

BỘ GIÁO DỤC VÀ ĐÀO TẠO
ĐẠI HỌC ĐÀ NẴNG

HUỲNH THỊ YẾN GIANG

**CÁC CẤU TRÚC ĐẠI SỐ CỦA TẬP THÔ
VÀ NGŨ NGHĨA CỦA TẬP MỜ TRONG
LÝ THUYẾT TẬP THÔ**

Chuyên ngành: Phương pháp Toán sơ cấp

Mã số: 60.46.40

TÓM TẮT LUẬN VĂN THẠC SĨ KHOA HỌC

Đà Nẵng - 2011

Công trình được hoàn thành tại
ĐẠI HỌC ĐÀ NẴNG

Người hướng dẫn khoa học: **PGS.TS. NGUYỄN GIA ĐỊNH**

Phản biện 1: **TS. NGUYỄN DUY THÁI SƠN**

Phản biện 2: **GS.TSKH. NGUYỄN VĂN MẬU**

Luận văn được bảo vệ tại Hội đồng chấm luận văn tốt nghiệp Thạc sĩ khoa học họp tại Đại học Đà Nẵng vào ngày 22 tháng 10 năm 2011.

** Có thể tìm hiểu luận văn tại:*

- Trung tâm Thông tin - Học liệu, Đại học Đà Nẵng
- Thư viện trường Đại học Sư Phạm, Đại học Đà Nẵng.

MỞ ĐẦU

1. Lý do chọn đề tài

Vào năm 1982, trong International Journal of Information and Computer Sciences, bài báo với tựa đề Rough sets của Z. Pawlak đã đánh dấu sự ra đời của một lĩnh vực hoàn toàn mới của toán học và tin học, đó là lý thuyết tập thô và Z. Pawlak được xem như là cha đẻ của lý thuyết này. Tiếp theo vào những năm 1985, 1992, 1995 và 1998, những công trình sáng giá của Z. Pawlak đã khẳng định vai trò quan trọng và hấp dẫn của lĩnh vực này. Sự đóng góp to lớn về mặt toán học cũng như các ứng dụng tuyệt vời, đa dạng và phong phú của lý thuyết tập thô không thể không kể đến các tác giả Slowinski (1992), Ziarko (1993), Szkladow (1993), Jackson (1994, 1996), Lin-Wildberger (1995), Wang (1995), Tsumoto (1996), Pagliani (1996, 1998), Kryszkiewicz (1998), Lin-Cercorn (1997), Cattaneo (1997, 1998), Cattaneo-Ciucci (2002, 2004), Polkowski (2006), Ivo Dýntsch (2006), Hassanien - Suraj - Slezak - Lingras (2008).

Trong lý thuyết tập thô, dữ liệu được biểu diễn thông qua hệ thống tin hay bảng quyết định và chất lượng của thông tin được đo bằng cách sử dụng khái niệm tập xấp xỉ trên và xấp xỉ dưới. Từ những bảng dữ liệu lớn với dữ liệu dư thừa, không hoàn hảo, dữ liệu liên tục hay dữ liệu biểu diễn dưới dạng ký hiệu, lý thuyết tập thô cho phép khám phá tri thức từ những loại dữ liệu như vậy nhằm phát hiện ra những quy luật tiềm ẩn từ khối dữ liệu này.

2. Mục đích nghiên cứu

Bằng việc sử dụng hệ thống tin chưa đầy đủ với sự hỗ trợ tập các đối tượng X , đề tài sẽ bàn đến việc đại số hoá có thể được về đại số cụ thể của tập lũy thừa của X thông qua dàn tựa BZ. Cấu trúc này cho phép chúng ta xác định được hai xấp xỉ thô dựa trên quan hệ tương tự và quan hệ loại trừ, với quan hệ thứ hai luôn tốt hơn quan hệ thứ nhất. Tiếp đến, đề tài chuyển hướng quan tâm

đến các tập thô Pawlak và xét một số cấu trúc đại số có thể được của chúng. Sau đó để tôn thêm vẻ đẹp và tầm quan trọng của lý thuyết tập thô, đề tài bàn đến ngữ nghĩa của tập mờ trong lý thuyết tập thô. Cuối cùng là hai ứng dụng để minh họa là khám phá tri thức theo tiếp cận tập thô và vai trò của tập thô trong bài toán nhận dạng mặt người.

3. Đối tượng và phạm vi nghiên cứu

Nghiên cứu các khái niệm về tập thô, đại số hóa của đại số cụ thể của tập lũy thừa của X thông qua dàn tựa BZ.

Tìm hiểu mối quan hệ giữa lý thuyết tập mờ và lý thuyết tập thô qua việc nhận được một khung ngữ nghĩa của tập mờ .

Nghiên cứu từ các tài liệu liên quan đến tập thô, tập mờ của PGS. TS Nguyễn Gia Định (2008), Jan Bazan, Nguyen Hung Son, Marcin Szczuka (2004), G.Cattaneo (1997, 1998), Z.Pawlak (1982, 1985, 1992, 1998), Y. Yao (2004), W. P. Ziarko(1994), L. A. Zadeh (1965).

4. Phương pháp nghiên cứu

Phương pháp nghiên cứu chủ yếu của luận văn sẽ là khảo sát, phân tích, tổng hợp và làm sáng tỏ các kết quả trong các bài báo khoa học về lý thuyết tập thô và ứng dụng được công bố vào những năm gần đây, để từ đó tạo ra được tài liệu cần thiết và những đề xuất hữu ích đáp ứng trong việc nghiên cứu về lý thuyết tập thô.

5. Ý nghĩa khoa học và thực tiễn của đề tài

Tổng quan các kết quả của các tác giả đã nghiên cứu liên quan đến các cấu trúc đại số của tập thô, ngữ nghĩa của tập mờ trong lý thuyết tập thô và hai ứng dụng đặc trưng và quan trọng trong lý thuyết tập thô.

Chứng minh chi tiết và làm rõ một số mệnh đề, cũng như đưa ra một số ví dụ minh họa đặc sắc nhằm làm cho người đọc dễ dàng tiếp cận vấn đề được đề cập.

6. Cấu trúc luận văn

Chương 1: Các cấu trúc đại số của tập thô

Chương 2: Ngữ nghĩa của tập mờ trong lý thuyết tập thô

Chương 3: Các ứng dụng đặc trưng của lý thuyết tập thô

Chương 1

CÁC CẤU TRÚC ĐẠI SỐ CỦA TẬP THƠ

1.1 Hệ thống tin và cấu trúc đại số sinh từ quan hệ tương đương và loại trừ

Định nghĩa 1.1.1. Một hệ thống thông tin là một cấu trúc dạng:

$K(X) = \langle X, Att(X), val(X), F \rangle$ trong đó:

- X (gọi là tập vũ trụ) là một tập khác rỗng các đối tượng (các tình huống, thực thể, trạng thái)
- $Att(X)$ là tập khác rỗng các thuộc tính, đó là các giá trị của các đối tượng thuộc X
- $val(X)$ là tập các giá trị có thể có mà được quan sát đối với một đối tượng a từ $Att(X)$ trong trường hợp một đối tượng x từ X
- F gọi là hàm thông tin, là 1 ánh xạ $F : X \times Att(X) \longrightarrow val(X)$.

Định nghĩa 1.1.2. Không gian tương tự là cấu trúc dạng $S = \langle X, R \rangle$, trong đó X (gọi là vũ trụ của không gian) là một tập các đối tượng, $X \neq \emptyset$ và R (gọi là quan hệ tương tự của không gian) là quan hệ hai ngôi có tính đối xứng và phản xạ được xác định trên X . Nói cách khác:

$$\forall x \in X : xRx \quad (\text{phản xạ})$$

$$\forall x, y \in X : xRy \Rightarrow yRx \quad (\text{đối xứng})$$

Theo một cách hình thức hơn là:

$\forall x, y \in X : xR_D y$ nếu và chỉ nếu $\forall a_i \in D \subseteq Att(X)$, hoặc

$$F(x, a_i) = F(y, a_i) \quad \text{hoặc} \quad F(x, a_i) = * \quad \text{hoặc} \quad F(y, a_i) = * \quad (1.1)$$

Định nghĩa 1.1.3. Cho một không gian tương tự $\langle X, R \rangle$ và một tập các đối tượng $A \subseteq X$, xấp xỉ thô của A theo tính tương tự được định nghĩa là một cặp $\langle L_R(A), U_R(A) \rangle$, trong đó:

$$\begin{aligned} L_R(A) &:= \{x \in X : S(x) \subseteq A\} = \{x \in X : \forall z (xRz \Rightarrow z \in A)\} \\ U_R(A) &:= \{x \in X : S(x) \cap A \neq \emptyset\} = \{x \in X : \exists z (xRz \text{ và } z \in A)\} \end{aligned} \quad (1.2)$$

Từ đó, ta có:

$$L_R(A) \subseteq A \subseteq U_R(A) \quad (1.3)$$

$L_R(A)$ được gọi là xấp xỉ dưới tương tự của A và $U_R(A)$ được gọi là xấp xỉ trên tương tự của A.

Định nghĩa 1.1.4. Một không gian loại trừ là một cấu trúc dạng $S = \langle X, \# \rangle$, trong đó X (gọi là vũ trụ của không gian) là một tập khác rỗng và $\#$ (gọi là quan hệ loại trừ của không gian) là một quan hệ có tính phản phản xạ và đối xứng trên X. Nói cách khác:

- (i) $\forall x \in X : \text{not } x \# x$ (phản phản xạ)
- (ii) $\forall x, y \in X : x \# y \Rightarrow y \# x$ (đối xứng)

Mệnh đề 1.1.1. Cho $S = \langle X, \# \rangle$ là một không gian loại trừ và xét cấu trúc $\langle P(X), \cap, \cup, ^c, \#, \emptyset, X \rangle$. Khi đó ta có:

1. Cấu trúc con $\langle P(X), \cap, \cup, \#, \emptyset, X \rangle$ là một dàn phân phối (đầy đủ), đối với phép giao $\cap$ và phép hợp $\cup$, bị chặn bởi phần tử nhỏ nhất là $\emptyset$ và phần tử lớn nhất là X. Quan hệ thứ tự bộ phận cảm sinh từ cấu trúc dàn này gọi là quan hệ bao hàm $\subseteq$.

2. Phép toán $^c : P(X) \longrightarrow P(X)$ biến một tập con bất kỳ H của X thành phần bù $H^c = X \setminus H$ là phần bù chuẩn, nghĩa là thỏa mãn:

- (C-1) $H = (H^c)^c$ (tính đối lập)
- (C-2a) $H \subseteq K$ nếu và chỉ nếu $K^c \subseteq H^c$ (tính phản đảo)
- (C-2b) $H^c \cap K^c = (H \cup K)^c$ (tính $\cap$ de Morgan)
- (C-2c) $H^c \cup K^c = (H \cap K)^c$ (tính $\cup$ de Morgan)
- (C-3a) $H \cap H^c = \emptyset$ (tính phi mâu thuẫn)
- (C-3b) $H \cup H^c = X$ (tính bài trung)

3. Phép toán $\# : P(X) \longrightarrow P(X)$ biến một tập con H của X thành phần bù loại trừ $H^\#$ của nó là phần bù không thông dụng và thỏa mãn:

- (B1) $H \subseteq (H^\#)^\#$ (tính phủ định kép yếu)
 (B2a) Nếu $H \subseteq K$ thì $K^\# \subseteq H^\#$ (tính phản đảo)
 (B2b) $H^\# \cap K^\# = (H \cup K)^\#$ (tính $\cap$ de Morgan đối với $\#$)
 (B3) $H \cap H^\# = \emptyset$ (tính phi mâu thuẫn)

4. Quy tắc kết nối trong: $H^\# \subseteq H^c$

Trong trường hợp đặc biệt của quan hệ tương tự (1.1), quan hệ loại trừ tương ứng là phủ định logic của nó:

$\forall x, y \in X \quad x \#_D y$ nếu và chỉ nếu $\text{not } x R_D y$ nếu và chỉ nếu

$$\exists a_i \in D \subseteq \text{Att}(X) : F(x, a_i) \neq F(y, a_i) \text{ và } F(x, a_i) \neq * \text{ và } F(y, a_i) \neq * \quad (1.4)$$

Mệnh đề 1.1.2. Cho $\langle P(X), \cap, \cup, ^c, \#, \emptyset, X \rangle$ là cấu trúc đại số xét trên tập lũy thừa của X và sinh bởi không gian loại trừ $\langle X, \# \rangle$. Khi đó ánh xạ:

$L_\# : P(X) \rightarrow P(X), H \rightarrow L_\#(H) := H^{c\#\#^c}$, là một phép toán phần trong, nghĩa là:

- (I₀) $X = L_\#(X)$ (tính chuẩn hoá)
 (I₁) $L_\#(H) \subseteq H$ (tính giảm)
 (I₂) $L_\#(H) = L_\#(L_\#(H))$ (tính lũy đẳng)
 (I₃) $L_\#(H \cap K) \subseteq L_\#(H) \cap L_\#(K)$ (nhân tính)

Mệnh đề 1.1.3. Cho $\langle P(X), \cap, \cup, ^c, \#, \emptyset, X \rangle$ là cấu trúc đại số được sinh bởi không gian loại trừ $\langle X, \# \rangle$ và $\#$ là phần bù loại trừ trên $P(X)$. Khi đó ánh xạ:

$U_\# : P(X) \rightarrow P(X), H \rightarrow U_\#(H) := H^{\#\#}$, là một phép toán bao đóng, nghĩa là:

- (C₀) $\emptyset = U_\#(\emptyset)$ (tính chuẩn hoá)
 (C₁) $H \subseteq U_\#(H)$ (tính tăng)
 (C₂) $U_\#(H) = U_\#(U_\#(H))$ (tính lũy đẳng)
 (C₃) $U_\#(H) \cup U_\#(K) \subseteq U_\#(H \cup K)$ (cộng tính)

Tập hợp của tất cả các tập $\#$ - mở được định nghĩa như sau:

$$O(X, \#) := \{A \subseteq X : A = L_\#(A) = A^{c\#\#^c}\}$$

Tập hợp tất cả các tập #- đóng được định nghĩa như sau:

$$C(X, \#) := \{B \subseteq X : B = U_{\#}(B) = B^{\#\#}\}$$

Tập hợp tất cả các tập #- đóng-mở được định nghĩa như sau:

$$CO(X, \#) = C(X, \#) \cap O(X, \#)$$

Do tính tăng (C_1) của phép toán bao đóng ta có:

$$L_{\#}(H) \subseteq H \subseteq U_{\#}(H) \quad (1.5)$$

Xấp xỉ trên loại trừ của tập H có thể được biểu diễn:

$$U_{\#}(H) = \cap \{B \in C(X, \#) : H \subseteq B\} \quad (1.6)$$

Xấp xỉ dưới loại trừ của H có thể được biểu diễn:

$$L_{\#}(H) = \cup \{B \in C(X, \#) : B \subseteq H\} \quad (1.7)$$

Như có thể được thấy trong mọi trường hợp đặc biệt, dãy bao hàm sau đây là đúng.

$$L_R(H) \subseteq L_{\#}(H) \subseteq H \subseteq U_{\#}(H) \subseteq U_R(H) \quad (1.8)$$

1.2 Tính đơn điệu tăng của tri thức theo thời gian

Định nghĩa 1.2.1. Cho $K^{(t_0)}(X)$ và $K^{(t_1)}(X)$ với $t_0, t_1 \in R$ là hai hệ thông tin không đầy đủ dựa trên cùng bộ $\langle X, Att(X), val(X) \rangle$ và được đặc trưng bởi hai hàm thông tin khác nhau: $K^{(t_0)} : X \times Att(X) \longrightarrow val(X)$ và $K^{(t_1)} : X \times Att(X) \longrightarrow val(X)$. Ta nói rằng, có sự đơn điệu tăng của thông tin nếu $t_0 < t_1$ và $\forall(x, a)(F^{(t_0)}(x, a) \neq * \Rightarrow F^{(t_1)}(x, a) = F^{(t_0)}(x, a))$. Trong trường hợp như thế, ta sẽ viết $K^{(t_0)}(X) \leq K^{(t_1)}(X)$.

Định nghĩa 1.2.2. Cho $K^{(t_0)}(X)$ và $K^{(t_1)}(X)$ với $t_0, t_1 \in R$ là hai hệ thông tin không đầy đủ và $K^{(t_0)}(X) \leq K^{(t_1)}(X)$. Ta nói rằng có sự đơn điệu tăng của tri thức nếu $C^{(t_0)}(X) \subseteq C^{(t_1)}(X)$.

Mệnh đề 1.2.1. Cho $K^{(t_0)}(X)$ và $K^{(t_1)}(X)$ là hai hệ thông tin không đầy đủ và $K^{(t_0)}(X) \leq K^{(t_1)}(X)$ được đặc trưng bởi phép đơn điệu tăng của tri thức. Khi đó:

$$\forall H \subseteq X, L_{\#}^{(t_0)}(H) \subseteq L_{\#}^{(t_1)}(H) \subseteq H \subseteq U_{\#}^{(t_1)}(H) \subseteq U_{\#}^{(t_0)}(H) \quad (1.9)$$

Mệnh đề 1.2.2. Cho $K^{(t_0)}(X)$ và $K^{(t_1)}(X)$ là hai hệ thông tin không đầy đủ và $K^{(t_0)}(X) \leq K^{(t_1)}(X)$ được đặc trưng bởi phép đơn điệu tăng của thông tin. Khi đó:

$$\forall H \subseteq X, L_R^{(t_0)}(H) \subseteq L_R^{(t_1)}(H) \subseteq H \subseteq U_R^{(t_1)}(H) \subseteq U_R^{(t_0)}(H) \quad (1.10)$$

Cho K là một hệ thông tin với các thuộc tính giá trị số, nghĩa là $Att(X) \subseteq R$. Khi đó, cho thuộc tính $a \in Att(X)$, định nghĩa được một giả metric cho a là:

$$d_a(x, y) := \frac{|F(x, a) - F(y, a)|}{\max\{F(z, a) : z \in X\} - \min\{F(w, a) : w \in X\}} \quad (1.11)$$

Khi giá trị cố định $\varepsilon \in [0, 1]$ ta có thể xét quan hệ tương tự định nghĩa như sau:

$$xR_D^\varepsilon y \quad \text{nếu và chỉ nếu} \quad d_D(x, y) \leq \varepsilon \quad (1.12)$$

Mệnh đề 1.2.3. Cho $K(X)$ là một hệ thông tin giá trị thực (nghĩa là $val(X) \subseteq R$) và $\varepsilon_1, \varepsilon_2 \in [0, 1]$. Nếu, với quan hệ tương tự (12), ta có:

$$C^{\varepsilon_2}(X) \subseteq C^{\varepsilon_1}(X) \text{ thì } \forall H \subseteq X :$$

$$L_{\#}^{\varepsilon_2}(H) \subseteq L_{\#}^{\varepsilon_1}(H) \subseteq H \subseteq U_{\#}^{\varepsilon_1}(H) \subseteq U_{\#}^{\varepsilon_2}(H).$$

Mệnh đề 1.2.4. Cho $K(X)$ là một hệ thông tin giá trị thực (nghĩa là $val(X) \subseteq R$) và $\varepsilon_1, \varepsilon_2 \in [0, 1]$ với $\varepsilon_1 \leq \varepsilon_2$. Khi đó, với một tập các thuộc tính $D \subseteq Att(X)$ được chọn và quan hệ tương tự R_D^ε được định nghĩa trong phương trình (12) ta có:

$$L_R^{\varepsilon_2}(H) \subseteq L_R^{\varepsilon_1}(H) \subseteq H \subseteq U_R^{\varepsilon_1}(H) \subseteq U_R^{\varepsilon_2}(H) .$$

1.3 Dàn phân phối tựa - BZ

Định nghĩa 1.3.1. Hệ thống $(\Sigma, \wedge, \vee, ', \sim, 0, 1)$ được gọi là dàn phân phối tựa Brouwer-Zadeh nếu thỏa các điều sau:

1. Σ là một dàn phân phối với các phép toán "hợp" và "giao" $\vee, \wedge$ mà quan hệ thứ tự bộ phận cảm sinh là $a \leq b$ nếu và chỉ nếu $a = a \wedge b$ (tương đương $b = a \vee b$). Ngoài ra, Σ bị chặn bởi phần tử nhỏ nhất là 0 và phần tử lớn nhất là 1: $\forall a \in \Sigma, 0 \leq a \leq 1$.

2. Phép toán một ngôi $' : \Sigma \rightarrow \Sigma$ là phần bù Kleene (Zadeh hoặc mờ), nghĩa là thỏa mãn $\forall a, b \in \Sigma$,

$$(K1) \quad a'' = a$$

$$(K2) \quad (a \vee b)' = a' \wedge b'$$

$$(K3) \quad a \wedge a' \leq b \vee b'$$

3. Phép toán một ngôi $\sim: \Sigma \rightarrow \Sigma$ là phần bù Brouwer (hoặc trực giác), nghĩa là thỏa mãn $\forall a, b \in \Sigma$,

$$(B1) \quad a \sim \sim = a$$

$$(B2) \quad (a \vee b)^\sim = a^\sim \wedge b^\sim$$

$$(B3) \quad a \wedge a^\sim = 0$$

4. Hai phép bù được liên kết bởi nguyên tắc nối trong, nghĩa là thỏa mãn $\forall a \in \Sigma$,

$$(in) \quad a^\sim \leq a'$$

Mệnh đề 1.3.1. Cho $(\Sigma, \wedge, \vee, ', \sim, 0, 1)$ được gọi là dàn phân phối tựa BZ. Khi đó, ánh xạ $i: \Sigma \rightarrow \Sigma$ mà $i(a) := a^{bb} = a'^{\sim}$ là phép toán phần trong, nghĩa là thỏa mãn:

$$(I_0) \quad 1 = i(1) \quad (\text{chuẩn hóa})$$

$$(I_1) \quad i(a) \leq a \quad (\text{giảm})$$

$$(I_2) \quad i(a) = i(i(a)) \quad (\text{lũy đẳng})$$

$$(I_3) \quad i(a \wedge b) \leq i(a) \wedge i(b) \quad (\text{nhân tính})$$

Đối ngẫu, ánh xạ $c: \Sigma \rightarrow \Sigma$ mà $c(a) := a^{\sim\sim}$ là toán tử bao đóng, nghĩa là thỏa mãn:

$$(C_0) \quad 0 = c(0) \quad (\text{chuẩn hóa})$$

$$(C_1) \quad a \leq c(a) \quad (\text{tăng})$$

$$(C_2) \quad c(a) = c(c(a)) \quad (\text{lũy đẳng})$$

$$(C_3) \quad c(a) \vee c(b) \leq c(a \vee b) \quad (\text{cộng tính})$$

Định nghĩa 1.3.2. Cho $(\Sigma, \wedge, \vee, ', \sim, 0, 1)$ là dàn phân phối tựa BZ. Không gian xấp xỉ thô cảm sinh là cấu trúc dạng $\langle \Sigma, O(\Sigma), C(\Sigma), i, c \rangle$ trong đó:

- Σ là tập các phần tử có thể xấp xỉ
- $O(\Sigma) \subseteq \Sigma$ là tập các phần tử có thể xác định trong sao cho 0 và $1 \in O(\Sigma)$
- $C(\Sigma) \subseteq \Sigma$ là tập các phần tử có thể xác định ngoài sao cho 0 và $1 \in C(\Sigma)$
- $i: \Sigma \rightarrow O(\Sigma)$ là ánh xạ xấp xỉ trong
- $c: \Sigma \rightarrow C(\Sigma)$ là ánh xạ xấp xỉ ngoài

Với mọi phần tử $a \in \Sigma$, xấp xỉ thô được định nghĩa là một cặp:

$$r(a) := \langle i(a), c(a) \rangle \text{ (với } i(a) \leq a \leq c(a) \text{)}$$

Mệnh đề 1.3.2. Trong dàn phân phối tựa BZ bất kỳ, các điều kiện sau là thỏa mãn:

$$\begin{aligned} (\text{mod-1p}) \quad & i(1) = 1 \quad (\text{quy tắc tất yếu}) \\ (\text{mod-2p}) \quad & i(a) \leq a \leq c(a) \quad (\text{nguyên tắc đặc trưng của hệ mô thái } T) \\ (\text{mod-3p}) \quad & i(a) = i(i(a)); c(a) = c(c(a)) \quad (S_4 - \text{nguyên tắc đặc trưng}) \end{aligned}$$

Các phép toán phần trong và bao đóng cho xấp xỉ tốt nhất bởi các phần tử thì chúng không phải là duy nhất. Xét hai ánh xạ sau:

$$\begin{aligned} \nu : \Sigma &\rightarrow C(\Sigma) \quad \nu(a) := a'^{\sim} \\ \mu : \Sigma &\rightarrow O(\Sigma) \quad \mu(a) := a^{\sim'} \end{aligned} \quad (1.13)$$

Mệnh đề 1.3.3. Trong dàn phân phối tựa BZ bất kỳ các điều kiện sau là thỏa mãn:

$$\begin{aligned} (\text{mod-1s}) \quad & \nu(1) = 1 \quad (\text{quy tắc tất yếu}) \\ (\text{mod-2s}) \quad & \nu(a) \leq a \leq \mu(a) \quad (\text{nguyên tắc đặc trưng của hệ mô thái } T) \\ (\text{mod-3s}) \quad & a \leq \nu(\mu(a)) \quad (B - \text{nguyên tắc đặc trưng}) \end{aligned}$$

Thực tế, đối với mọi phần tử $a \in X$, ta có:

$$\nu(a) \leq i(a) \leq a \leq c(a) \leq \mu(a) \quad (1.14)$$

Mệnh đề 1.3.4. Cấu trúc $\langle A, \sqcap, \sqcup, -, \approx, \langle 0, 1 \rangle, \langle 1, 0 \rangle \rangle$ trong đó:

$$\begin{aligned} \langle a_i, a_e \rangle^{\sqcap} \langle b_i, b_e \rangle &:= \langle i(a_i \wedge b_i), a_e \vee b_e \rangle \\ \langle a_i, a_e \rangle^{\sqcup} \langle b_i, b_e \rangle &:= \langle a_i \vee b_i, i(a_e \wedge b_e) \rangle \\ \langle a_i, a_e \rangle^- &:= \langle (a_e, a_i) \\ \langle a_i, a_e \rangle^{\approx} &:= \langle e((a_e)b), (a_e)b \rangle \end{aligned}$$

là một dàn phân phối tựa BZ.

Thứ tự bộ phận cảm sinh ra bởi các toán tử dàn $\sqcap, \sqcup$ là

$$\langle a_i, a_e \rangle \sqsubseteq \langle b_i, b_e \rangle \text{ nếu và chỉ nếu } a_i \leq b_i \text{ và } b_e \leq a_e.$$

Trong bối cảnh này, các phép toán phần trong, bao đóng, phần ngoài lần lượt được định nghĩa như sau:

$$I(\langle a_i, a_e \rangle) := \langle i(a_i), e(i(a_i)) \rangle = \langle a_i, e(a_i) \rangle$$

$$C(\langle a_i, a_e \rangle) := \langle e(i(a_e)), i(a_e) \rangle = \langle e(a_e), a_e \rangle$$

$$E(\langle a_i, a_e \rangle) := \langle a_e, e(a_e) \rangle$$

Tập hợp của các tập mở và đóng lần lượt là

$$O(A) = \{ \langle a_i, e(a_i) \rangle : a \in \Sigma \}$$

$$C(A) = \{ \langle e(b_e), b_e \rangle : b \in \Sigma \}$$

1.4 Dàn BZ và tập thô Pawlak

Cho phần tử $x \in X$, ta có thể xác định lớp tương đương sinh bởi x là tập:

$$E(x) = \{y \in X : xRy\}$$

Cho tập con các thuộc tính $D \subseteq Att(X)$, quan hệ tương đương của hai đối tượng bất kỳ được định nghĩa là:

$$x \equiv y \quad \text{nếu và chỉ nếu} \quad \forall a \in D, F(x, a) = F(y, a) \quad (1.15)$$

Quan hệ loại trừ tương ứng được định nghĩa là

$$x \not\equiv y \quad \text{nếu và chỉ nếu} \quad \exists a \in D, F(x, a) \neq F(y, a) \quad (1.16)$$

Định nghĩa 1.4.1. Cấu trúc $(\Sigma, \wedge, \vee, ', \sim, 0, 1)$ là một dàn phân phối BZ nếu nó là dàn phân phối tựa BZ thỏa mãn nguyên tắc kết nối trong mạnh:

$$(s\text{-in}) \quad a^{\sim\sim} = a^{\sim'}$$

Định nghĩa 1.4.2. Dàn phân phối BZ cũng thỏa mãn tính chất $\wedge$ De Morgan.

$$(B2a) \quad (a \wedge b)^{\sim} = a^{\sim} \vee b^{\sim}$$

được gọi là dàn phân phối BZ De Morgan $(BZ)^{dM}$

Mệnh đề 1.4.1. Cho $(\Sigma, \wedge, \vee, ', \sim, 0, 1)$ là một dàn phân phối BZ. khi đó các điều sau là đúng:

$$\forall a \in \Sigma, \nu(a) = a'^{\sim} = a'^{\sim\sim} = i(a)$$

$$\forall a \in \Sigma, \mu(a) = a^{\sim'} = a^{\sim\sim} = c(a)$$

Định nghĩa 1.4.3. Cho $\langle \Sigma, \leq, ', \sim, 0, 1, \rangle$ là dàn phân phối BZ. Không gian xấp xỉ thô là cấu trúc dạng: $\langle \Sigma, \Sigma_e, \nu, \mu \rangle$ trong đó:

- Σ : là tập các phần tử xấp xỉ được
- $\Sigma_e \subseteq \Sigma$: là tập các phần tử chính xác

- $\nu : \Sigma \rightarrow \Sigma_e$ là ánh xạ xấp xỉ trong
- $\mu : \Sigma \rightarrow \Sigma_e$ là ánh xạ xấp xỉ ngoài

Với bất kỳ $a \in \Sigma$, xấp xỉ thô của nó được định nghĩa là cặp:

$$r(a) := \langle \nu(a), \mu(a) \rangle \text{ [với } \nu(a) \leq a \leq \mu(a)]$$

Mệnh đề 1.4.2. *Cấu trúc $\langle A, \sqcap, \sqcup, -, \approx, \langle 0, 1 \rangle, \langle 1, 0 \rangle \rangle$ trong đó*

$$\langle a_i, a_e \rangle^\sqcap \langle b_i, b_e \rangle := \langle a_i \wedge b_i, a_e \vee b_e \rangle$$

$$\langle a_i, a_e \rangle^\sqcup \langle b_i, b_e \rangle := \langle a_i \vee b_i, a_e \wedge b_e \rangle$$

$$\langle a_i, a_e \rangle - := \langle (a_e, a_i); \quad \langle a_i, a_e \rangle \approx := \langle a_e, (a_e)' \rangle$$

là một dàn phân phối BZ^{dM} .

Các phép toán mô thái về tất yếu $\Box$ và khả năng $\Diamond$ chọn ra được một phần tử chính xác. Nghĩa là, tất yếu của xấp xỉ thô $\langle a_i, a_e \rangle$ bằng xấp xỉ thô tất yếu của a:

$$\Box(r_e(a)) := \Box \langle a_i, (a_e) \rangle = \langle a_i, (a_i)' \rangle = r_e(\nu(a)) \quad (1.17)$$

Một cách đối ngẫu, khả năng của xấp xỉ thô $\langle a_i, (a_e) \rangle$ bằng xấp xỉ thô khả năng của a:

$$\Diamond(r_e(a)) := \Diamond \langle a_i, (a_e) \rangle = \langle (a_e)', a_e \rangle = r_e(\mu(a)) \quad (1.18)$$

Mệnh đề 1.4.3. *Cho tập $\langle X, Att(X), val(X), F \rangle$ là một hệ thông tin đầy đủ. Khi đó, cấu trúc $P = \langle P(X), \cap, \cup, ^c, \neq, \emptyset, X \rangle$ trong đó tập thuộc tính $D \subseteq Att(X)$ cố định, quan hệ loại trừ $\neq$,*

$H^\neq = \{x \in X : \forall y \in H, \exists a \in D, F(x, a) \neq F(y, a)\}$ là một dàn phân phối BZ. Nói chung cấu trúc P không phải là một dàn BZ De Morgan.

Chương 2

NGŨ NGHĨA CỦA TẬP MỜ TRONG LÝ THUYẾT TẬP THÔ

2.1 Khái niệm tập mờ

2.1.1 Hệ thống tập mờ

Cho U là một tập hữu hạn khác rỗng gọi là tập vũ trụ. Một tập con mờ của U được định nghĩa bởi hàm thuộc:

$$\mu_A : U \rightarrow [0, 1] \quad (2.1)$$

Tập mờ μ_A là tập con của tập mờ μ_B , ký hiệu $\mu_A \subseteq \mu_B$ nếu $\mu_A(x) \leq \mu_B(x)$ với $\forall x \in U$.

Tập mờ μ_A bằng tập mờ μ_B , ký hiệu $\mu_A = \mu_B$ nếu $\mu_A(x) = \mu_B(x)$ với $\forall x \in U$. Rõ ràng, $\mu_A = \mu_B$ nếu và chỉ nếu $\mu_A \subseteq \mu_B$ và $\mu_B \subseteq \mu_A$.

Có nhiều định nghĩa về phần bù, giao và hợp của tập mờ. Zadeh đưa ra hệ thống chuẩn min-max mà được cho theo từng thành phần:

$$\mu_{\neg A}(x) = 1 - \mu_A(x)$$

$$\mu_{A \cap B}(x) = \min(\mu_A(x), \mu_B(x))$$

$$\mu_{A \cup B}(x) = \max(\mu_A(x), \mu_B(x)) \quad (2.2)$$

Một t-norm là một hàm từ $[0, 1] \times [0, 1] \rightarrow [0, 1]$ và thỏa mãn các điều kiện sau:
 $\forall a, b, c \in [0, 1]$.

(i) Các điều kiện biên

$$t(0, 0) = 0, t(1, a) = t(a, 1) = a$$

(ii) Tính đơn điệu

$$(a \leq c, b \leq d) \Rightarrow t(a, b) \leq t(c, d)$$

(iii) Tính đối xứng

$$t(a, b) = t(b, a)$$

(4i) Tính kết hợp

$$t(a, t(b, c)) = t(t(a, b), c)$$

Một số t-norm thường sử dụng là $t_b(a, b) = \max(0, a + b - 1)$, $t_{\min}(a, b) = \min(a, b)$, phép toán nhân $t_p(a, b) = ab$ và t_w được xác định bởi điều kiện biên và $t_w(a, b) = 0, \forall (a, b) \in [0, 1) \times [0, 1)$. Các t-norm này có mối quan hệ với bất đẳng thức:

$$t_w(a, b) \leq t_b(a, b) \leq t_p(a, b) \leq t_{\min}(a, b) \quad (2.3)$$

Ngoài ra, t-norm bất kỳ bị chặn bởi t_w và $t_{\min}$ nghĩa là:

$$t_w(a, b) \leq t(a, b) \leq t_{\min}(a, b) \quad (2.4)$$

Giả sử $n: [0, 1] \rightarrow [0, 1]$ là một phép toán gọi là phủ định. Đối với phép phủ định đối ngẫu của t-norm được cho bởi $n(t(n(a), n(b)))$ và được gọi là t-conorm, đó là hàm $s: [0, 1] \times [0, 1] \rightarrow [0, 1]$ và thỏa mãn các điều kiện sau:

(i'). Các điều kiện biên

$$s(1, 1) = 1, s(a, 0) = s(0, a) = a$$

và điều kiện đơn điệu, đối xứng và kết hợp. Giả sử phép phủ định được định nghĩa là $n(a) = 1 - a$. t-conorm tương ứng với t-norm t được cho bởi:

$$s(a, b) = n(t(n(a), n(b))) = 1 - t(1 - a, 1 - b) \quad (2.5)$$

t-conorm của $t_{\min}$ và t_b là $s_{\max}(a, b) = \max(a, b)$, $s_p(a, b) = a + b - ab$ và $s_b(a, b) = \min(1, a + b)$. t -conorm tương ứng t_w được cho bởi các điều kiện biên và $s_w(a, b) = 1, \forall a, b \in [0, 1] \times [0, 1]$. Tương tự, ta có:

$$s_{\max}(a, b) \leq s_p(a, b) \leq s_b(a, b) \leq s_w(a, b) \quad (2.6)$$

t-conorm bất kỳ có biên là $s_{\max}$ và s_w :

$$s_{\max}(a, b) \leq s(a, b) \leq s_w(a, b) \quad (2.7)$$

Kết hợp với phương trình (2.4) ta có

$$t(a, b) \leq \min(a, b) \leq \max(a, b) \leq s(a, b) \quad (2.8)$$

biểu diễn mối quan hệ giữa phép giao và phép hợp tập mờ.

Cho t và s là một cặp t -norm và t -conorm. Ta định nghĩa phép giao và phép hợp tập mờ cho từng thành phần:

$$\begin{aligned}\mu_{A \cap B}(x) &= t(\mu_A(x), \mu_B(x)) \\ \mu_{A \cup B}(x) &= s(\mu_A(x), \mu_B(x))\end{aligned}\tag{2.9}$$

2.1.2 Đặc trưng định tính của tập mờ

So sánh với các điều kiện định lượng trên t -norm và t -conorm, các tính chất định tính dễ dàng giải thích. Do các điều kiện biên và tính đơn điệu, ta có:

$$\mu_{A \cap B} \subset \mu_A \subset \mu_{A \cup B}\tag{2.10}$$

mà tương ứng với điều kiện định lượng trong (2.8)

Lõi của tập mờ là một tập con rõ của U gồm các phần tử với giá trị thuộc đầy:

$$Core(\mu_A) = \{x \in U \mid \mu_A(x) = 1\}\tag{2.11}$$

Giá là một tập con rõ của U gồm các phần tử có giá trị thuộc khác rỗng.

$$Support(\mu_A) = \{x \in U \mid \mu_A(x) > 0\}\tag{2.12}$$

2.2 Một khung ngữ nghĩa của tập mờ

2.3 Hàm thuộc thô cổ điển và xấp xỉ tập thô

2.3.1 Hàm thuộc thô

Cho $R \subseteq U \times U$ là quan hệ tương đương trên vũ trụ U , với U là tập hữu hạn khác rỗng. Nghĩa là, R có tính phản xạ, đối xứng và bắc cầu. Quan hệ R cảm sinh một phân hoạch U/R của vũ trụ U , đó là một họ các tập con rời nhau của U được biết là các lớp tương đương. Quan hệ tương đương biểu diễn tri thức sẵn có về tập vũ trụ.

Cặp $\text{apr} = (U, R)$ được gọi là một không gian xấp xỉ. Phần tử $x \in U$ thuộc vào một và chỉ một lớp tương đương. Ký hiệu:

$$[x]_R = \{y \mid xRy\}\tag{2.13}$$

Nghĩa là lớp tương đương chứa x . Với tập con $A \subseteq U$, định nghĩa hàm thuộc thô:

$$\mu_A(x) = \frac{|[x]R \cap A|}{|[x]R|} \quad (2.14)$$

trong đó $|\cdot|$ là bản số của một tập hợp. Giá trị thuộc thô $\mu_A(x)$ có thể được hiểu là xác suất có điều kiện. Tập A được gọi là tập sinh của giá trị thuộc thô μ_A .

Công thức sau để tính toán hàm thuộc của tập:

$$\mu_A(x) = \Sigma \mu_{\{y\}}(x), \forall y \in A \quad (2.15)$$

2.3.2 Xấp xỉ thô và các phép toán xấp xỉ

Trong một không gian xấp xỉ, một tập con $A \subseteq U$ được xấp xỉ bởi một cặp tập hợp gọi là xấp xỉ trên và xấp xỉ dưới:

$$\begin{aligned} \underline{apr}(A) &= \{x \in U | \mu_A(x) = 1\} \\ &= Core(\mu_A) \\ \overline{apr}(A) &= \{x \in U | \mu_A(x) > 0\} \\ &= Support(\mu_A) \end{aligned} \quad (2.16)$$

Thực ra chúng là lõi và giá của tập mờ μ_A .

Các xấp xỉ tập thô có thể được biểu diễn lại theo các dạng tương đương sau:

$$\begin{aligned} \underline{apr}(A) &= \{x \in U | [x]_R \subseteq A\} \\ &= \{x \in U | \forall y \in U (xRy \Rightarrow y \in A)\} \\ &= \cup \{[x]_R | [x]_R \subseteq A\} \\ \overline{apr}(A) &= \{x \in U | [x]_R \cap A \neq \emptyset\} \\ &= \{x \in U | \exists y \in U (xRy \Rightarrow y \in A)\} \\ &= \cup \{[x]_R | [x]_R \cap A \neq \emptyset\} \end{aligned} \quad (2.17)$$

2.3.3 Các đặc điểm chính của tập thô dựa trên các ngữ nghĩa của tập mờ

2.4 Tập thô được tổng quát hóa bằng quan hệ không tương đương

2.4.1 Quan hệ hai ngôi

Cho $R \subseteq U \times U$ là một quan hệ hai ngôi trên tập vũ trụ hữu hạn và khác rỗng U . Với hai phần tử $x, y \in R$ nếu xRy , chúng ta nói rằng x là R -quan hệ với y . Ký hiệu:

$$\begin{aligned} R_s(x) &= \{y \in Y | xRy\}, \\ R_p(x) &= \{y \in Y | yRx\} \end{aligned} \quad (2.18)$$

lần lượt là các lân cận liền sau và lân cận liền trước của x cả sinh bởi quan hệ hai ngôi R . Từ quan hệ hai ngôi R , ta có thể định nghĩa bốn quan hệ hai ngôi:

$$\begin{aligned} x \equiv_R y &\Leftrightarrow R_s(x) = R_s(y) \\ x \approx_R y &\Leftrightarrow R_s(x) \cap R_s(y) \neq \emptyset \\ x \simeq_R y &\Leftrightarrow R_p(x) = R_p(y) \\ x \sim_R y &\Leftrightarrow R_s(x) \cap R_s(y) \neq \emptyset \end{aligned} \quad (2.19)$$

2.4.2 Các hàm thuộc thô

Cặp $\text{apr} = (U, R)$ được gọi là một không gian xấp xỉ, với quan hệ R biểu diễn mối quan hệ giữa các phần tử của U . Ta có thể định nghĩa hàm thuộc thô bằng cách mở rộng phương trình (2.14) bằng việc sử dụng các lân cận cả sinh bởi quan hệ hai ngôi. Xét định nghĩa dựa trên các lân cận liền sau với tập con $A \subseteq U$, một hàm thuộc thô có thể được xác định bằng phép thay thế $[x]_R$ với $R_s(x)$ trong phương trình (2.14) như sau:

$$\mu_A(x) = \frac{|R_s(x) \cap A|}{|R_s(X)|} \quad (2.20)$$

2.4.3 Xấp xỉ tập thô và các phép toán xấp xỉ

Tương tự trường hợp cổ điển, một cặp các xấp xỉ tập thô có thể được xác định bởi lõi và giá của μ_A . Nghĩa là, với một tập $A \subseteq U$, ta có:

$$\begin{aligned}
 \underline{apr}(A) &= Core(\mu_A) \\
 &= \{x \in U | \mu_A(x) = 1\} \\
 &= \{x \in U | R_s(x) \subseteq A\} \\
 &= \{x \in U | \forall y \in U (xRy \Rightarrow y \in A)\} \\
 \overline{apr}(A) &= Support(\mu_A) \\
 &= \{x \in U | \mu_A(x) > 0\} \\
 &= \{x \in U | R_s(x) \cap A \neq \emptyset\} \\
 &= \{x \in U | \exists y \in U (xRy, y \in A)\}
 \end{aligned} \tag{2.21}$$

2.4.4 Lớp của các mô hình tập thô

Mô hình phản xạ.

Mô hình đối xứng.

Mô hình bắc cầu.

Đối với quan hệ bắc cầu R , với mọi $x, y \in U$, nếu xRy thì $R_s(y) \subseteq R_s(x)$. Khi đó ta có kết nối giữa R và $\approx_R$ và $\sim_R$:

$$\begin{aligned}
 R \text{ là nối tiếp và bắc cầu} &\Rightarrow [xRy \Rightarrow x \approx_R y] \\
 R \text{ là nối tiếp đảo và bắc cầu} &\Rightarrow [xRy \Rightarrow x \sim_R y]
 \end{aligned} \tag{2.22}$$

Mô hình Euclide.

Đối với quan hệ Euclide, với mọi $x, y \in U$, nếu xRy thì $R_s(x) \subseteq R_s(y)$. Trong trường hợp đó ta có:

$$\begin{aligned}
 R \text{ là Euclide} &\Rightarrow [xRy \Rightarrow x \approx_R y] \\
 R \text{ là nối tiếp đảo và Euclide} &\Rightarrow [xRy \Rightarrow x \sim_R y]
 \end{aligned} \tag{2.23}$$

Mô hình Pawlak.

2.5 Tập thô được tổng quát hóa dựa trên phủ của tập vũ trụ

2.5.1 Phủ

Một phủ của vũ trụ, $\mathfrak{C} = \{C_1, C_2, \dots, C_n\}$ là họ các tập con của U sao cho $U = \cup\{C_i, i = 1, \dots, n\}$. Hai tập phân biệt trong $\mathfrak{C}$ có thể có giao khác rỗng. Một phần tử có thể có nhiều hơn một lớp trong $\mathfrak{C}$. Họ $\mathfrak{C}(x) = \{C \in \mathfrak{C} | x \in C\}$ gồm các tập trong $\mathfrak{C}$ chứa x . Đối với một phủ, ta xây dựng quan hệ tương đương như sau:

$$x \equiv_C y \Leftrightarrow \forall C \in \mathfrak{C} (x \in C \Leftrightarrow y \in C) \Leftrightarrow \mathfrak{C}(x) = \mathfrak{C}(y) \quad (2.24)$$

Hai phần tử được xem là tương đương nếu chúng xuất hiện trong cùng họ của các tập con trong $\mathfrak{C}$. Một quan hệ tương tự $\sim_C$ được định nghĩa như sau:

$$x \sim_C y \Leftrightarrow \exists C \in \mathfrak{C} (x \in C \Leftrightarrow y \in C) \Leftrightarrow \mathfrak{C}(x) \cap \mathfrak{C}(y) \neq \emptyset \quad (2.25)$$

Nghĩa là, x và y được xem là tương tự nếu chúng xuất hiện đồng thời trong ít nhất một lớp của phủ vũ trụ.

2.5.2 Hàm thuộc thô

Các tập trong $\mathfrak{C}(x)$ mô tả các kiểu khác nhau hoặc bậc tương tự khác nhau của các phần tử trong U . Với một tập $C \in \mathfrak{C}(x)$, ta có:

(nhỏ nhất)

$$\underline{\mu}_A(x) = \min\left\{\frac{|C \cap A|}{|C|} \mid x \in C, C \in \mathfrak{C}\right\} \quad (2.26)$$

(lớn nhất)

$$\overline{\mu}_A(x) = \max\left\{\frac{|C \cap A|}{|C|} \mid x \in C, C \in \mathfrak{C}\right\} \quad (2.27)$$

(trung bình)

$$\mu_A^*(x) = \text{avg}\left\{\frac{|C_i \cap A|}{|C_i|} \mid x \in C, C \in \mathfrak{C}\right\} \quad (2.28)$$

Ta có tính chất sau:

$$\mu_{A \cap B}^*(x) = \mu_A^*(x) + \mu_B^*(x) - \mu_{A \cup B}^*(x) \quad (2.29)$$

2.5.3 Xấp xỉ thô và các phép toán xấp xỉ

Từ ba hàm thuộc thô, ta định nghĩa ba cặp xấp xỉ trên và xấp xỉ dưới. Đối với định nghĩa nhỏ nhất, ta có:

$$\begin{aligned} \underline{apr}^m(A) &= Core(\underline{\mu}_A) \\ &= \{x \in U \mid \underline{\mu}_A(x) = 1\} \\ &= \{x \in U \mid \forall C \in \mathfrak{C}, (x \in C \Rightarrow C \subseteq A)\} \\ \overline{apr}^m(A) &= Support(\underline{\mu}_A) \\ &= \{x \in U \mid \underline{\mu}_A(x) > 0\} \\ &= \{x \in U \mid \forall C \in \mathfrak{C}, (x \in C \Rightarrow C \cap A \neq \emptyset)\} \end{aligned} \quad (2.30)$$

Đối với định nghĩa lớn nhất, ta có:

$$\begin{aligned} \underline{apr}^M(A) &= Core(\overline{\mu}_A) \\ &= \{x \in U \mid \overline{\mu}_A(x) = 1\} \\ &= \{x \in U \mid \exists C \in \mathfrak{C}, (x \in C \Rightarrow C \subseteq A)\} \\ &= \cup \{C \in \mathfrak{C} \mid C \subseteq A\} \\ \overline{apr}^M(A) &= Support(\overline{\mu}_A) \\ &= \{x \in U \mid \overline{\mu}_A(x) > 0\} \\ &= \{x \in U \mid \exists C \in \mathfrak{C}, (x \in C \Rightarrow C \cap A \neq \emptyset)\} \\ &= \cup \{C \in \mathfrak{C} \mid C \cap A \neq \emptyset\} \end{aligned} \quad (2.31)$$

Các xấp xỉ trên và xấp xỉ dưới trong mỗi cặp không còn các phép toán đối ngẫu. Tuy nhiên, $(\underline{apr}^m, \overline{apr}^M)$ và $(\underline{apr}^M, \overline{apr}^m)$ là hai cặp phép toán đối ngẫu. Cặp thứ nhất có thể được dẫn xuất từ định nghĩa trung bình, cụ thể là:

$$\underline{apr}^*(A) = \underline{apr}^m(A) \quad \overline{apr}^*(A) = \overline{apr}^M(A) \quad (2.32)$$

Các phép toán xấp xỉ này được nghiên cứu rộng rãi trong lý thuyết tập thô.

Chương 3

CÁC ỨNG DỤNG ĐẶC TRƯNG CỦA LÝ THUYẾT TẬP THÔ

3.1 Khám phá tri thức theo tiếp cận tập thô

Cho hệ thông tin đầy đủ $K(U) = \langle U, Att(U), val(U), F \rangle$. Với tập con bất kỳ $B \subseteq A = Att(U)$, tồn tại quan hệ tương đương, ký hiệu là $Ind_{K(U)}(B)$ được xác định như sau:

$$Ind_{K(U)}(B) = \{(x, x') \in U \times U \mid F(x, a) = F(x', a), \forall a \in B\}.$$

$Ind_{K(U)}(B)$ được gọi là quan hệ không phân biệt được theo B. Ký hiệu $[x]_B$ là lớp tương đương của x theo quan hệ tương đương này. Ở đây, ta còn viết $a(x) = F(x, a)$, $a(x') = F(x', a)$ và $Ind(B)$ thay cho $Ind_{K(U)}(B)$ khi $K(U)$ đã được hoàn toàn xác định.

Bây giờ, cho $B \subseteq A$ và $X \subseteq U$. Các tập xấp xỉ của X theo thông tin có từ B, được xác định như dưới đây:

1. Tập B - xấp xỉ dưới của X, ký hiệu $\underline{BX}$ là tập $\underline{BX} = \{x \mid [x]_B \subseteq X\}$
2. Tập B - xấp xỉ trên của X, ký hiệu $\overline{BX}$ là tập $\overline{BX} = \{x \mid [x]_B \cap X \neq \emptyset\}$

Đối tượng trong $\underline{BX}$ chắc chắn được phân lớp là thành viên của X theo tri thức cơ sở từ B, tập $\underline{BX}$ còn gọi là tập chắc chắn, trong khi đối tượng trong $\overline{BX}$ chỉ có khả năng được phân lớp là thành viên của X theo tri thức cơ sở trong B, tập $\overline{BX}$ có thể được gọi là tập khả năng.

Tập $BN_B(X) = \overline{BX} \setminus \underline{BX}$ được gọi là B - vùng biên của X. Tập $U \setminus \overline{BX}$ được gọi là B - vùng ngoài của X bao gồm các đối tượng chắc chắn không thuộc X. Một tập được gọi là thô hoàn toàn nếu vùng biên của nó là khác rỗng.

3.1.1 Tính phụ thuộc thuộc tính trong hệ thông tin

Giả sử D và C là các tập con của A , ta nói rằng D phụ thuộc vào C với mức k ($k \in [0, \dots, 1]$) biểu thị $C \Rightarrow_k D$ nếu:

$$k = \gamma(C, D) = \frac{|Pos_C(D)|}{|U|}, \text{ với } Pos_C(D) = \bigcup_{x \in U/D} \underline{C}(X)$$

được gọi là một C - vùng dương của phân hoạch U/D đối với C , là tập tất cả các phần tử của U mà có thể được phân loại duy nhất thành khối của phân hoạch U/D với ý nghĩa của C .

$$\gamma(C, D) = \sum_{x \in U/D} \frac{|\underline{C}(X)|}{|U|}$$

Nếu $k=1$ ta nói rằng D phụ thuộc hoàn toàn vào C và nếu $k < 1$, ta nói rằng D phụ thuộc một phần vào C .

Hệ số k diễn tả tỉ lệ của các thành phần trong tập tổng thể, với sự phân loại thành khối của phân hoạch U/D , các thuộc tính sử dụng trong C gọi là mức phụ thuộc.

3.1.2 Tập thuộc tính rút gọn và tập thuộc tính lõi

3.1.3 Ma trận phân biệt được và hàm phân biệt được

Ma trận phân biệt được của A ký hiệu là $M(A)$ là một ma trận đối xứng $n \times n$, với các phần tử c_{ij} cho như sau:

$$c_{ij} := \begin{cases} \{a \in A : a(x_i) \neq a(x_j)\} & \text{nếu } \exists d \in D [d(x_i) \neq d(x_j)] \\ \lambda & \text{nếu } \forall d \in D [d(x_i) = d(x_j)] \end{cases}$$

c_{ij} là tập tất cả các thuộc tính điều kiện mà phân loại x_i, x_j thành các lớp khác nhau.

Hàm phân biệt được f_A cho một hệ thông tin A là một hàm boole của m biến logic $a_1^*, a_2^*, \dots, a_m^*$ (tương ứng với các thuộc tính $a_1, a_2, \dots, a_m$) được xác định như sau với $c_{ij} = \{a^* \mid a \in c_{ij}\}$

$$f_A(a_1^*, a_2^*, \dots, a_m^*) = \bigwedge \{ \bigvee c_{ij}^* \mid 1 \leq j \leq i \leq n, c_{ij} \neq \emptyset \} \text{ với}$$

$$\bigvee c_{ij} = \perp \text{ (false) nếu } c_{ij} \neq \emptyset$$

$$\bigvee c_{ij} = t(true) \text{ nếu } c_{ij} = \lambda$$

3.1.4 Quá trình khám phá tri thức theo cách tiếp cận thô

3.2 Vai trò của tập thô trong bài toán nhận dạng mặt người

3.2.1 Giới thiệu bài toán

Hiện nay, những công nghệ hiện đại đã cho phép việc xác thực dựa vào "bản chất" của từng cá nhân. Công nghệ này dựa trên lĩnh vực được gọi là sinh trắc học. Kiểm soát sinh trắc học là những phương pháp tự động cho phép xác thực hay nhận dạng một cá nhân dựa vào các đặc trưng sinh lý học của người đó như đặc điểm vân tay, gương mặt, gen,... hoặc dựa trên những đặc điểm liên quan đến đặc trưng hành vi như dạng chữ viết, cách gõ phím, giọng nói,... Vì những hệ thống nhận dạng bằng trắc học sử dụng thông tin sinh học của con người nên kết quả chính xác và đặc biệt là rất khó bị giả mạo.

Nhận dạng khuôn mặt là một trong ít các phương pháp nhận dạng dựa vào đặc trưng sinh lý cho kết quả chính xác cao đồng thời rất thuận tiện khi sử dụng. Khả năng nhận dạng nói chung và khả năng nhận biết khuôn mặt người nói riêng của con người thật đáng kinh ngạc. Do đó, việc nghiên cứu các đặc tính của gương mặt người đã thu hút rất nhiều nhà triết học, nhà khoa học qua nhiều thế kỷ.

3.2.2 Mô hình nhận dạng mặt người tiêu biểu

3.2.3 Rút trích đặc trưng

Giả sử rằng mỗi bức ảnh gương mặt được thể hiện dưới dạng một ma trận số hai chiều các giá trị điểm ảnh, hay mỗi ảnh được viết dưới dạng một vectơ $X = \{x_i \in S\}$ với S là lưới vuông đại diện cho lưới ảnh. Đôi khi người ta sử dụng X dưới dạng vectơ một chiều theo từng dòng của bức ảnh: $X = [x_1 x_2 \dots x_N]^T$ với N là tổng số điểm ảnh. Như vậy, với một ảnh kích thước 320×240 , kích thước ảnh là $N = 76800$. Một vectơ có số chiều lớn như vậy thường không hiệu quả trong tính toán và hạn chế khả năng nhận dạng. Chính vì vậy người ta đã đưa ra nhiều phương pháp nhằm đưa vectơ X về một vectơ đặc trưng $f(X) = [f_1(X) f_2(X) \dots f_M(X)]^T$ trong đó f_i với $i = 1, 2, \dots, M$ có thể là các

hàm tuyến tính hoặc phi tuyến. Trong đa số trường hợp, nhằm làm tăng hiệu quả tính toán, kích thước vectơ đặc trưng n thường nhỏ hơn rất nhiều kích thước của vectơ ban đầu N .

3.2.4 Ứng dụng tập thô trong lựa chọn đặc trưng

3.2.4.1. Phương pháp chung

3.2.4.2. Kết hợp heuristic và lý thuyết tập thô

Các thuộc tính được chọn để phát sinh ra các luật tuân theo chiến lược nêu sau đây:

1. Để nhận được tập thuộc tính nhỏ nhất có thể, ta ưu tiên chọn thuộc tính a_0 mà việc thêm nó vào tập rút gọn R hiện có sẽ làm cho số lượng đối tượng bền vững tăng lên nhanh nhất, tức là:

$$a_0 = \arg \max Pos_{R \cup \{a\}}(D)$$

2. Khi thêm thuộc tính a_0 và R , tập các phân hoạch các đối tượng bền vững theo tập thuộc tính được chọn, tức là tập hợp $Pos_{R \cup \{a\}}(D) | Ind(R \cup \{a\} \cup (D))$ sẽ thay đổi, từ đó làm thay đổi tập các luật phát sinh. Trong các lớp tương đương thuộc tập phân hoạch mới, giả sử M là lớp có nhiều phần tử nhất và r là luật được phát sinh tương ứng với tập các đối tượng M . Ta nhận xét rằng, kích thước của tập M càng lớn bao nhiêu thì tính bao phủ của luật r càng lớn bấy nhiêu, cụ thể hơn là số lượng đối tượng thỏa mãn r càng lớn. Như vậy, ta có thể lấy kích thước của M như là tiêu chuẩn thứ hai trong lựa chọn thuộc tính.

Tóm lại, ta sử dụng hai chỉ số sau:

- Số lượng đối tượng bền vững: $\nu_a = card(Pos_{R \cup \{a\}}(D))$
- Kích thước lớp tương đương lớn nhất: $m_a = \max-size(Pos_{R \cup \{a\}}(D) | Ind(R \cup \{a\} \cup (D)))$ trong đó a là thuộc tính chưa được chọn: $a \in C \setminus R$

Hai chỉ số trên có xu hướng cạnh tranh với nhau, do đó ta sử dụng tích của chúng làm tiêu chuẩn cuối cùng để chọn thuộc tính.

KẾT LUẬN

Qua việc tìm hiểu và nghiên cứu các khái niệm và kết quả về cấu trúc đại số, ngữ nghĩa tập mờ trong lý thuyết tập thô và các ứng dụng của lý thuyết tập thô, luận văn đã đạt được một số kết quả chính sau:

1. Bằng việc sử dụng hệ thông tin không đầy đủ với sự hỗ trợ một tập các đối tượng X , chúng tôi tìm hiểu và nghiên cứu một đại số hóa của đại số cụ thể của tập lũy thừa của X thông qua những dàn tựa BZ. Cấu trúc này cho phép xác định hai xấp xỉ thô dựa trên quan hệ tương tự và quan hệ loại trừ, với quan hệ loại trừ luôn tốt hơn quan hệ tương tự.

Đặc biệt, khi thông tin là đầy đủ thì quan hệ hai ngôi được sử dụng trở thành quan hệ tương đương và ta nhận được tập thô cổ điển Pawlak. Trong trường hợp này, chúng tôi tìm hiểu những không gian xấp xỉ thô dưới dạng những dàn BZ.

2. Tìm hiểu mối quan hệ giữa lý thuyết tập mờ và lý thuyết tập thô qua việc nhận được một khung ngữ nghĩa đối với tập mờ theo lý thuyết tập thô. Ở đây, hàm thuộc thô được xem như là một kiểu đặc biệt của hàm thuộc mờ mà có thể giải thích bằng cách dùng xác suất có điều kiện. Mối quan hệ giữa những hàm thuộc mờ và những hàm thuộc thô, giữa lõi và giá của lý thuyết tập mờ và xấp xỉ dưới, xấp xỉ trên của lý thuyết tập thô.

Do hạn chế về mặt thời gian và khuôn khổ của luận văn được ấn định nên có một số vấn đề thú vị và hấp dẫn không đưa và được trong luận văn, đó là các vấn đề:

- a. Tìm hiểu các không gian xấp xỉ thô dưới dạng HW - đại số và mối liên hệ giữa lý thuyết tập mờ và lý thuyết tập thô thông qua các HW - đại số.
- b. Tìm hiểu các ứng dụng sâu hơn và cụ thể hơn của tập thô vào lĩnh vực khai phá dữ liệu.

Chúng tôi hy vọng sẽ tiếp tục nghiên cứu phát triển đề tài theo hướng này.