metric
$PCORE \ C$	Tập lõi dựa trên miền dương
$SCORE \ C$	Tập lõi sử dụng ma trận phân biệt
$ECORE \ C$	Tập lõi dựa trên entropy Liang có điều kiện

$MCORE C$	Tập lỗi dựa trên metric
$E P$	Entropy Liang của tập thuộc tính $P$
$E(Q P)$	Entropy Liang có điều kiện của $Q$ khi đã biết $P$
$IE P$	Entropy Liang mở rộng của tập thuộc tính $P$ trong hệ thông tin không đầy đủ
$IE(Q P)$	Entropy Liang mở rộng có điều kiện của $Q$ khi đã biết $P$ trong hệ thông tin không đầy đủ.
$K P$	Trong hệ thông tin đầy đủ, ký hiệu $K P$ là tri thức sinh bởi tập thuộc tính $P$ . Trong hệ thông tin không đầy đủ, ký hiệu $K P$ là phủ sinh bởi tập thuộc tính $P$ .
$d_j K P , K Q$	Khoảng cách giữa $K P$ và $K Q$ trong hệ thông tin đầy đủ dựa trên khoảng cách Jaccard giữa hai tập hợp.
$d_E K P , K Q$	Khoảng cách giữa $K P$ và $K Q$ trong hệ thông tin không đầy đủ dựa trên entropy Liang mở rộng



**DANH SÁCH BẢNG**

<b>Bảng 1.1.</b> Bảng thông tin về bệnh cúm .....	5
<b>Bảng 1.2.</b> Bảng quyết định về bệnh cúm.....	9
<b>Bảng 1.3.</b> Bảng thông tin về các xe hơi.....	13
<b>Bảng 1.4.</b> Bảng quyết định về các xe hơi .....	18
<b>Bảng 1.5.</b> Ký hiệu các tập rút gọn trong bảng quyết định không đầy đủ. ....	19
<b>Bảng 2.1.</b> Hệ thông tin không đầy đủ về các xe hơi.....	35
<b>Bảng 2.2.</b> Bảng quyết định không đầy đủ về tivi .....	38

**DANH SÁCH HÌNH**

<b>Hình 1.1.</b> Mối liên hệ giữa các tập rút gọn của bảng quyết định không đầy đủ.....	20
<b>Hình 3.1:</b> Giao diện lớp MaxtrixDiscern.....	45
<b>Hình 3.2:</b> Giao diện lớp SqlExecute.....	45
<b>Hình 3.3:</b> Giao diện lớp ImportData .....	46
<b>Hình 3.4:</b> Giao diện lớp MainForm.....	46
<b>Hình 3.5.</b> Dữ liệu adult-stretch gốc .....	48
<b>Hình 3.6.</b> Dữ liệu adult-stretch sau khi chuyển đổi.....	49
<b>Hình 3.7.</b> Giao diện chọn tệp dữ liệu .....	50
<b>Hình 3.8.</b> Kết quả thử nghiệm với bộ dữ liệu adult-stretch.....	50
<b>Hình 3.9.</b> Lưu kết quả rút gọn thành dạng tệp.....	51