BỘ GIÁO DỤC VÀ ĐÀO TẠO

VIỆN KHOA HỌC
VÀ CÔNG NGHỆ VIỆT NAM

VIỆN CÔNG NGHỆ THÔNG TIN

NGUYỄN ĐỨC THUẦN

PHỦ TẬP THÔ
VÀ ĐỘ ĐO ĐÁNH GIÁ HIỆU NĂNG
TẬP LUẬT QUYẾT ĐỊNH

Chuyên ngành: BẢO ĐẢM TOÁN HỌC CHO MÁY TÍNH
VÀ HỆ THỐNG TÍNH TOÁN

Mã số: 62.46.35.01

TÓM TẮT LUẬN ÁN TIẾN SĨ TOÁN HỌC

HÀ NỘI - 2010

Công trình được hoàn thành tại:

VIỆN CÔNG NGHỆ THÔNG TIN

VIỆN KHOA HỌC VÀ CÔNG NGHỆ VIỆT NAM

Người hướng dẫn khoa học

1. PGS. TSKH NGUYỄN XUÂN HUY
2. PGS. TS LÊ HẢI KHÔI

Phản biện 1: GS.TS NGUYỄN THANH THỦY

Phản biện 2: PGS.TS ĐẶNG QUANG Á

Phản biện 3: PGS.TS NGUYỄN BÁ TƯỜNG

Luận án sẽ được bảo vệ trước Hội đồng chấm luận án cấp viện

họp tại: Hội trường Viện Công nghệ Thông tin

18 Hoàng Quốc Việt, Cầu Giấy, Hà Nội

vào hồi 15 giờ 00 ngày 12 tháng 01 năm 2011

Có thể tìm hiểu luận án tại thư viên:

Thư viện Quốc Gia, Thư viện Viện Công Nghệ Thông Tin

DANH MỤC CÔNG TRÌNH ĐÃ CÔNG BỐ LIÊN QUAN LUẬN ÁN

[1]	Nguyễn Đức Thuần, Nguyễn Xuân Huy (2009), " CÁC XẤP XỈ TRÊN CỦA PHỦ TẬP THÔ VÀ ÁNH XẠ ĐÓNG". *Kỷ yếu Hội Nghị Khoa Học Kỷ Niệm 25 Năm Thành Lập Viện Cơ học & Tin học Ứng dụng Tp Hồ Chí Minh*. Nxb Khoa Học Tự Nhiên và Công Nghệ, 329-333.
[2]	Nguyễn Đức Thuần, Nguyễn Xuân Huy (2009), "RÚT GỌN TẬP THUỘC TÍNH CỦA HỆ QUYẾT ĐỊNH DỰA VÀO HỌ PHỦ TẬP THÔ", *Tạp chí Khoa học & Công nghệ, ĐH Đà Nẵng*. Vol (4)-33, 64-69. Nguyen Duc Thuan (2010), "A Family of Covering Rough Sets Based Algorithm for Reduction of Attributes ", *International Journal of Computer Theory and Engineering (IJCTE)*. Vol 2(2) 180-184.
[3]	Nguyen Duc Thuan, Nguyen Xuan Huy (2009), "A New Measure to Evaluate the Consistency of a Set of Decision Rules Extracted from a Decision Table", *International Journal of Computer Electrical Engineering (IJCEE)*. Vol 1(4) 447- 451.
[4]	Nguyen Duc Thuan (2009), "Covering Rough Sets From a Topological Point of View", *International Journal of Computer Theory and Engineering (IJCTE)*. Vol 1(5) 601-604

KẾT LUẬN

Luận án đã thực hiện được các kết quả sau

1. Khảo sát tính chất toán học của các phép xấp xỉ ứng với ba loại phủ do W. Zhu, F.Y Wang đề xuất. Chỉ ra hai phủ sinh cùng một phép xấp xỉ loại 2, Trình bày điều kiện để các phép xấp xỉ là đồng nhất khi tiếp cận bằng không gian topo.

2. Xác lập mối quan hệ giữa các phép xấp xỉ dựa trên phủ với ánh xạ đóng. Đây là cơ sở để mở rộng các phép xấp xỉ cũng như kết thừa các kết quả đã có nhằm ứng dụng tập thô hiệu quả hơn.

3. Đề xuất thuật toán FC-Reduct: *tìm một rút gọn tối thiểu tập thuộc tính ứng với một họ phủ quyết định tập thô*. Độ phức tạp của thuật toán là $O(|\Delta||U|^2)$ (*tương đương với các thuật toán tìm một rút gọn tập thuộc tính trong lý thuyết tập thô cổ điển*).

4. Xây dựng được độ đo mới đánh giá hiệu năng của tập luật quyết định khắc phục được những hạn chế của các hệ độ đo trước đó.

5. Thử nghiệm các kết quả đạt được: thuật toán FC-Reduct và độ đo đánh giá hiệu năng tập luật quyết định trên các bộ cơ sở dữ liệu của UCI. Tích hợp độ đo vào phần mềm rút trích tập luật quyết định hỗ trợ xử lý thông tin dạy và học tại Đại học Nha Trang.

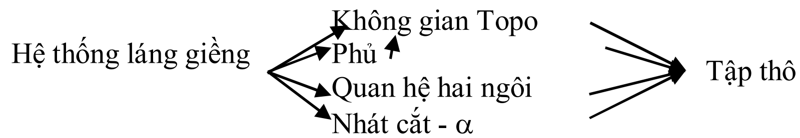
Ngoài các kết quả thu được trong luận án, các vấn đề còn phải tiếp tục, nghiên cứu và phát triển liên quan là:

- Mối quan hệ giữa phủ tập thô và các không gian toán học, lớp phụ thuộc bool dương, CSDL quan hệ...
- Phối hợp các khái niệm tập thô với các công cụ toán học xấp xỉ khác như tập mờ, xác suất, tập mơ hồ (*vague set*) ..dựa trên phủ tập thô.
- Ứng dụng tập thô vào khai thác dữ liệu (*Datamining*)
- Xây dựng các phần mềm ứng dụng, giải quyết các bài toán thực tiễn dựa vào lý thuyết tập thô mở rộng.

MỞ ĐẦU

Trong thời gian gần đây, lý thuyết tập thô do Pawlak đề xuất (1982) đã cung cấp công cụ toán học hữu ích phục vụ cho việc nghiên cứu các hệ thống thông minh, các hệ thống thông tin không đầy đủ ... Phương pháp tập thô được ứng dụng trong nhiều lĩnh vực: kinh tế, ngân hàng, tài chính, y học điều khiển tiến trình…

Lý thuyết Tập thô dựa trên cơ sở toán học là các phép xấp xỉ, quan hệ tương đương và phép phân hoạch. Chính những yếu tố đơn giản này làm tập thô dễ tiếp cận. Tuy nhiên, đó cũng là yếu tố làm hạn chế sự ứng dụng vì lớp quan hệ tương đương và phân hoạch là không lớn. Nhiều mở rộng thú vị và ý nghĩa dựa trên sự mở rộng hai khái niệm: quan hệ hai ngôi và phân hoạch hay phối hợp với các phương pháp khác. Chúng ta có thể thấy hướng mở rộng tập thô qua tổng kết của T.Y.Lin



Các hướng tiếp cận mở rộng tập thô

Sự mở rộng tập thô đã phát sinh nhiều bài toán thú vị cần nghiên cứu và giải quyết. Với mong muốn phát triển, mở rộng lý thuyết tập thô và ứng dụng, luận án đóng góp một số kết quả tập trung vào các vấn đề sau:

1. *Khảo sát, phân tích các loại phủ tập thô. Phát hiện tính chất và mối quan hệ giữa các loại phủ và các phép xấp xỉ.*
2. *Phát hiện các tính chất của ánh xạ đóng mà các phép xấp xỉ trên, xấp xỉ dưới xây dựng trên các mô hình phủ tập thô có được.*
3. *Đề xuất thuật toán rút gọn tập thuộc tính dựa vào phủ.*
4. *Xây dựng độ đo đánh giá hiệu năng tập luật quyết định được rút trích từ các bảng quyết định.*
5. *Xây dựng một ứng dụng trên mô hình lý thuyết tập thô dựa vào phủ.*

BỐ CỤC CỦA LUẬN ÁN

Luận án gồm ba chương, phần kết luận, các công trình đã công bố và tài liệu tham khảo

Chương 1: **CÁC KHÁI NIỆM CƠ SỞ**

- Trình bày các khái niệm cơ sở làm nền tảng toán học cho các chương sau.

Chương 2: **PHỦ TẬP THÔ**

Đóng góp một số kết quả về phủ tập thô:

- Điều kiện để hai phủ cùng sinh ra một phép xấp xỉ loại 2.
- Tính chất ánh xạ đóng của các phép xấp xỉ loại 1, 2, 3 ứng với ba loại phủ: đơn vị, nửa thu gọn, nửa tựa điểm.
- Một số điều kiện để các phép xấp xỉ là đồng nhất khi tiếp cận bằng không gian topo.
- Thuật toán mới *FC_reduct* rút gọn tập thuộc tính dựa vào họ phủ tập thô.
- Ứng dụng thuật toán cho bài toán xử lý thông tin dạy và học tại Đ.H Nha Trang.

Chương 3: **ĐỘ ĐO ĐÁNH GIÁ HIỆU NĂNG TẬP LUẬT QUYẾT ĐỊNH**

- Đề xuất một hệ độ đo mới đánh giá hiệu năng của tập luật quyết định. Ứng dụng hệ độ đo cho bài toán xử lý thông tin dạy và học tại Đ.H Nha Trang.

Phần Kết luận : Tổng kết những kết quả đạt được và hướng nghiên cứu tiếp theo.

Các công trình đã công bố và Tài liệu tham Khảo

Phụ lục: 1. Một số kết quả ứng dụng đạt được.

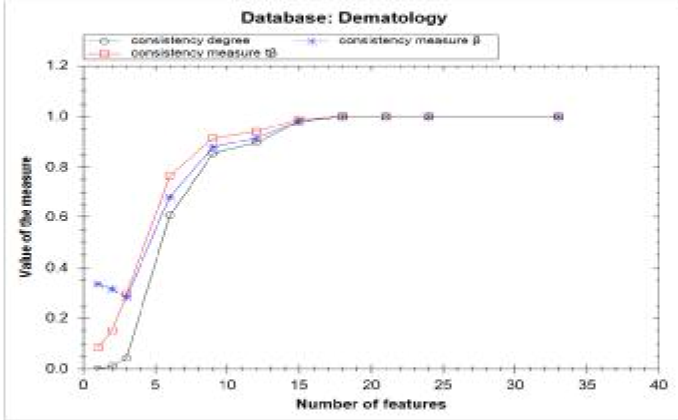
2. Biểu mẫu phiếu khảo sát thông tin dạy & học tại Đ.H Nha Trang

NỘI DUNG LUẬN ÁN

Chương 1: CÁC KHÁI NIỆM CƠ BẢN

1.1 Hệ thống thông tin và tập thô

1.1.1 Hệ thống thông tin



Hình 3.2 *Sự biến thiên của các độ đo độ nhất quán: τβ, β,* $c_c(D)$ *ứng với tập dữ liệu Dermatology*

3.4 Ứng dụng hệ độ đo cho bài toán xử lý thông tin dạy và học tại Đ.H Nha Trang

Như đã trình bày trong 2.8.1, tập luật quyết định rút trích được từ cơ sở dữ liệu khảo sát chất lượng giảng dạy tại ĐH Nha Trang được đánh giá hiệu năng theo độ đo luận án đề xuất. Các giá trị ứng với các độ đo nhằm hỗ trợ nhóm chuyên gia giáo dục đưa ra các kết luận về thông tin dạy và học, đồng thời là cơ sở để so sánh với các kết quả thu được bằng phương pháp thống kê mà các nhóm chuyên gia giáo dục tiến hành trước đây.

3.5 Kết luận chương 3

Chương này trình bày đô đo đánh giá hiệu năng tập luật quyết định được đề xuất. Hệ độ đo mới khắc phục được những hạn chế của các hệ độ đo đã có trước đó. Việc minh họa bằng cách cài đặt thực nghiệm độ đo trên ba bộ dữ liệu của UCI là Tic-tac-toe, Dermatology, Nursery đã cho thấy sự tương quan và khác biệt của các hệ độ đo. Quá trình tích hợp độ đo đề xuất vào bài toán xử lý thông tin dạy và học tại ĐH Nha Trang cho thấy khả năng ứng dụng của đô đo.

Định lý 3.11 Cho $\mathbf{T}_1 = (U, C \cup D)$, $\mathbf{T}_2 = (U, B \cup D)$ là 2 bảng quyết định, nếu $B \subseteq C$ và B là một rút gọn miền dương của D ứng với C thì $\tau\beta(\boldsymbol{T}_2) \geq \tau\beta(\boldsymbol{T}_1)$.

Bảng 3. 1 *Mô tả các tập dữ liệu thử nghiệm*

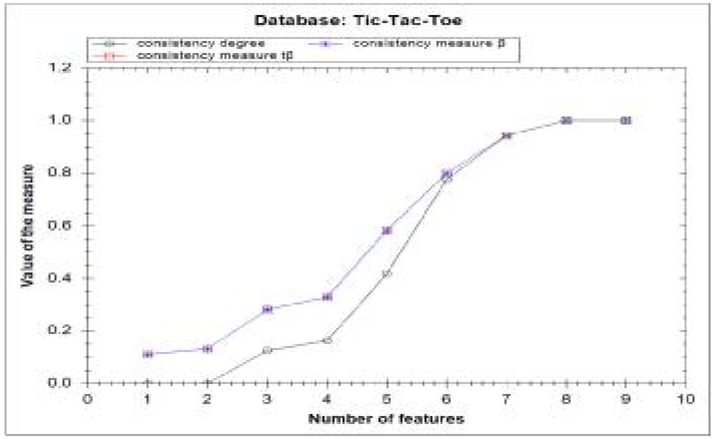
Data sets	Số lượng mẫu	Số thuộc tính ĐK	Số lớp Q.định
Tic-tac-toe	958	9	2
Dermatology	366	33	6

Bảng 3.2 *Số liệu chỉ ra sự khác biệt của $c_c(D)$, β và τβ đối với dữ liệu Tic-tac-toe*

Độ đo	Đặc trưng (Features)								
	1	2	3	4	5	6	7	8	9
$c_c(D)$	0.0000	0.0000	0.1253	0.1628	0.4186	0.7766	0.9436	1.0000	1.0000
β	0.1114	0.1322	0.2827	0.3300	0.5832	0.8000	0.9436	1.0000	1.0000
τβ	0.1114	0.1322	0.2827	0.3300	0.5832	0.8000	0.9436	1.0000	1.0000

Bảng 3.3 *S. liệu chỉ ra sự khác biệt của $c_c(D)$, β và τβ đối với dữ liệu Dermatology*

Độ đo	(Đặc trưng)Features									
	1	2	3	6	9	12	15	18	21	33
$c_c(D)$	0.0000	0.0109	0.0437	0.6066	0.8552	0.8962	0.9809	1.0000	1.0000	1.0000
β	0.3350	0.3164	0.2821	0.6826	0.8797	0.9153	0.9818	1.0000	1.0000	1.0000
τβ	0.0854	0.1581	0.2960	0.7661	0.9148	0.9415	0.9891	1.0000	1.0000	1.0000



Hình 3.1 *Sự b. thiên của các độ đo độ nhất quán: τβ, β, $c_c(D)$ ứng với dữ liệu Tic-Tac-Toe*

Hệ thống thông tin là một cặp $\mathbb{S} = (U, A)$, U là một tập hữu hạn khác rỗng các đối tượng gọi là tập vũ trụ hay là tập phổ dụng, A là một tập hữu hạn khác rỗng các thuộc tính.

1.1.2 Quan hệ không phân biệt được

Xét hệ thống thông tin $\mathbb{S} = (U, A)$. Khi đó mỗi tập thuộc tính $B \subseteq A$ đều tạo ra tương ứng một quan hệ tương đương IND(B):

$$\text{IND}(B) = \{(u,v) \in U \times U \mid a(u) = a(v), \forall a \in B\}$$

IND(B) được gọi là *quan hệ B_không phân biệt.* Lớp tương đương của $u \in U$ trong quan hệ IND(B) được kí hiệu bởi $[u]_B$. Tập thương xác định bởi quan hệ IND(B) được ký hiệu *U/IND(B)* hay *U/B*.

1.1.3 Tập thô

Cho một hệ thống thông tin $\mathbb{S} = (U, A)$. Với mỗi tập con $X \subseteq U$ và $B \subseteq A$, đặt R= IND(B), ta có 2 tập con sau

$$\underline{R}X = \{u \in U \mid [u]_B \subseteq X\}$$

$$\overline{R}X = \{u \in U \mid [u]_B \cap X \neq \emptyset\}$$

$\underline{R}X, \overline{R}X$ lần lượt gọi là *R-xấp xỉ dưới và R- tập xấp xỉ trên* của tập *X*. Từ hai tập xấp xỉ trên, người ta định nghĩa các tập

$BN_B(X) = \overline{R}X - \underline{R}X$: *biên của X trên R.*

$POS_B(X) = \bigcup_{V \in U/B} BX$: *B-vùng dương của X.*

$NEG_B(X) = U - \overline{R}X$: *B-vùng âm của X.*

Trong trường hợp *$BN_B(X) \neq \emptyset$, X được gọi là tập thô*, ngược lại *X* được gọi là *tập rõ*.

Với *B,D ⊆ A,* người ta gọi *B*-miền khẳng định dương của D là tập được xác định

$$POS_B(D) = \bigcup_{V \in U/D} (\underline{R}(V))$$

1.1.4 Các tính chất của xấp xỉ (do *Pawlak công bố 1991*)

1.1.5 Độ chính xác của xấp xỉ

Cho một hệ thống thông tin $\mathbb{S} = (U, A)$. Với mỗi tập con $X \subseteq U$ và $B \subseteq A$, đặt $R = \text{IND}(B)$, đại lượng đo sự chính xác của tập xấp xỉ X đối với phân hoạch R là giá trị

$$\alpha_R(X) = \frac{Card(\underline{R}X)}{Card(\overline{R}X)}$$

1.1.6 Bảng quyết định

Bảng quyết định là một hệ thống thông tin có dạng $\mathbb{T} = (U, C \cup D)$, với $C \cap D = \phi$, C được gọi là *tập thuộc tính điều kiện*, còn D là *tập thuộc tính quyết định.*

Cho bảng quyết định $\mathbb{T} = (U, C \cup D)$, giả sử $U/C = \{X_1, X_2, .., X_m\}$ và $U/D = \{Y_1, Y_2, .., Y_n\}$. Một lớp $X_i \in U/C$ được gọi là *nhất quán* nếu $u(d) = v(d)$, $\forall u, v \in X_i$ và $\forall d \in D$; một lớp $Y_j \in U/D$ được gọi là *nhất quán ngược* nếu $u(a) = v(a)$, $\forall u, v \in Y_j$ và $\forall a \in C$. Bảng quyết định $\mathbb{T} = (U, C \cup D)$ là *nhất quán* nếu mọi lớp $X_i \in U/C$ là nhất quán, ngược lại là không nhất quán.

Một quan hệ bộ phận $\preceq$ trên họ $\{U/B \mid B \subseteq A\}$ được định nghĩa

$U/P \preceq U/Q$ nếu và chỉ nếu : $\forall P_i \in U/P, \exists Q_j \in U/Q : P_i \subseteq Q_j$

Khi đó ta nói *Q là thô hơn P* hay *P là mịn hơn Q*.

Dễ thấy rằng, nếu $U/C \preceq U/D$ thì $\mathbb{T} = (U, C \cup D)$ được gọi là nhất quán.

1.1.7 Rút gọn và nhân

Xét một bảng quyết định $\mathbb{T} = (U, A \cup D)$. Tập thuộc tính $R \subseteq A$ được gọi là một *rút gọn* của A nếu $\text{POS}_R(D) = \text{POS}_A(D)$.

Nhân của một tập thuộc tính điều kiện A ký hiệu CORE(A), được định nghĩa CORE(A) = $\cap$RED(A).;RED(A) là tập các rút gọn của A.

Ngoài ra, người ta còn định nghĩa rút gọn C-miền khẳng định dương của D như sau: Nếu $B \subseteq C$ thỏa

Định lý 3.7 (*Cực trị cho $\tau\beta$*) Cho $\mathbf{T} = (U, C \cup D)$ là một bảng quyết định và RULE $= \{Z_{ij} \mid Z_{ij}: \text{des}(X_i) \to \text{des}(Y_j), X_i \in U/C, Y_j \in U/D\}$

Với mọi $Z_{ij} \in$ RULE, nếu $\mu(Z_{ij}) = 1$ thì độ đo $\tau\beta(\mathbf{T})$ đạt giá trị lớn nhất là 1.

Nếu m=1 và n= |U| thì $\tau\beta(\mathbf{T})$ đạt giá trị nhỏ nhất là 0.

Tính đơn điệu của độ đo $\tau\beta$ đối với bảng quyết định nhất quán ngược có thể thấy qua các định lý

Định lý 3.8 Cho $\mathbf{T}_1 = (U, C_1 \cup D_1)$, $\mathbf{T}_2 = (U, C_2 \cup D_2)$ là hai bảng nhất quán ngược. Nếu $U/C_1 = U/C_2$ và $U/D_2 \preceq U/D_1$ thì $\tau\beta(\mathbf{T}_1) \geq \tau\beta(\mathbf{T}_2)$.

Bổ đề 3.1 Cho $\mathbf{T} = (U, C \cup D)$ là một bảng quyết định, và 2 tập khác rỗng X, $Y \subseteq U$. Giả sử $X = \bigcup_{j=1}^{k} X_j$, $X_p \cap X_q = \varnothing$ với mọi p≠q, có nghĩa $\{X_1, X_2, .., X_k\}$ là một phân hoạch của X, thì

$$\frac{|X \cap Y|}{|U|}\frac{|X \cap Y^C|}{|X|} \geq \sum_{j=1}^{k} \frac{|X_j \cap Y|}{|U|}\frac{|X_j \cap Y^C|}{|X|}$$

và ta có đẳng thức nếu $|X_p \cap Y||X_q \cap Y^C| = 0$, $\forall$p≠q và p,q = 1, 2, ..,k.

Định lý 3.9 Cho $\mathbf{T}_1 = (U, C_1 \cup D_1)$, $\mathbf{T}_2 = (U, C_2 \cup D_2)$ là hai bảng nhất quán ngược. Nếu $U/D_1 = U/D_2$ và $U/C_2 \preceq U/C_1$ thì $\tau\beta(\mathbf{T}_1) \leq \tau\beta(\mathbf{T}_2)$.

Từ định nghĩa của độ nhất quán và nhất quán ngược, ta thấy độ nhất quán và nhất quán ngược tỉ lệ thuận với độ chắc chắn. Nhưng độ đo của Yuhua Qian và cộng sự không thỏa, vì vậy độ đo $\tau\beta$ là phù hợp hơn β.

Bổ đề 3.2 Cho $\mathbf{T}_1 = (U, C \cup D_1)$, $\mathbf{T}_2 = (U, B \cup D_2)$ là hai bảng quyết định, nếu $C \subseteq B$ thì $U/B \preceq U/C$ và $\tau\beta(\boldsymbol{T}_2) \leq \tau\beta(\boldsymbol{T}_1)$, dấu đẳng thức xảy ra $(\tau\beta(\boldsymbol{T}_2) = \tau\beta(\boldsymbol{T}_1))$ nếu $\boldsymbol{T}_1$, $\boldsymbol{T}_2$ là hai bảng nhất quán.

Định lý 3.10 Cho $\mathbf{T}_1 = (U, C \cup D)$, $\mathbf{T}_2 = (U, B \cup D)$ là 2 bảng quyết định, nếu $\mathbf{T}_1$ là bảng nhất quán và $B \subseteq C$. Nếu B là một C-rút gọn miền dương của D thì $\tau\beta(\boldsymbol{T}_2) = \tau\beta(\boldsymbol{T}_1)$

có một số nhược điểm hoặc là áp dụng cho mỗi luật đơn, hoặc là khi giá trị độ đo độ chắc chắn bằng 0 thì sẽ không có một luật quyết định nào của bảng quyết định là chất nhận. Yuhua Qian và cộng sự đã phân tích, đề xuất ba độ đo mới khắc phục được nhược điểm này. Tuy có được nhiều tính chất khá tốt, nhưng công thức của họ khá phức tạp, đáng lưu ý là độ đo độ nhất quán có hạn chế là không đồng biến như độ đo cổ điển. Luận án đề xuất độ đo mới khắc phục những nhược điểm của các hệ độ đo đã có.

3.2 Độ đo hiệu năng của tập luật quyết định *(Xem ở bảng dưới)*

3.3 Đề xuất độ đo hiệu năng của tập luật quyết định

Cho bảng quyết định **T**=(U, $C\cup D$) và RULE = {Z_{ij} | Z_{ij}: des(X_i) → des(Y_j), $X_i \in U/C$, $Y_j \in U/D$}

Độ đo độ chắc chắn (*certainty measure*)	
$\alpha(S) = \sum_{i=1}^{m}\sum_{j=1}^{n} s(Z_{ij})\mu(Z_{ij}) = \sum_{i=1}^{m}\sum_{j=1}^{n} \frac{\lvert X_i \cap Y_j\rvert^2}{\lvert U\rvert \lvert X_i\rvert}$	Yuhua & cộng sự
$\tau\alpha(S) = 1 - \sum_{i-1}^{m}\sum_{j=1}^{n} \frac{\lvert X_i \cap Y_j\rvert}{\lvert X_i\rvert} \frac{\lvert X_i \cap Y_j^C\rvert}{\lvert U\rvert}$	***Độ đo đề xuất***
Độ đo độ nhất quán *(consistency measure)*	
$\beta(S) = \sum_{i=1}^{m} \frac{\lvert X_i\rvert}{\lvert U\rvert}[1 - \frac{4}{\lvert \mathrm{X_i}\rvert} \sum_{j=1}^{N_i} \lvert X_i \cap Y_j\rvert \mu(Z_{ij})(1-\mu(Z_{ij}))]$	Yuhua & cộng sự
$\tau\beta(S) = 1 - \frac{n}{n-1}\sum_{i=1}^{m}\sum_{j=1}^{n} \frac{\lvert X_i \cap Y_j\rvert}{\lvert X_i\rvert} \frac{\lvert X_i \cap Y_j^C\rvert}{\lvert U\rvert}$	***Độ đo đề xuất***
Độ đo độ hỗ trợ*(consistency measure)*	
$\gamma(S) = \sum_{i=1}^{m}\sum_{j=1}^{n} s^2(Z_{ij}) = \sum_{i=1}^{m}\sum_{j=1}^{n} \frac{\lvert X_i \cap Y_j\rvert^2}{\lvert U\rvert^2}$	Yuhua & cộng sự

Định lý 3.1 Độ đo độ chắc chắn $\tau\alpha$ của **T** chính là độ đo α.

(*Các định lý 3.2-3.6 là tính chất của độ đo α do Yuhua Qian và cộng sự công bố*).

Độ đo độ nhất quán $\tau\beta$ và một số tính chất

$$1. POS_B(D) = POS_C(D)$$
$$2. \forall a \in B, POS_C(D) \neq POS_{C-\{a\}}(D)$$

B được gọi lả rút gọn C-miền khẳng định dương của D.

1.1.8 Ma trận phân biệt được và hàm phân biệt được

Xét bảng quyết định $\mathbb{T}$=(U, $C\cup D$), với U={u_1, u_2, .., u_n}. Ma trận phân biệt được của $\mathbb{T}$, ký hiệu M($\mathbb{T}$) = $(m_{ij})_{n\times n}$, là một ma trận đối xứng trong đó mỗi phần tử của nó là một tập thuộc tính được xác định như sau

$$m_{ij} = \begin{cases} \{c \in C \mid u_i(c) \neq u_j(c)\}, & u_i(D) \neq u_j(D) \\ \varnothing, & u_i(D) = u_j(D) \end{cases}$$

Hàm phân biệt được f_{T} là một hàm logic, được xác định từ ma trận phân biệt M($\mathbb{T}$) như sau $f_T(u_i) = \wedge_{i\neq j}(\vee m_{ij})$, với mỗi $u_i \in U$.

Trong đó, mỗi thuộc tính được đặt tương ứng một biến logic cùng tên và

(1) *$\vee m_{ij}$ là biểu thức tuyển của tất cả các biến c $\in$ m_{ij}, nếu $m_{ij} \neq \varnothing$,*

(2) *$\vee m_{ij}$ = true, nếu $m_{ij} = \varnothing$ và $u_i(D) = u_j(D)$,*

(3) *$\vee m_{ij}$ = false, nếu $m_{ij} = \varnothing$ và $u_i(D) \neq u_j(D)$,*

1.1.9 Luật quyết định

Cho bảng quyết định $\mathbb{T}$= (U, $C\cup D$), giả sử U/C = {X_1, X_2,.., X_m} và U/D = {Y_1, Y_2,.., Y_n}. Nếu $X_i \cap Y_j \neq \varnothing$, ký hiệu des($X_i$), des($Y_j$) lần lượt là các mô tả của các lớp tương đương tương ứng với X_i, Y_j. Một luật quyết định xác định bởi X_i, Y_j có dạng Z_{ij}: *des(X_i) → des(Y_j).*

Độ đo độ chắc chắn và độ hỗ trợ của luật quyết định Z_{ij} được định nghĩa

$$\mu(Z_{ij}) = \lvert X_i \cap Y_j\rvert / \lvert X_i\rvert \quad \text{và} \quad s(Z_{ij}) = \lvert X_i \cap Y_j\rvert / \lvert U\rvert$$

Ở đây |.| là bản số hay lực lượng của tập hợp. Để thuận tiện trong trình bày, ký hiệu $|Z_{ij}|$ thay cho $\lvert X_i \cap Y_j\rvert$.

1.1.10 Phụ thuộc độ k

Cho hệ thống thông tin $\mathbb{S} = (U, A)$, $X, Y \subseteq A$. Chúng ta nói rằng tập thuộc tính Y *phụ thuộc độ* $k \in [0,1]$ vào tập thuộc tính X, ký hiệu $X \xrightarrow{k} Y$, với k được xác định như sau (|.| là ký hiệu bản số của tập hợp.)

$$k = \frac{|POS_X(Y)|}{|U|}$$

1.2 Phủ tập thô

1.2.1 Phủ và không gian xấp xỉ phủ

Định nghĩa 1.2.1 (*Phủ*) Cho U là một tập phổ dụng, $\mathscr{C}$ là họ các tập con khác rỗng của U, $\cup\mathscr{C} = U$, khi đó $\mathscr{C}$ được gọi là một *phủ* của U.

Định nghĩa 1.2.2 (*Không gian xấp xỉ phủ*) Cho U là một tập phổ dụng, $\mathscr{C}$ là một phủ của U. Cặp thứ tự $(U, \mathscr{C})$ được gọi là một *không gian xấp xỉ phủ* (*CAS*).

Định nghĩa 1.2.3 (*Mô tả tối thiểu*) Cho một không gian *xấp xỉ phủ* $(U, \mathscr{C})$, họ các tập hợp được xác định bởi $x \in U$:

$$\mathrm{Md}(x) = \{K \in \mathscr{C} \mid x \in K \wedge (\forall S \in \mathscr{C} \wedge x \in S \wedge S \subseteq K \Rightarrow K = S)\}$$

được gọi là *mô tả tối thiểu* của x.

Định nghĩa 1.2.4 (*Nửa thu gọn*) Cho $\mathscr{C}$ là một phủ của U. $\mathscr{C}$ được gọi là (*phủ*) *nửa thu gọn* hay *nửa không dư thừa* nếu nó thỏa điều kiện

$$\forall K_1, K_2 \in \mathscr{C} \text{ và } K_1 \subseteq K_2 \Rightarrow K_1 = K_2.$$

Định nghĩa 1.2.5 (*Đơn vị*) Cho $\mathscr{C}$ là một phủ của U. $\mathscr{C}$ được gọi là (phủ) *đơn vị* nếu $\forall x \in U$, $|\mathrm{Md}(x)| = 1$.

Định nghĩa 1.2.6 *(Phủ tựa điểm)* Cho $\mathscr{C}$ là một phủ của U. $\mathscr{C}$ được gọi là *phủ tựa điểm* nếu $\forall K \in \mathscr{C}$ và $x \in K$, $K \subseteq \cup\mathrm{Md}(x)$

Định nghĩa 1.2.7 (*Phần tử loại được của một phủ*)

2.8 Ứng dụng thuật toán FC-Reduct cho bài toán xử lý thông tin dạy và học tại Đại học Nha Trang

Thuật toán FC-Reduct được sử dụng để thu gọn tập thuộc tính nhằm giảm bớt kích thước tập luật quyết định. Kết quả cũng là một kênh để các chuyên gia giáo dục tham khảo phục vụ đánh giá bộ tiêu chí khảo sát.

2.9 Kết luận chương 2

Chương này luận án trình bày những kết quả đạt được liên quan đến phủ tập thô. Cụ thể là:

Một số tính chất cơ bản về rút gọn và phép phủ xấp xỉ trên của ba loại phủ tập thô: Nửa thu gọn (*Semi-reduced*), Phủ tựa điểm (*Pointwise-covered*), Đơn vị (*Unary*). Điều kiện để hai phủ sinh ra cùng một phép xấp xỉ trên loại 2 được đề xuất và chứng minh (*định lý 2.13, hệ quả 2.1, nhận xét 2.1*). Mối liên hệ, tính chất của các phép xấp xỉ dựa vào các loại phủ này và ánh xạ đóng (*mệnh đề 2.1-2.3,nhận xét 2.2, hệ quả 2.2*). Chỉ ra một số điều kiện để các phép xấp xỉ là đồng nhất khi phủ là một không gian topo (*mệnh đề 2.4-2.6, hệ quả 2.3*).

Thuật toán FC_Reduct rút gọn tập thuộc tính dựa vào một họ phủ được đề xuất. Độ phức tạp của thuật toán $O(|\Delta||U|^2)$ (*tương đương với các giải thuật trên tập thô cổ điển*). Ứng dụng thực tế thuật toán cho bài toán xử lý thông tin dạy và học tại Đại học Nha Trang cho thấy khả năng ứng dụng và tính đúng đắn của thuật toán.

Chương 3

ĐỘ ĐO ĐÁNH GIÁ HIỆU NĂNG TẬP LUẬT QUYẾT ĐỊNH

3.1 Hạn chế của các độ đo cổ điển trên các bảng quyết định

Rút trích và đánh giá hiệu năng của tập luật quyết định là bài toán được nhiều người quan tâm. Các độ đo cổ điển và của một số tác giả

2.7.1 Thuật toán FC_Reduct rút gọn thuộc tính của họ quyết định phủ tập thô

Đầu vào: Hệ QĐ phủ $\mathbb{T} = (U, \Delta, D=\{d\})$.

Đầu ra: Một rút gọn tập thuộc tính RD of Δ..

Bước 1: Tính $CI = \sum_{x\in U} \frac{|\Delta_x \cap [x]_D|}{|\Delta_x|}$

Bước 2: *If* CI = |U| {**T** *là một hệ quyết định nhất quán*} then *goto* Bước 3 else goto Bước 5.

Bước 3: Tính Δ_x, $d(\Delta_x)$, $\forall x\in U$.

Bước 4: *begin*

for each $\mathscr{C}_i \in \Delta$ *do if*

$$\sum_{x_i\in U}\sum_{x_j\in U}\left|(\Delta_{x_i}\cap\Delta_{x_j})\cup(P_{x_i}\cap P_{x_j})\right|\left|d(\Delta_{x_i})-d(\Delta_{x_j})\right| = 0$$

then $\Delta := \Delta - \{\mathscr{C}_i\}$; {ở đây $Cov(\Delta - \{\mathscr{C}_i\}) = \{P_x \mid x\in U\}$}

endif; endfor

goto Bước 6.

end;

Bước 5: *begin*

for each $\mathscr{C}_i \in \Delta$ *d oif* $\sum_{x_i\in U}\left|\frac{|\Delta_{x_i}\cap[x_i]_D|}{|\Delta_{x_i}|} - \frac{|P_{x_i}\cap[x_i]_D|}{|P_{x_i}|}\right| = 0$

{ở đây $Cov(\Delta - \{\mathscr{C}_i\}) = \{Px \mid x\in U\}$} *then* $\Delta := \Delta - \{\mathscr{C}_i\}$;

endif; endfor

end;

Bước 6: RD= Δ; thuật toán kết thúc.

2.7.2 Đánh giá độ phức tạp thuật toán FC_Reduct

Thuật toán này có độ phức tạp là $O(|\Delta||U|^2)$ (ở đây chúng ta bỏ qua thời gian tính Δ_{xi}, P_{xi}, với i= 1..|Δ|). So sánh kết quả thử nghiệm của thuật toán với kết quả của Chen Degang,

Hệ quyết định	Thuật toán Chen Degang	Thuật toán mới
Nhất quán	Red($\{\mathscr{C}_3, \mathscr{C}_4\}, \{\mathscr{C}_2, \mathscr{C}_3\}$)	$\{\mathscr{C}_3, \mathscr{C}_4\}$
Không nhất quán	Red($\{\mathscr{C}_2, \mathscr{C}_4\}, \{\mathscr{C}_2, \mathscr{C}_3\}$)	$\{\mathscr{C}_2, \mathscr{C}_4\}$

Cho $(U, \mathscr{C})$ là một *CAS* và $K\in\mathscr{C}$. Nếu K là hợp của một số tập hợp nào đó của $\mathscr{C}$- $\{K\}$, thì chúng ta nói rằng K là một *phần tử loại được* của $\mathscr{C}$, ngược lại K là *phần tử không loại được.*

Định nghĩa 1.2.8 (*Phủ rút gọn được*) Cho $(U, \mathscr{C})$ là một *CAS*. Nếu mọi phần tử của $\mathscr{C}$ là phần tử *không loại được* thì $\mathscr{C}$ là *phủ không rút gọn được*, ngược lại $\mathscr{C}$ là *phủ rút gọn được*.

Định nghĩa 1.2.9 (*Rút gọn của một phủ*) Đối với một phủ $\mathscr{C}$ của U. Một phủ không rút gọn được có được từ việc loại bỏ các phần tử loại được của $\mathscr{C}$ gọi là một *rút gọn của phủ* $\mathscr{C}$, ký hiệu *reduct*($\mathscr{C}$).

Mệnh đề 1.2.1 Cho $\mathscr{C}$ là một phủ của U, $K\in\mathscr{C}$. K là phần tử loại được trong $\mathscr{C}$, và $K_1\in\mathscr{C}-\{K\}$, K_1 là một phần tử loại được trong $\mathscr{C}$ khi và chỉ khi nó là phần tử loại được trong $\mathscr{C}-\{K\}$.

1.2.2 Thuật toán tìm rút gọn của một phủ

Do W.Zhu & FY.Wang đề xuất. Ý tưởng: Duyệt tuần tự và loại bỏ dần các phần tử loại được (*dựa vào Định nghĩa 1.2.7-1.2.9*).

1.2.3 Các phép xấp xỉ dựa vào phủ tập thô

Cho $(U, \mathscr{C})$ là một CAS. Một tập X ⊆U. Xấp xỉ dưới, xấp xỉ trên phủ loại 1, 2, 3 của X được định nghĩa như sau

Xấp xỉ phủ dưới loại 1, 2, 3 lần lượt là $X_* = \underline{X} = X_\#$	$\cup\{K\in\mathscr{C} \mid K\subseteq X\}$	Ký hiệu FL(X), SL(X),TL(X) *K.h chung* CL(X)
Xấp xỉ phủ trên loại 1: X*	$X_*\cup\{Md(x) \mid x\in X-X_*\}$	FH(X)
Xấp xỉ phủ trên loại 2 : $\overline{X}$	$\cup\{K\in\mathscr{C} \mid K\cap X\neq\varnothing\}$	SH(X)
Xấp xỉ phủ trên loại 3 : $X^\#$	$\cup\{Md(x) \mid x\in X\}$	TH(X)

Bảng 1.2 *Các phép xấp xỉ dựa vào phủ tập thô*

1.3 Ánh xạ đóng

Cho U là một tập khác rỗng. Toán tử H: $\mathscr{P}(U)\rightarrow\mathscr{P}(U)$ ($\mathscr{P}(U)$ *là tập tất cả các tập con của* U) được gọi là một ánh xạ đóng nếu H thỏa: $\forall\ X,Y\subseteq U$

(Cl$_1$) $X \subseteq$ H(X) (*tính phản xạ*)

(Cl$_2$) $X \subseteq Y \Rightarrow$ H(X) $\subseteq$ H(Y) (*tính đồng biến*)

(Cl$_3$) H(H(X)) = H(X) (*tính lũy đẳng*)

1.4 Không gian topo

Xét tập hợp X, một họ τ các tập con của X gọi là một topo trên X, nếu thỏa các điều kiện:

1. X và ∅ thuộc τ
2. Hợp tùy ý các tập thuộc τ là thuộc τ
3. Giao của hữu hạn các tập thuộc τ là thuộc τ.

Một tập X cùng một topo τ trên X gọi là một không gian topo. Tập $G\in\tau$ được gọi là tập mở của X. Tập con F của X được gọi là tập đóng, nếu $X\backslash F$ là tập mở. Các khái niệm kinh điển liên quan cũng được trình bày: *Lân cận,Bao đóng,Phần trong,Biên, Cơ sở và Tiền cơ sở* (*Base, Subbase*).

1.5 Kết luận Chương 1

Chương này trình bày một số khái niệm cơ bản làm cơ sở toán học cần thiết để trình bày các kết quả trong các chương sau.

Chương 2: PHỦ TẬP THÔ

Các kết quả trong 2.1, 2.2 được công bố bởi W. Zhu và F.Y. Wang (2006,2007)

2.1 Tính chất của xấp xỉ phủ loại 1, 2, 3

2.1.1 Xấp xỉ phủ tập thô loại 1

A. *Sự phụ thuộc xấp xỉ dưới và xấp xỉ trên loại 1.* Cho $\mathscr{C}_1$, $\mathscr{C}_2$ là hai phủ của *U*

Định lý 2.1	$\mathscr{C}_1$, $\mathscr{C}_2$ sinh ra cùng phép xấp xỉ dưới	⇔	reduct($\mathscr{C}_1$) = reduct($\mathscr{C}_2$)
Định lý 2.2	$\mathscr{C}_1$, $\mathscr{C}_2$ sinh ra cùng phép xấp xỉ trên FH	⇔	
Định lý 2.3	$\mathscr{C}_1$, $\mathscr{C}_2$ sinh ra cùng phép xấp xỉ dưới CL	⇔	$\mathscr{C}_1$, $\mathscr{C}_2$ sinh ra cùng phép xấp xỉ trên FH

2.7 Thuật toán FC_Reduct rút gọn tập thuộc tính dựa vào họ phủ tập thô

Nhận xét 2.3 Từ định nghĩa của đại lượng Δ_x. Với (U, Δ, D={d}) là một hệ quyết định phủ nhất quán, d là một hàm quyết định d: $U \rightarrow I_d$ xác định từ tập vũ trụ U vào tập giá trị I_d. Ta có các kết quả sau

- Với mỗi x_i, $x_j \in U$, nếu $\Delta_{x_i} \subseteq \Delta_{x_j}$ thì

$$d(x_i) = d([x_i]_D) = d(\Delta_{x_i}) = d(\Delta_{x_j}) = d(x_j) = d([x_j]_D)$$

- Nếu $d(x_i) \neq d(x_j)$ thì $\Delta_{x_i} \cap \Delta_{x_j} = \varnothing$ có nghĩa là $\Delta_{x_i} \not\subset \Delta_{x_j}$ và $\Delta_{x_j} \not\subset \Delta_{x_i}$.

Định lý 2.19 Cho (U, Δ, D={d}) là một hệ quyết định phủ, ta có

(U, Δ, D={d}) là một hệ quyết định phủ nhất quán khi và chỉ khi thỏa

$$\sum_{x\in U} \frac{\left|\Delta_x \cap [x]_D\right|}{\left|\Delta_x\right|} = |U|$$

Giả sử Cov(Δ)≤U/D, $\mathscr{C}_i\in\Delta$, $\mathscr{C}_i$ là không cần thiết khi và chỉ khi thỏa

$$\sum_{x_i\in U}\sum_{x_j\in U}\left|(\Delta_{x_i}\cap\Delta_{x_j})\cup(P_{x_i}\cap P_{x_j})\right|\left|d(\Delta_{x_i}) - d(\Delta_{x_j})\right| = 0$$

Ở đây, Cov(Δ-{$\mathscr{C}i$})={P_x | $x\in U$}=Cov(P), Cov(Δ)={Δ_x | $x\in U$}.

Định lý 2.20 Cho (U, Δ, D={d}) là một hệ quyết định không nhất quán. $P\subseteq\Delta$, $POS_P(D)$= $POS_\Delta(D)$ nếu và chỉ nếu$\forall x_i\in U$, ta có

$$\left|\frac{\left|\Delta_{x_i}\cap[x_i]_D\right|}{\left|\Delta_{x_i}\right|} - \frac{\left|P_{x_i}\cap[x_i]_D\right|}{\left|P_{x_i}\right|}\right| = 0$$

Định lý 2.21 Cho hệ quyết định phủ nhất quán T=(U,Δ,D). Xét hai họ phủ P^1, P^2 : $P^2 \subseteq P^1 \subseteq \Delta$, Cov($P^i$)≤U/D, i=1,2, ∀ $\mathscr{C}_k \in P^2 \subseteq P^1$, nếu $\mathscr{C}_k$ không dư thừa trong P^1 thì $\mathscr{C}_k$ không dư thừa trong P^2.

Định lý 2.22 Cho hệ quyết định Không nhất quán T=(U,Δ,D) Xét hai họ phủ P^1, P^2 : $P^2 \subseteq P^1 \subseteq \Delta$, $POS_{P^1}(D) = POS_{P^2}(D) \neq U$,∀ $\mathscr{C}_k \in P^2 \subseteq P^1$, nếu $\mathscr{C}_k$ không dư thừa trong P^1 thì $\mathscr{C}_k$ không dư thừa trong P^2.

đối với D, và $POS_P(D)=POS_\Delta(D)$ thì P được gọi là một rút gọn của Δ ứng với D.

2.6.3 Một số kết quả liên quan giữa họ phủ và phủ suy dẫn

Cheng Degang và cộng sự đã đưa ra các kết quả sau

Định lý 2.15 Giả sử U là tập phổ dụng hữu hạn và $\Delta=\{\mathscr{C}_i : i=1,..m\}$ là một họ các phủ của U, các mệnh đề sau là đúng

(1) $\Delta_x=\Delta_y$ nếu và chỉ nếu với mọi $\mathscr{C}_i\in\Delta$ ta có $C_{ix}=C_{iy}$.

(2) $\Delta_x\supset\Delta_y$ nếu và chỉ nếu với mọi $\mathscr{C}_i\in\Delta$ ta có $C_{ix}\supseteq C_{iy}$ và tồn tại tối thiểu một $\mathscr{C}_k\in\Delta$ mà $C_{kx}\supset C_{ky}$.

(3) $\Delta_x\not\subset\Delta_y$ và $\Delta_y\not\subset\Delta_x$ nếu và chỉ nếu tồn tại $\mathscr{C}_i$, $\mathscr{C}_j\in\Delta$ mà $C_{ix}\subset C_{iy}$ và $C_{jx}\supset C_{jy}$ hay tồn tại $\mathscr{C}_k\in\Delta$ mà $C_{kx}\not\subset C_{ky}$ và $C_{ky}\not\subset C_{kx}$.

Định lý 2.16 Giả sử $Cov(\Delta)\leq U/D$, $\mathscr{C}_i\in\Delta$, $\mathscr{C}_i$ là cần thiết có nghĩa $Cov(\Delta-\{\mathscr{C}_i\})\leq U/D$ là sai nếu và chỉ nếu tồn tại ít nhất một cặp x_i, $x_j\in U$ thỏa $d([x_i]_D)\neq d([x_j]_D)$, quan hệ giữa chúng tương ứng với Δ sẽ thay đổi sau khi $\mathscr{C}_i$ bị loại bỏ khỏi Δ..

Định lý 2.17 Giả sử $Cov(\Delta)\leq U/D$, $P\subseteq\Delta$ thì $Cov(P)\leq U/D$ nếu và chỉ nếu với mọi cặp x_i, $x_j\in U$ thỏa $d([x_i]_D)\neq d([x_j]_D)$, quan hệ giữa x_i, x_j ứng với Δ tương đương với quan hệ của chúng đối với P, nghĩa là $\Delta_{x_i}\not\subset\Delta_{x_j}$ và $\Delta_{x_j}\not\subset\Delta_{x_i}\Leftrightarrow P_{x_i}\not\subset P_{x_j}$ và $P_{x_j}\not\subset P_{x_i}$

Định lý 2.18 Hệ quyết định không nhất quán (U, Δ, $D=\{d\}$) có các tính chất sau

(1) $\forall x_i\in U$, nếu $\Delta_{x_i}\subset POS_\Delta(D)$ thì $\Delta_{x_i}\subseteq[x_i]_D$; nếu $\Delta_{x_i}\not\subset POS_\Delta(D)$ thì $\forall x_k\in U, \Delta_{x_i}\subseteq[x_k]_D$ là không đúng.

(2) $\forall P\subseteq\Delta, POS_P(D)=POS_\Delta(D)$ nếu và chỉ nếu

$$\underline{P}(X)=\underline{\Delta}(X), \forall X\in U/D\ .$$

(3) $\forall P\subseteq\Delta$, $POS_P(D)=POS_\Delta(D)$ nếu và chỉ nếu

$$\forall x_i\in U,\ \Delta_{x_i}\subseteq[x_i]_D\Leftrightarrow P_{x_i}\subseteq[x_i]_D$$

B. *Tiên đề cho phép xấp xỉ dưới*

Định lý 2.4 Cho U là một tập khác rỗng. Nếu tồn tại 1 toán tử $L: \mathscr{P}(U)\rightarrow\mathscr{P}(U)$ thỏa các tính chất sau: $\forall X, Y\subseteq U$

(1L) $L(U) = U$

(3L) $L(X)\subseteq X$

(5L) $L(L(X)) = L(X)$

(7L) $X\subseteq Y\Rightarrow L(X)\subseteq L(Y)$

thì tồn tại một phủ $\mathscr{C}$ của U có tính chất toán tử xấp xỉ dưới CL được sinh bởi $\mathscr{C}$ là L. (*Chú ý: ký hiệu (1L) – (7L) là số thứ tự các tính chất của phép xấp xỉ dưới, xấp xỉ trên do Pawlak công bố*)

C. *Tiên đề cho phép phủ xấp xỉ trên loại 1*

Bài toán tiên đề hóa cho xấp xỉ phủ trên loại 1 vẫn còn là bài toán mở.

2.1.2 Xấp xỉ phủ tập thô loại 2

A. *Sự phụ thuộc xấp xỉ dưới và xấp xỉ trên tập thô loại 2*

Định lý 2.5 Phép xấp xỉ phủ dưới và xấp xỉ phủ trên loại 2 không xác định duy nhất lẫn nhau.

B. *Tiên đề các phép phủ xấp xỉ trên loại 2*

Bài toán tiên đề hóa cho xấp xỉ phủ trên loại 2 vẫn còn là bài toán mở.

2.1.3 Xấp xỉ phủ tập thô loại 3

A. *Sự phụ thuộc xấp xỉ phủ dưới và xấp xỉ phủ trên loại 3*

Cho $\mathscr{C}_1$, $\mathscr{C}_2$ là hai phủ của U

Định lý 2.6	$\mathscr{C}_1$, reduct($\mathscr{C}_1$) sinh ra cùng phép xấp xỉ dưới và xấp xỉ trên loại 3		
Định lý 2.7	$\mathscr{C}_1$, $\mathscr{C}_2$ sinh ra cùng phép xấp xỉ trên TH	$\Leftrightarrow$	reduct($\mathscr{C}_1$)= reduct($\mathscr{C}_2$)

Chú ý 2.1: Hai phủ cùng sinh ra xấp xỉ trên loại 3 nhưng không có cùng rút gon.

B. *Tiên đề các phép xấp xỉ phủ trên loại 3*

Bài toán tiên đề hóa phép xấp xỉ trên phủ loại 3 vẫn còn là bài toán mở.

2.2 Mối quan hệ giữa ba loại phủ tập thô

		FH	TH	SH
$\mathscr{C}$ là một đơn vị	$\Leftrightarrow$	FH = TH		
$\mathscr{C}$ là phủ tựa điểm	$\Leftrightarrow$		TH = SH	
*$\mathscr{C}$ là một nửa thu gọn **	$\Rightarrow$		TH = SH	
$\mathscr{C}$ là một phân hoạch	$\Leftrightarrow$	FH = TH = SH		

Bảng 2.1 *Điều kiện để các phép xấp xỉ phủ trên bằng nhau*

2.3 Một số kết quả về xấp xỉ phủ loại 2

Định lý 2.13 Cho $\mathscr{C}_1$, $\mathscr{C}_2$ là các phủ của U, $\mathscr{C}_1$ và $\mathscr{C}_2$ cùng xác định xấp xỉ phủ dưới và xấp xỉ phủ trên loại 2 nếu chúng thỏa các điều kiện sau

1. reduct($\mathscr{C}_1$) = reduct($\mathscr{C}_2$)
2. $\mathscr{C}_1$ và $\mathscr{C}_2$ là các phủ tựa điểm

Hệ quả 2.1 Cho $\mathscr{C}_1$, $\mathscr{C}_2$ là các phủ của U, $\mathscr{C}_1$ và $\mathscr{C}_2$ sinh ra cùng xấp xỉ dưới và xấp xỉ trên phủ loại 2, nếu chúng thỏa các điều kiện sau

1. reduct($\mathscr{C}_1$) = reduct($\mathscr{C}_2$)
2. $\mathscr{C}_1$ và $\mathscr{C}_2$ là các phủ nửa thu gọn

Nhận xét 2.1 Cho $\mathscr{C}$ là một phủ của U, $\mathscr{C}$ và reduct(C) chưa chắc sinh ra cùng xấp xỉ trên loại 2 (*ngay cả khi reduct($\mathscr{C}$) là một phân hoạch*).

2.4 Tính chất ánh xạ đóng của ba phép xấp xỉ dựa vào phủ

2.4.1 Tính chất giữa ánh xạ đóng của ba phép xấp xỉ phủ trên ứng với phủ Đơn vị

Mệnh đề 2.1 Cho $\mathscr{C}$ là một phủ của U, nếu $\mathscr{C}$ là (phủ) đơn vị thì FH sinh bởi $\mathscr{C}$ thỏa tính chất

$\forall X, Y \subseteq U, X \subseteq Y \Rightarrow FH(X) \subseteq FH(Y)$ (*tính đồng biến*)

và TH sinh bởi $\mathscr{C}$ thỏa:

$TH(TH(X)) = TH(X)$ (*tính lũy đẳng*)

2.6 Rút gọn tập thuộc tính dựa vào họ phủ tập thô

Các khái niệm và kết quả trong 2.6.1, 2.6.2 do Cheng Degang và cộng sự đề xuất.

2.6.1 Một số khái niệm và kết quả cơ sở

Với $\mathscr{C} = \{C_1, C_2, .., C_n\}$ là một phủ của U. Với mọi $x \in U$, đặt $C_x = \cap\{C_j \in \mathscr{C}: x \in C_j\}$. Cov($\mathscr{C}$)=$\{C_x: x \in U\}$ cũng là một phủ của U được gọi là một phủ suy dẫn của $\mathscr{C}$. Khái niệm phủ suy dẫn của một họ phủ tập thô cũng được định nghĩa tương tự:

Cho $\Delta = \{\mathscr{C}_i \mid i = 1,..,m\}$ là một họ phủ của U. Với mọi $x \in U$, đặt $\Delta_x = \cap\{C_{ix} \mid C_{ix} \in Cov(C_i), x \in C_{ix}\}$ thì Cov(Δ)=$\{\Delta_x: x \in U\}$ cũng là một phủ của U được gọi là một phủ suy dẫn của Δ.

2.6.2 Rút gọn tập thuộc tính các hệ thống quyết định nhất quán và không nhất quán

Xét $(U, \Delta, D=\{d\})$ là một hệ quyết định nhất quán. Với $\mathscr{C}_i \in \Delta$, nếu Cov(Δ-$\{\mathscr{C}_i\}$) $\leq U/D$, thì $\mathscr{C}_i$ thuộc Δ được nói là không cần thiết đối với D, ngược lại $\mathscr{C}_i$ được nói là cần thiết đối với D. Tập $P \subseteq \Delta$ thỏa Cov(P) $\leq U/D$, nếu mọi phần tử thuộc P là cần thiết, có nghĩa là $\forall \mathscr{C}_i \in P$, Cov($\Delta$-$\{\mathscr{C}_i\}$)$\leq U/D$ là sai thì P được gọi là một rút gọn của D.

Tập tất cả các phần tử cần thiết trong Δ tương ứng với D được gọi là nhân của Δ ứng với D, ký hiệu $Core_D(\Delta)$. Rút gọn của một hệ quyết định nhất quán là một tập tối thiểu các thuộc tính điều kiện đảm bảo chắc chắn các luật quyết định là nhất quán.

Xét d: $U \rightarrow I_d$ là hàm quyết định được định nghĩa $d(u) = u(D)$, $\forall u \in U$. Ta có $\forall x_i, x_j \in [u]_D \Leftrightarrow x_i(D) = x_j(D) = u(D)$, vì vậy không nhầm lẫn có thể viết $d(x_i) = d(x_j) = d([u]_D) = d(u)$.

Tương tự như 1.1.6, một hệ quyết định phủ (U, Δ, D) là không nhất quán khi $POS_\Delta(D) \neq U$. Nếu $POS_\Delta(D) = POS_{\Delta-\{C_i\}}(D)$, thì $\mathscr{C}_i$ là phần tử không cần thiết tương ứng với D. Ngược lại, $\mathscr{C}_i$ là phần tử cần thiết tương ứng với D. Với mỗi $P \subseteq \Delta$, nếu mọi phần tử trong P là phần tử cần thiết

Hệ quả 2.3 Cho (U, τ) là một không gian topo được cảm sinh từ một quan hệ hai ngôi R có tính phản xạ và bắc cầu. Xét một phủ của U là $\mathscr{C}=\{ r_R(x) | x\in U\}$, ta có:

(1) $X_+ = C_+X = \underline{X} = \oint_R X = \tau\underline{X}$

(2) $C^+X = \oiint_R X$

(3) $\overline{X} = \tau\overline{X}$

Hệ quả này cho thấy mối quan hệ giữa các xấp xỉ của Yao(3), A.Mkozae và cộng sự (5). Trong trường hợp tổng quát $\overline{X}, \tau\overline{X}$ là khác nhau. Tuy nhiên ta có tính chất sau

Cho (U, τ_S) là một không gian topo được xây dựng theo 2.5.1 b. Xét một phủ của U là $\mathscr{C} = \tau_S$. Mọi tập con X $\subseteq \mathscr{P}$(U), thì $\tau\overline{X} \subseteq \overline{X}$

Việc rút gọn phủ khi phủ là một không gian topo. Có thể thực hiện việc rút gọn bằng thuật toán của W.Zhu & Wang. Ngoài ra, ta còn có thể sử dụng chuyển đổi phủ do Guilong Liu, Ying Sai đề xuất. Phép chuyển đổi này được định nghĩa:

Gọi $\mathbb{C}(U)$ là tập tất cả các phủ của U, định nghĩa một phép chuyển đổi $\mathscr{F}$ từ $\mathbb{C}(U)$ đến $\mathbb{C}(U)$: $\mathscr{F}: \mathbb{C}(U) \rightarrow \mathbb{C}(U)$, với $\mathscr{C}\in \mathbb{C}(U)$: $\mathscr{F}(\mathscr{C}) = \mathscr{C} = \{N(x) \mid x\in U\}$.

Đối với phép chuyển đổi phủ này các phép xấp xỉ phủ trên X^+, C^+X và xấp xỉ phủ dưới C_+X là không đổi. Nhưng, phép chuyển đổi này không bảo toàn không gian topo. Nói khác hơn, xấp xỉ của Yao(3) và xấp xỉ của A.M . Kozae và cộng sự (5) không bảo toàn với phép chuyển đổi này. Có thể thấy qua phản ví dụ sau

Giả sử $U=\{a, b, c, d\}$, topo τ được định nghĩa trên U: $\mathscr{C} = \tau = \{\varnothing, U, \{d\}, \{c, d\}\}$, $\mathscr{F}(\mathscr{C}) = \{N(a)= U= N(b), N(c) = \{c, d\}, N(d) = \{d\}\}$. Dễ thấy $\mathscr{F}(\mathscr{C})$ không còn là một topo.

Tuy nhiên SH chưa chắc thỏa tính lũy đẳng nếu $\mathscr{C}$ là (*phủ*) đơn vị.

Phản ví dụ 2.6 Cho $U = \{a, b, c, d\}$, $K_1= \{a, b\}$, $K_2= \{a, d, c\}$, $K_3= \{a, b, d\}$, $\mathscr{C}= \{K_1, K_2, K_3\}$. $\mathscr{C}$ là một phủ đơn vị của U. Với $X= \{c\}$, chúng ta có $SH(X) = \cup \{K \mid K\in\mathscr{C}, K\cap X \neq \varnothing\} = \{a, d, c\} \neq SH(SH(X)) = \{a, b, c, d\}$.

2.4.2 Tính chất ánh xạ đóng của ba phép xấp xỉ phủ trên ứng với phủ Tựa điểm

Mệnh đề 2.2 Cho $\mathscr{C}$ là một phủ của U, nếu $\mathscr{C}$ là phủ tựa điểm thì FH sinh bởi $\mathscr{C}$ thỏa tính chất

$\forall\, X, Y \subseteq U, X \subseteq Y \Rightarrow FH(X) \subseteq FH(Y)$ (tính đồng biến)

Khi $\mathscr{C}$ là một phủ tựa điểm của U, nhưng TH, SH chưa chắc thỏa tính lũy đẳng

Phản ví dụ 2.7 Cho $U = \{a, b, c, d\}$, $K_1= \{a, b\}$, $K_2= \{a, c\}$, $K_3= \{b, d\}$, $K_4= \{d\}$. $\mathscr{C}= \{K_1, K_2, K_3, K_4\}$. $\mathscr{C}$ là một phủ tựa điểm của U. Với $X= \{a\}$, chúng ta có $TH(X) = SH(X) = \{a, b, c\} \neq SH(SH(X)) = \{a, b, c, d\}$.

2.4.3 Tính chất ánh xạ đóng của ba phép xấp xỉ phủ trên ứng với phủ Nửa thu gọn

Nhận xét 2.2 Cho $\mathscr{C}$ là một phủ của U, nếu $\mathscr{C}$ nửa thu gọn, thì TH, SH sinh bởi $\mathscr{C}$ chưa chắc thỏa tính lũy đẳng.

Hệ quả 2.2 Cho $\mathscr{C}$ là một phủ của U, nếu $\mathscr{C}$ là một nửa thu gọn thì FH có tính đơn điệu.

	Phủ đơn vị	Phủ tựa điểm	Phủ nửa thu gọn
FH	Ánh xạ đóng	Ánh xạ đóng	Ánh xạ đóng
SH	Ánh xạ đóng	-	-
TH	-	-	-

Bảng 2.3 *Tính chất ánh xạ đóng của ba phép xấp xỉ phủ trên sinh bởi ba loại phủ*

2.5 Mối quan hệ giữa các phép xấp xỉ phủ dựa vào không gian topo

2.5.1 Quan hệ hai ngôi và không gian topo

a. Không gian topo được xây dựng từ một quan hệ hai ngôi

Giả sử R là một quan hệ hai ngôi tùy ý xác định trên U, cặp (U, R) được gọi là một không gian xấp xỉ xác định bởi quan hệ hai ngôi R. Ứng với R, có thể định nghĩa láng giềng trái, phải của một phần tử x thuộc U lần lượt như sau: $l_R(x) = \{y \mid y \in U,\ yRx\}$ và $r_R(x) = \{y \mid y \in U,\ xRy\}$.

Xây dựng topo τ_1 sử dụng R-láng giềng phải (tương tự, topo τ_2 sử dụng R-láng giềng trái), chúng ta xem họ $S_1 = \{r_R(x) \mid x \in U\}$ là một tiền cơ sở của topo τ_1 và ký hiệu $S_x = \{G \in S_1 \mid x \in G\}$. Topo τ_1 được gọi là cảm sinh từ quan hệ hai ngôi R.

b. Không gian topo được xây dựng từ một họ phủ

Một hệ thống thông tin $\mathbb{S} = (U, A)$, U là một tập hữu hạn khác rỗng các đối tượng, A là một tập hữu hạn khác rỗng các thuộc tính. Với mỗi thuộc tính $a \in A$ xác định một quan hệ hai ngôi R_a trên UxU như sau

$\forall u, v \in U,\ u\ R_a\ v$ khi và chỉ khi $u(a) \cap v(a) \neq \varnothing$

Với định nghĩa này, R_a xác định một phủ $\mathscr{C}_a$ của U là một topo cảm sinh từ quan hệ hai ngôi R_a. Với tất cả các thuộc tính thuộc A, chúng ta sẽ có một topo τ_S được sinh từ tiền cơ sở $\bigcup_{a \in A} S_a$. Trong đó, S_a là một tiền cơ sở của topo τ_a.

Nếu (U, τ_S) là một không gian topo được xây dựng từ một họ phủ $\{\mathscr{C}_a \mid \forall a \in A\}$ sinh ra từ tập các quan hệ $\{R_a \mid \forall a \in A\}$ thì τ_S được gọi là phủ được sản sinh từ hệ thống thông tin $\mathbb{S}$.

c. Khái niệm rút gọn không gian topo sản sinh từ một tập các quan hệ hai ngôi

Xét không gian topo (U, τ) sinh ra từ tập các quan hệ hai ngôi RA, ký hiệu β_{RA} là cơ sở của (U, τ). Với $P \subseteq RA$, $r \in P$, r được gọi là không cần thiết trong P nếu và chỉ nếu: $\beta_P = \beta_{(P-r)}$. Tập M được gọi là một rút gọn của P, nếu và chỉ nếu: (i) $\beta_P = \beta_M$, (ii) $\beta_{P-\{r\}} \neq \beta_M, \forall r \in M$

d. Danh sách các phép xấp xỉ đã được các tác giả định nghĩa

Cho $(U, \mathscr{C})$ là một không gian xấp xỉ phủ. $N(x) = \cap\{K \in \mathscr{C} \mid x \in K\}$ là một lân cận của x. Ký hiệu X^C cho phần bù của X đối với U. (U, τ) là một không gian topo sử dụng các R-láng giềng phải. Khảo sát các phép xấp xỉ sau

W.Zhu (1)	
$X_+ = \cup\{K \in \mathscr{C} \mid K \subseteq X\}$	$X^+ = X_+ \cup \{N(x) \mid x \in X - X_+\}$
Xu, Zhang (2)	
$C_+X = \{x \in U \mid (\cap Md(x)) \subseteq X\}$	$C^+X = \{x \in U \mid (\cap Md(x)) \cap X \neq \varnothing\}$
Yao (3)	
$\underline{X} = \bigcup_{r_R(x) \subseteq X} r_R(x)$	$\overline{X} = ((\underline{X^C})^C)$
Yao (4)	
$\oint_R X = \{x \in U : r_R(x) \subseteq X\}$	$\oiint_R X = \{x \in U : r_R(x) \cap X \neq \varnothing\}$
A.M . Kozae, A.A. Abo Khadra, T. Medhat (5)	
$\tau \underline{X} = X^0$	$\tau \overline{X} = \cap\{F \subseteq U : X \subseteq F \wedge F \text{ dóng}\}$

Bảng 2.4 *Các phép xấp xỉ phủ định nghĩa trên không gian topo*

2.5.2 Mối quan hệ giữa các xấp xỉ dựa vào không gian topo

Xét hai tập con đáng chú ý của $\mathscr{P}(U)$:

$G = \{X \mid X \in \mathscr{P}(U),\ \oiint_R X = \varnothing\}$ và $H = \{X \in \mathscr{P}(U) \mid \exists Y \in \mathscr{P}(U),\ X = \oiint_R Y\}$

Mệnh đề 2.4 Nếu R có tính bắc cầu thì

1. $\oiint_R \oiint_R X \subseteq \oiint_R X \quad \forall X \in \mathscr{P}(U)$
2. Nếu $\oiint_R$ có tính lũy đẳng thì $G \cap H = \varnothing$

Trong phần sau, xét một phủ đặc biệt $\mathscr{C} = \tau_S$.Ở đây (U, τ_S) là một không gian topo được xây dựng trong 2.5.1 b.

Mệnh đề 2.5 Cho (U, τ) là một không gian topo sinh bởi quan hệ hai ngôi R. Nếu R là một quan hệ hai ngôi có tính phản xạ thì hai phép xấp xỉ Yao (3) và Yao (4) là đồng nhất.

Mệnh đề 2.6 Cho (U, τ_S) là một không gian topo được xây dựng như trong 2.5.1 b. Xét một phủ đặc biệt của U là $\mathscr{C} = \tau_S$, chúng ta có $X_+ = \tau \underline{X}$ với mọi $X \subseteq U$.