

ĐẠI HỌC THÁI NGUYÊN
TRƯỜNG ĐẠI HỌC CNTT VÀ TRUYỀN THÔNG

HOÀNG THỊ NGỌC MAI

MỘT SỐ PHƯƠNG PHÁP RÚT GỌN THUỘC TÍNH
TRONG BẢNG QUYẾT ĐỊNH

LUẬN VĂN THẠC SĨ CÔNG NGHỆ THÔNG TIN

Thái Nguyên - Năm 2013

**ĐẠI HỌC THÁI NGUYÊN
TRƯỜNG ĐẠI HỌC CNTT VÀ TRUYỀN THÔNG**

HOÀNG THỊ NGỌC MAI

**MỘT SỐ PHƯƠNG PHÁP RÚT GỌN THUỘC TÍNH
TRONG BẢNG QUYẾT ĐỊNH**

LUẬN VĂN THẠC SĨ CÔNG NGHỆ THÔNG TIN

**Chuyên ngành: Khoa học máy tính
Mã số: 60.48.01**

NGƯỜI HƯỚNG DẪN KHOA HỌC: GS.TS Vũ Đức Thi

Thái Nguyên - Năm 2013

MỤC LỤC

LỜI CẢM ƠN	I
LỜI CAM ĐOAN.....	II
DANH MỤC CÁC THUẬT NGỮ	III
BẢNG CÁC KÝ HIỆU	IV
DANH SÁCH BẢNG.....	VI
LỜI MỞ ĐẦU	1
Chương 1. KHÁI QUÁT VỀ TẬP THÔ VÀ RÚT GỌN THUỘC TÍNH.....	5
1.1. Hệ thông tin.....	5
1.2. Tập thô	7
1.3. Bảng quyết định.....	9
1.4. Tập rút gọn và lõi	9
1.5. Ma trận phân biệt và hàm phân biệt	10
1.6. Mối liên hệ giữa các tập rút gọn của các phương pháp rút gọn thuộc tính.	11
1.6.1. Entropy trong hệ thông tin và các tính chất.	12
1.6.2. Tập rút gọn dựa trên entropy thông tin.....	14
1.6.3. Mối liên hệ của tập rút gọn dựa trên Shannon entropy	15
1.6.4. Mối liên hệ của tập rút gọn dựa trên độ khác biệt giữa các tri thức.....	19
1.7. Sự thay đổi các độ đo đánh giá hiệu năng bảng quyết định khi rút gọn thuộc tính.....	22
1.7.1. Luật quyết định và các độ đo cổ điển	23
1.7.2. Độ đo hiệu năng cải tiến của bảng quyết định	24
1.7.3. Đề xuất độ đo hiệu năng mới của bảng quyết định.....	25
1.7.4. Sự thay đổi các độ đo khi thực hiện các phương pháp rút gọn thuộc tính ...	29
1.8. Kết luận Chương 1	31
Chương 2. MỘT SỐ PHƯƠNG PHÁP RÚT GỌN THUỘC TÍNH TRONG BẢNG QUYẾT ĐỊNH.	32
2.1. Mở đầu	32

2.2. Thuật toán tìm tập rút gọn sử dụng Liang entropy.....	39
2.2.1. Tập rút gọn dựa trên Liang entropy với phân hoạch cải tiến.....	40
2.2.2. Thuật toán tìm tập rút gọn sử dụng Liang entropy.....	43
2.3. Thuật toán tìm tập rút gọn sử dụng metric	48
2.3.1. Khoảng cách Jaccard giữa hai tập hợp hữu hạn.....	49
2.3.2. Metric trên hệ thông tin	50
2.3.3. Tập rút gọn dựa trên metric	51
2.3.4. Thuật toán tìm tập rút gọn sử dụng metric	54
2.3.5. Thuật toán tìm tập rút gọn theo ngưỡng chắc chắn của bảng quyết định	59
2.4. Kết luận Chương 2	61
Chương 3: CHƯƠNG TRÌNH THỬ NGHIỆM	62
3.1. Bài toán	62
3.2. Phương pháp	62
3.3. Xây dựng chương trình thử nghiệm	63
3.4. Kết quả thử nghiệm	64
3.5. Kết luận chương 3	65
KẾT LUẬN.....	66
TÀI LIỆU THAM KHẢO	67

LỜI CẢM ƠN

Tôi xin chân thành cảm ơn đến:

- Trường Đại học Công nghệ thông tin và Truyền thông, Đại học Thái Nguyên

- Viện Công nghệ Thông tin và các thầy cô giáo đã trực tiếp giảng dạy, hướng dẫn tôi trong quá trình học tập và định hướng quan trọng trong việc hình thành ý tưởng nghiên cứu.

Tôi xin chân thành cảm ơn Chi bộ, BGH, BCH Công đoàn, Tổ Khoa học tự nhiên và cán bộ giáo viên, nhân viên Trường THPT Bình Độ đã động viên, giúp đỡ, tạo điều kiện thuận lợi cho tôi trong quá trình học tập và nghiên cứu.

Đặc biệt, tôi xin bày tỏ lòng biết ơn sâu sắc đến GS.TS Vũ Đức Thi, người thầy đã trực tiếp hướng dẫn và giúp đỡ tôi hoàn thành luận văn tốt nghiệp.

Cuối cùng xin chân thành cảm ơn những người thân và gia đình đã luôn chia sẻ mọi khó khăn và là chỗ dựa vững chắc về vật chất, tinh thần để tôi hoàn thành chương trình khóa học cũng như trong suốt thời gian hoàn thành luận văn.

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

Thái Nguyên, tháng 01 năm 2013

Tác giả

Hoàng Thị Ngọc Mai

II

LỜI CAM ĐOAN

Tôi xin cam đoan luận văn này là công trình do tôi tổng hợp và nghiên cứu.

Trong luận văn có sử dụng một số tài liệu tham khảo như đã nêu trong phần tài liệu tham khảo.

Tác giả Luận văn

Hoàng Thị Ngọc Mai

III

DANH MỤC CÁC THUẬT NGỮ

Tập thô	Rough Set
Hệ thống tin	Information System
Hệ thống tin đầy đủ	Complete Information System
Bảng quyết định	Decision Table
Bảng quyết định đầy đủ	Comple Decision Table
Bảng quyết định không nhất quán	Inconsistent Decision Table
Quan hệ không phân biệt được	Indiscernibility Relation
Rút gọn thuộc tính	Attribute Reduction
Tập rút gọn	Reduct
Tập lõi	Core
Shannon entropy	Entropy
Liang entropy	Entropy mới của Jiye Liang trong [28]

BẢNG CÁC KÝ HIỆU

$IS = (U, A, V, f)$	Hệ thông tin
$DS = (U, C \cup D, V, f)$	Cho bảng quyết định
$ U $	Số đối tượng
$ C $	Số thuộc tính điều kiện trong bảng quyết định
$u(a)$	Giá trị đối tượng của u của thuộc tính a
$[u]_B$	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 U sinh bởi tập thuộc tính 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
$PRED(C)$	Tập tất cả các rút gọn dựa trên miền dương
$HRED(C)$	Tập tất cả các rút gọn dựa trên Shannon entropy
$SRED(C)$	Tập tất cả các rút gọn của phương pháp ma trận phân biệt
$ERED(C)$	Tập tất cả các rút gọn dựa trên Liang entropy
$NERED(C)$	Tập tất cả các rút gọn dựa trên Liang entropy với phân hoạch cải tiến.
$MRED(C)$	Tập tất cả các rút gọn dựa trên metric
$KRED(C)$	Tập tất cả các rút gọn dựa trên độ đo lượng tri thức khác nhau.

$CORE_P(C)$	Tập lõi dựa trên miền dương
$CORE_H(C)$	Tập lõi dựa trên Shannon entropy
$CORE_S(C)$	Tập lõi của phương pháp ma trận phân biệt.
$CORE_E(C)$	Tập lõi dựa trên Liang entropy.
$CORE_M(C)$	Tập lõi dựa trên metric
$CORE_K(C)$	Tập lõi dựa trên độ đo lượng tri thức khác nhau.
$H(P)$	Shannon entropy của tập thuộc tính P
$H(Q \setminus P)$	Shannon entropy có điều kiện của Q khi đã biết P
$E(P)$	Liang entropy của tập thuộc tính P
$E(Q \setminus P)$	Liang entropy có điều kiện của Q khi đã biết P
$K(P)$	Tri thức sinh bởi tập thuộc tính P
$d(K(P), K(Q))$	Metric giữa hai tri thức $K(P)$ và $K(Q)$ trên hệ thông tin đầy đủ sử dụng khoảng cách Jaccard giữa hai tập hợp.
$DQP(K(P), K(Q))$	Lượng tri thức khác nhau giữa $K(P)$ và $K(Q)$

DANH SÁCH BẢNG

<i>Bảng 1.1. Bảng thông tin về bệnh cúm.....</i>	<i>6</i>
<i>Bảng 1.3. Bảng quyết định minh họa Ví dụ 1.3</i>	<i>18</i>
<i>Bảng 1.4. Bảng quyết định minh họa Ví dụ 1.4</i>	<i>46</i>
<i>Bảng 2.1. Bảng quyết định minh họa Ví dụ 2.1.</i>	<i>46</i>
<i>Bảng 2.2. Bảng quyết định về bệnh cảm cúm</i>	<i>53</i>
<i>Bảng 2.3. Bảng quyết định minh họa Ví dụ 2.5</i>	<i>57</i>