

TRƯỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN

KHOA HỆ THỐNG THÔNG TIN

Tài liệu bài giảng:

KHAI THÁC DỮ LIỆU – IS252

Chương 5:

Tập thô và ứng dụng phân lớp

ThS. Dương Phi Long – Email: longdp@uit.edu.vn

NỘI DUNG BÀI HỌC

01



Các khái niệm cơ bản

02



Rút gọn thuộc tính và luật phân lớp

Ứng dụng của tập thô trong KTDL

- Khắc phục hiện tượng dữ liệu bị nhiễu
- Rút gọn dữ liệu (khử dữ liệu thừa)
- Tạo luật phân lớp
- Nhận diện phụ thuộc riêng phần và toàn phần của các thuộc tính



1

Các khái niệm cơ bản

1. Hệ thống tin
2. Hệ quyết định
3. Quan hệ tương đương
4. Quan hệ bất khả phân
5. Xấp xỉ tập hợp
6. Phụ thuộc thuộc tính

1. Hệ thống tin (Information system)

- **Hệ thống tin** (IS): cặp **(U, A)**
 - U: là tập các đối tượng
 - A: là tập các thuộc tính
 - V_a : tập giá trị của thuộc tính a ($a \in A$)
- **VD1:**
 - $U = \{x1, x2, x3, x4, x5, x6, x7\}$
 - $A = \{\text{"Age"}, \text{"LEMS"}\}$

	Age	LEMS
x1	16-30	50
x2	16-30	0
x3	31-45	1-25
x4	31-45	1-25
x5	46-60	26-49
x6	16-30	26-49
x7	46-60	26-49

LEMS: Hội chứng nhược cơ Lambert-eaton

2. Hệ quyết định (Decision system)

- **Hệ quyết định** (DS): cặp **(U, C ∪ D)**

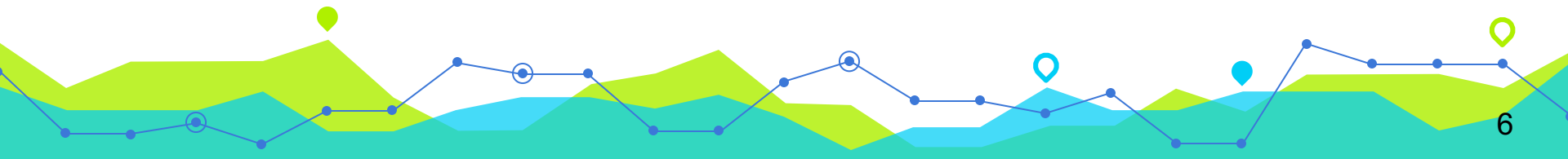
- U: tập các đối tượng
- C: tập thuộc tính điều kiện
- D: tập thuộc tính quyết định

- **VD2:**

- $U = \{x1, x2, x3, x4, x5, x6, x7\}$
- $A = \{\text{"Age"}, \text{"LEMS"}\}$
- $D = \{\text{"Walk"}\}$

	Age	LEMS	Walk
x1	16-30	50	Yes
x2	16-30	0	No
x3	31-45	1-25	No
x4	31-45	1-25	Yes
x5	46-60	26-49	No
x6	16-30	26-49	Yes
x7	46-60	26-49	No

LEMS: Hội chứng nhược cơ Lambert-eaton



2. Hệ quyết định (Decision system)

- Quan sát VD2:

- {x3, x4} có cùng giá trị thuộc tính điều kiện, nhưng khác giá trị thuộc tính quyết định
- {x5, x7} có cùng giá trị thuộc tính điều kiện và cùng giá trị thuộc tính quyết định

⇒ “Nếu Age = “16-30” và LEMS = “50” thì Walk = “Yes”.

	Age	LEMS	Walk
x1	16-30	50	Yes
x2	16-30	0	No
x3	31-45	1-25	No
x4	31-45	1-25	Yes
x5	46-60	26-49	No
x6	16-30	26-49	Yes
x7	46-60	26-49	No

3. Quan hệ tương đương

- Cho U : tập các đối tượng và R : một quan hệ 2 ngôi trên $U \times U$ // $R \subseteq U \times U$
- **R** được gọi là **quan hệ tương đương** nếu:
 - (i) Phản xạ: $\forall x \in U: xRx$ // $(x, x) \in R$
 - (ii) Phản xứng: $\forall x, y \in U$: nếu xRy thì yRx
 - (iii) bắc cầu: $\forall x, y, z \in U$: nếu xRy và yRz thì xRz
- **Lớp tương đương** của x : $[x]_R = \{x \in U / xRy\}$
- Tập các lớp tương đương theo quan hệ R : U/R

3. Quan hệ tương đương

VD3: Cho U là tập các sinh viên trong lớp.

Định nghĩa quan hệ 2 ngôi $R \subseteq U \times U$ như sau:

Với $(x, y) \in U \times U$, $(x, y) \in R \Leftrightarrow x$ **cùng giới tính** với y

- Nhận xét: R là quan hệ tương đương
- Ta có:
 - Giới tính(Tài) = "Nam"; Giới tính(Cường) = "Nam"
 - \Rightarrow Tài cùng giới tính với Cường
 - $\Leftrightarrow (\text{Tài}, \text{Cường}) \in R$ // Tài R Cường
 - Lớp tương đương của Tài: $[\text{Tài}]_R = \{\text{Tài}, \text{Cường}, \text{Quang}\} = [\text{Cường}]_R$

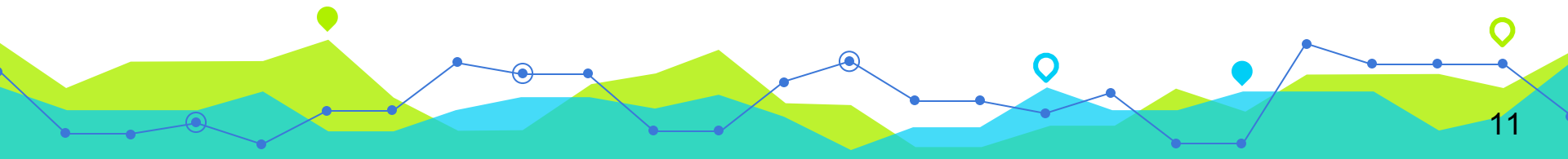
3. Quan hệ tương đương

VD3:

- Ta có:
 - Giới tính(Liên) = "Nữ"; Giới tính(Lan) = "Nữ"
 - \Rightarrow Liên cùng giới tính với Lan
 - $\Leftrightarrow (\text{Liên}, \text{Lan}) \in R$ // Liên R Lan
 - Lớp tương đương của Liên: $[\text{Liên}]_R = \{\text{Liên}, \text{Lan}, \text{Mai}, \dots\}$
 - $(\text{Tài}, \text{Liên}) \notin R$
 - Tập các lớp tương đương theo R:
 $U/R = \{\{\text{Tài}, \text{Cường}, \text{Quang}\}, \{\text{Liên}, \text{Lan}, \text{Mai}, \dots\}\} = \{[\text{Tài}]_R, [\text{Liên}]_R\}$

3. Quan hệ tương đương

- Nhận xét:
 - Mỗi đối tượng đều thuộc 1 lớp duy nhất
 - Hai đối tượng trong 1 lớp có quan hệ với nhau
 - Hai đối tượng thuộc 2 lớp khác nhau không quan hệ với nhau



3. Quan hệ bất khả phân

- Cho hệ thông tin $IS = (U, A)$, tập thuộc tính $B \subseteq A$
- R là một quan hệ trên $U \times U$, được gọi là **quan hệ bất khả phân theo B** (B-indiscernibility relation) **khi**:

$$R = \{ (x, y) \in U^2 \mid \forall b \in B, b(x) = b(y) \}$$

- Ký hiệu: $IND_{IS}(B)$
- Quan hệ bất khả phân là quan hệ tương đương
- Lớp tương đương của x theo quan hệ bất khả phân theo B:

$$[x]_R = [x]_{IND_{IS}(B)} = [x]_B = [x]$$

3. Quan hệ bất khả phân

VD4: Cho hệ thông tin IS

- Với $B = \{\text{"Age"}\}$

- Xét cặp $(x1, x2)$:

Ta có: $\text{Age}(x1) = \text{Age}(x2) = \text{"16-30"}$

Do đó: $(x1, x2) \in \text{IND}_{\text{IS}}(\{\text{"Age"}\})$ // quan hệ bất khả phân theo "Age"

Nghĩa là: Xét về "Age" thì $x1$ và $x2$ là giống nhau. $x1$ và $x2$ không thể phân biệt được qua "Age"

- Xét cặp $(x2, x3)$: $(x2, x3) \notin \text{IND}_{\text{IS}}(\{\text{"Age"}\})$

	Age	LEMS
x1	16-30	50
x2	16-30	0
x3	31-45	1-25
x4	31-45	1-25
x5	46-60	26-49
x6	16-30	26-49
x7	46-60	26-49

3. Quan hệ bất khả phân

VD4:

- Với $B = \{\text{"Age"}, \text{"LEMS"}\}$
 - Tập các lớp tương đương theo B
 $U/B = \{\{x1\}, \{x2\}, \{x3, x4\}, \{x5, x7\}, \{x6\}\}$
 - Lớp tương đương của $x3$ theo B:
 $[x3]_B = \{x3, x4\} = [x4]_B$
 - Các đối tượng trong cùng một lớp sẽ có "Age" và "LEMS" bằng nhau.
- VD: $\{x3, x4\}$ có $\text{Age}(x3) = \text{Age}(x4) = \text{"31-45"}$
và $\text{LEMS}(x3) = \text{LEMS}(x4) = \text{"1-25"}$

	Age	LEMS
x1	16-30	50
x2	16-30	0
x3	31-45	1-25
x4	31-45	1-25
x5	46-60	26-49
x6	16-30	26-49
x7	46-60	26-49

3. Quan hệ bất khả phân

VD4:

- Ta có:

- $IND(\{Age\}) = \{\{x1, x2, x6\}, \{x3, x4\}, \{x5, x7\}\}$
- $IND(\{LEMS\}) = \{\{x1\}, \{x2\}, \{x3, x4\}, \{x5, x6, x7\}\}$
- $IND(\{Age, LEMS\}) = \{\{x1\}, \{x2\}, \{x3, x4\}, \{x5, x7\}, \{x6\}\}$

	Age	LEMS
x1	16-30	50
x2	16-30	0
x3	31-45	1-25
x4	31-45	1-25
x5	46-60	26-49
x6	16-30	26-49
x7	46-60	26-49

4. Xấp xỉ tập hợp

- Ta thấy: Không thể định nghĩa rõ ràng các đối tượng có thuộc tính quyết định dương (Walk = "Yes") từ các thuộc tính khác.
- Gặp khó khăn ở x3, x4. Không thể có 1 định nghĩa chính xác nhưng đối tượng như vậy
→ Cần tập thô

	Age	LEMS	Walk
x1	16-30	50	Yes
x2	16-30	0	No
x3	31-45	1-25	No
x4	31-45	1-25	Yes
x5	46-60	26-49	No
x6	16-30	26-49	Yes
x7	46-60	26-49	No

4. Xấp xỉ tập hợp

Mục đích:

- Chỉ ra những đối tượng có thuộc tính quyết định dương
- Chỉ ra những đối tượng có thuộc tính quyết định không dương
- Những đối tượng nào thuộc vùng biên giữa các trường hợp chắc chắn

	Age	LEMS	Walk
x1	16-30	50	Yes
x2	16-30	0	No
x3	31-45	1-25	No
x4	31-45	1-25	Yes
x5	46-60	26-49	No
x6	16-30	26-49	Yes
x7	46-60	26-49	No

4. Xấp xỉ tập hợp

Định nghĩa

- Gọi $T = (U, A)$ và tập thuộc tính $B \subseteq A$, tập đối tượng $X \subseteq U$
- Chúng ta có thể xấp xỉ X dùng các thông tin chứa trong B bằng cách tạo các xấp xỉ dưới và trên của B
 - **Xấp xỉ dưới:** $\underline{B}X = \{x \mid [x]_B \subseteq X\}$
 - **Xấp xỉ trên:** $\overline{B}X = \{x \mid [x]_B \cap X \neq \emptyset\}$

4. Xấp xỉ tập hợp

- $\underline{B}X$: Tập các đối tượng **chắc chắn** được phân vào lớp X theo B
- $\overline{B}X$: Tập các đối tượng **có thể** được phân vào lớp X theo B
- Vùng B-biên của X: Tập các đối tượng **không thể** phân lớp chắc chắn là lớp X hoặc $\neg X$ theo B

$$BN_B(X) = \overline{B}X - \underline{B}X$$

- Vùng B-ngoài của X: $U - \overline{B}X$: Tập các đối tượng **chắc chắn không thể** phân vào lớp X theo B
- Một tập được gọi là **thô** (rough set) nếu **vùng biên của nó khác rỗng**, ngược lại tập là **rõ**.

4. Xấp xỉ tập hợp

- Các bước thực hiện

- B1. Phân lớp U theo B: U/B (hoặc $\text{IND}(B)$)

- B2. Xấp xỉ dưới: // $\text{LOWER}(B, X)$

$$\underline{B}X = \{x \mid [x]_B \subseteq X\}$$

- B3. Xấp xỉ trên: // $\text{UPPER}(B, X)$

$$\overline{B}X = \{x \mid [x]_B \cap X \neq \emptyset\}$$

- B4. Vùng biên $\text{BN}_B(X)$: $\text{B_Biên} = \overline{B}X - \underline{B}X$

- B5. Vùng ngoài: $\text{B_Ngoài} = U - \overline{B}X$

4. Xấp xỉ tập hợp

- Các bước thực hiện

- B6. Hệ số xấp xỉ:

$$\alpha_B(X) = \frac{|\underline{BX}|}{|\overline{BX}|} \quad \text{Với } |A|: \text{lực lượng của } A \neq \emptyset$$

- $0 \leq \alpha_B(X) \leq 1$
- Nếu $\alpha_B(X) = 1$: X là rõ so với B
- Nếu $\alpha_B(X) < 1$: X là thô so với B

4. Xấp xỉ tập hợp

VD5: Cho hệ IS = (U, A)

Với $B = \{\text{Age, LEMS}\}$

Và $X = \{x1, x4, x6\}$ // các bệnh nhân có Walk = "Yes"

Hãy xấp xỉ X theo B

	Age	LEMS	Walk
x1	16-30	50	Yes
x2	16-30	0	No
x3	31-45	1-25	No
x4	31-45	1-25	Yes
x5	46-60	26-49	No
x6	16-30	26-49	Yes
x7	46-60	26-49	No

4. Xấp xỉ tập hợp

VD5:

Gọi X là tập các đối tượng Walk

$$X = \{x \mid \text{Walk}(x) = \text{"Yes"}\} = \{x1, x4, x6\}$$

- B1. Phân lớp theo $B = \{\text{Age}, \text{LEMS}\}$

$$U/B = \{\{x1\}, \{x2\}, \{x3, x4\}, \{x5, x7\}, \{x6\}\}$$

- B2. Xấp xỉ dưới

$$\underline{B}X = \{x \mid [x]_B \subseteq X\} = \{x1, x6\}$$

- B3. Xấp xỉ trên

$$\overline{B}X = \{x \mid [x]_B \cap X \neq \emptyset\} = \{x1, x3, x4, x6\}$$

	Age	LEMS	Walk
x1	16-30	50	Yes
x2	16-30	0	No
x3	31-45	1-25	No
x4	31-45	1-25	Yes
x5	46-60	26-49	No
x6	16-30	26-49	Yes
x7	46-60	26-49	No

4. Xấp xỉ tập hợp

VD5:

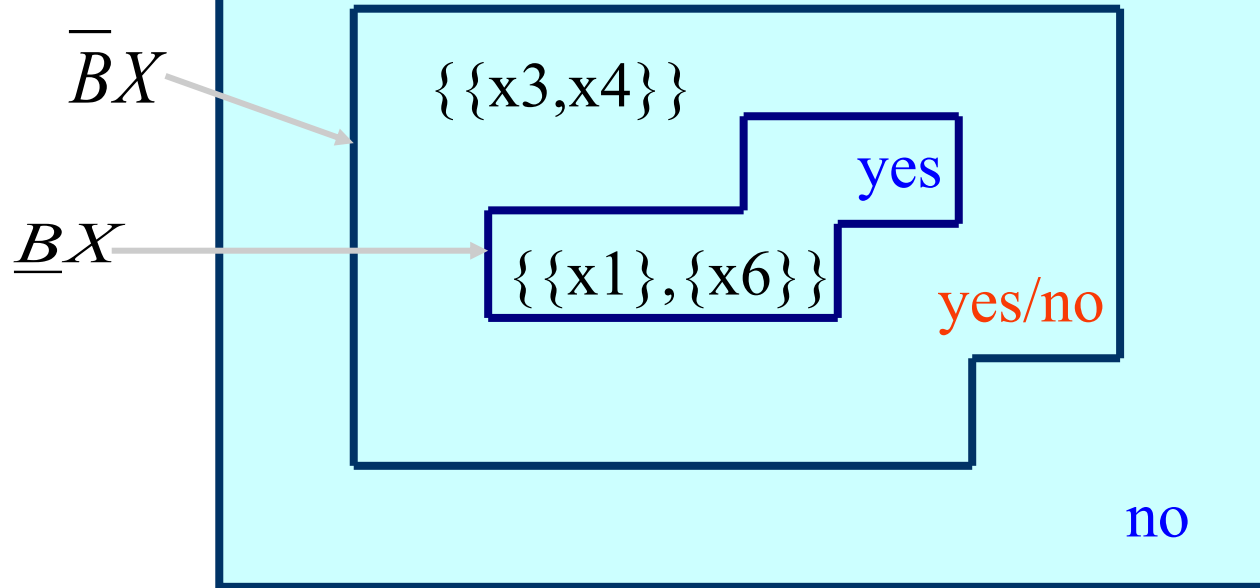
- B4. $B_Biên = \overline{BX} - \underline{BX} = \{x3, x4\}$
- B5. $B_Ngoài = U - \overline{BX} = \{x2, x5, x7\}$
- B6. Hệ số xấp xỉ

$$\alpha_B(X) = \frac{|\underline{BX}|}{|\overline{BX}|} = \frac{|\{x1, x6\}|}{|\{x1, x3, x4, x6\}|} = \frac{2}{4} = 0.5$$

- $\alpha_B(X) < 1 \Rightarrow$ Lớp quyết định Walk là thô

4. Xấp xỉ tập hợp

VD5:



Lớp quyết định
Walk là thô
vì vùng biên
khác rỗng

5. Phụ thuộc thuộc tính

- Tập thuộc tính D phụ thuộc hoàn toàn vào tập thuộc tính C nếu tất cả các thuộc tính của D đều được xác định duy nhất bởi giá trị của các thuộc tính trong C.
- Ký hiệu: $C \Rightarrow D$
- Hệ số phụ thuộc:

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

- Nếu $k = 1$: D phụ thuộc hoàn toàn vào C.
- Nếu $k < 1$: D phụ thuộc một phần (theo mức độ k) vào C

Xấp xỉ tập hợp: Bài tập

BT08

Cho hệ quyết định sau:

1. Tính xấp xỉ tập $X = \{o1, o3, o4\}$ qua tập thuộc tính $B = \{\text{trời, hướng gió}\}$
2. Khảo sát sự phụ thuộc của thuộc tính $C = \{\text{Kết quả}\}$ vào $B = \{\text{trời, hướng gió}\}$

Subject	Trời	Hướng Gió	Áp suất	Kết quả
o1	Trong	Bắc	Cao	Ko mưa
o2	Mây	Nam	Cao	Mưa
o3	Mây	Bắc	Trung bình	Mưa
o4	Trong	Bắc	Thấp	Ko mưa
o5	Mây	Bắc	Thấp	Mưa
o6	Mây	Bắc	Cao	Mưa
o7	Mây	Nam	Thấp	Ko mưa
o8	Trong	Nam	Cao	Ko mưa



2

Rút gọn thuộc tính và luật phân lớp

1. Rút gọn thuộc tính
2. Luật phân lớp

1. Rút gọn thuộc tính

- Một số vấn đề về bảng quyết định:
 - Có thể biểu diễn lặp lại các đối tượng giống nhau hay bất khả phân biệt
 - Một số thuộc tính có thể thừa

	Age	LEMS	Walk
x1	16-30	50	Yes
x2	16-30	0	No
x3	31-45	1-25	Yes
x4	31-45	1-25	No
x5	46-60	26-49	No
x6	16-30	26-49	Yes
x7	46-60	26-49	No

1. Rút gọn thuộc tính

- Chỉ giữ lại các thuộc tính điều kiện bảo toàn quan hệ bất khả phân và hệ quả là bảo toàn xấp xỉ tập hợp.
- Các tập con thuộc tính điều kiện được giữ lại gọi là **tập rút gọn** (reduct).



1. Rút gọn thuộc tính

- Các bước thực hiện

- B1. Xác định ma trận phân biệt
- B2. Xác định hàm phân biệt
- B3. Tìm rút gọn và lỗi



1.1. Ma trận phân biệt

Cho $DS = (U, C \cup D)$, với $U = \{u_1, u_2, \dots, u_n\}$

Ma trận phân biệt của DS: $M(DS) = (m_{ij})_{n \times n}$

- n : số đối tượng trong U
- m_{ij} : tập thuộc tính

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

1.2. Hàm phân biệt

Hàm phân biệt f_{DS} là một hàm logic, được xác định từ ma trận phân biệt $M(DS)$:

$$f_{DS}(u_i) = \bigwedge_{i \neq j} (\bigvee m_{ij}) \text{ với mỗi } u_i \in U \quad (*)$$

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

- (1) $\bigvee m_{ij}$ là biểu thức tuyển của tất cả các biến $c \in m_{ij}$, nếu $m_{ij} \neq \emptyset$
- (2) $\bigvee m_{ij} = \text{true}$, nếu $m_{ij} = \lambda$ và $D(u_i) = D(u_j)$
- (3) $\bigvee m_{ij} = \text{false}$, nếu $m_{ij} = \emptyset$ và $D(u_i) \neq D(u_j)$

(*): λ : true

Trong 1 ô: phép toán \vee
Giữa 2 ô: phép toán \wedge

1.3. Tìm rút gọn và lỗi

- **Đơn giản hàm:**

a) $(A) \wedge (\text{True}) = (A)$

b) $(A) \wedge (A \vee B) = (A)$

c) $(A) \vee (A \wedge B) = (A)$

- **Chuyển dạng hàm:**

Dạng cũ: $(\vee \vee \vee) \wedge (\vee \vee \vee) \wedge (\vee \vee \vee)$

Dạng mới: $(\wedge \wedge \wedge) \vee (\wedge \wedge \wedge) \vee (\wedge \wedge \wedge)$

- **Core (C):** Lỗi của tập thuộc tính điều kiện C là kết quả của phép giao các tập con của C

2. Luật phân lớp

- Từ các tập rút gọn, rút ra các luật phân lớp
- Các bước thực hiện:
 - B1. Phân lớp U theo D: $U/D = \{X_1, X_2, \dots\}$
 - B2. Phân lớp U theo R1: $U/R1 = \{Z_1, Z_2, Z_3, \dots, Z_k\}$
Phân lớp U theo R2: $U/R2 = \{Z_{k+1}, Z_{k+2}, Z_{k+3}, \dots\}$
 - B3. Nếu $Z_i \subseteq X_j$, ta có luật phân lớp **r**: **$Z_i \rightarrow X_j$**

Rút gọn thuộc tính và Luật phân lớp

VD6: Cho hệ quyết định DS (U, C \cup D)

Với U = {x1, x2, x3, x4, x5, x6, x7}

C = {H, M, T} là các thuộc tính điều kiện

D = {F} là thuộc tính quyết định

1. Rút gọn các thuộc tính điều kiện và tìm một số luật phân lớp có độ chính xác 100%
2. Khảo sát sự phụ thuộc thuộc tính của D = {F} vào B = {M, T}

U	H eadache	M usclepain	T emp	F lu
x1	Yes	Yes	Normal	No
x2	Yes	Yes	High	Yes
x3	Yes	Yes	Very-high	Yes
x4	No	Yes	Normal	No
x5	No	No	High	No
x6	No	Yes	Very-high	Yes

Rút gọn thuộc tính và Luật phân lớp

VD6: 1.1. Rút gọn các thuộc tính điều kiện

B1. Tìm ma trận phân biệt

U	H	M	T	F
x1	Yes	Yes	Normal	No
x2	Yes	Yes	High	Yes
x3	Yes	Yes	Very-high	Yes
x4	No	Yes	Normal	No
x5	No	No	High	No
x6	No	Yes	Very-high	Yes



	x1	x2	x3	x4	x5
x2	?				
x3	?	?			
x4	?	?	?		
x5	?	?	?	?	
x6	?	?	?	?	?

$$m_{ij} = \begin{cases} \{c \in C \mid c(x_i) \neq c(x_j)\}, & \text{nếu } D(x_i) \neq D(x_j) \\ \lambda & , \text{nếu } D(x_i) = D(x_j) \end{cases}$$

Rút gọn thuộc tính và Luật phân lớp

VD6: 1.1. Rút gọn các thuộc tính điều kiện

B1. Tìm ma trận phân biệt

U	H	M	T	F
x1	Yes	Yes	Normal	No
x2	Yes	Yes	High	Yes
x3	Yes	Yes	Very-high	Yes
x4	No	Yes	Normal	No
x5	No	No	High	No
x6	No	Yes	Very-high	Yes



	x1	x2	x3	x4	x5
x2	T				
x3	?	?			
x4	?	?	?		
x5	?	?	?	?	
x6	?	?	?	?	?

Xét m_{12} : $D(x1) = \text{No}$ và $D(x2) = \text{Yes}$
 $\Rightarrow m_{12} = \{c \in C \mid c(x_i) \neq c(x_j)\} = \mathbf{T}$

Rút gọn thuộc tính và Luật phân lớp

VD6: 1.1. Rút gọn các thuộc tính điều kiện

B1. Tìm ma trận phân biệt

U	H	M	T	F
x1	Yes	Yes	Normal	No
x2	Yes	Yes	High	Yes
x3	Yes	Yes	Very-high	Yes
x4	No	Yes	Normal	No
x5	No	No	High	No
x6	No	Yes	Very-high	Yes



	x1	x2	x3	x4	x5
x2	T				
x3	T	?			
x4	?	?	?		
x5	?	?	?	?	
x6	?	?	?	?	?

Xét m_{13} : $D(x1) = \text{No}$ và $D(x3) = \text{Yes}$
 $\Rightarrow m_{13} = \{c \in C \mid c(x_i) \neq c(x_j)\} = \mathbf{T}$

Rút gọn thuộc tính và Luật phân lớp

VD6: 1.1. Rút gọn các thuộc tính điều kiện

B1. Tìm ma trận phân biệt

U	H	M	T	F
x1	Yes	Yes	Normal	No
x2	Yes	Yes	High	Yes
x3	Yes	Yes	Very-high	Yes
x4	No	Yes	Normal	No
x5	No	No	High	No
x6	No	Yes	Very-high	Yes



	x1	x2	x3	x4	x5
x2	T				
x3	T	λ			
x4	?	?	?		
x5	?	?	?	?	
x6	?	?	?	?	?

Xét m_{23} : $D(x2) = \text{Yes}$ và $D(x3) = \text{Yes}$
 $\Rightarrow m_{23} = \lambda$

Rút gọn thuộc tính và Luật phân lớp

VD6: 1.1. Rút gọn các thuộc tính điều kiện

B1. Tìm ma trận phân biệt

U	H	M	T	F
x1	Yes	Yes	Normal	No
x2	Yes	Yes	High	Yes
x3	Yes	Yes	Very-high	Yes
x4	No	Yes	Normal	No
x5	No	No	High	No
x6	No	Yes	Very-high	Yes



	x1	x2	x3	x4	x5
x2	T				
x3	T	λ			
x4	λ	H,T	H,T		
x5	λ	H,M	H,M,T	λ	
x6	H,T	λ	λ	T	M,T

Xét m_{35} : $D(x3) = \text{Yes}$ và $D(x5) = \text{No}$
 $\Rightarrow m_{35} = \{c \in C \mid c(x_i) \neq c(x_j)\} = \mathbf{H, M, T}$

Rút gọn thuộc tính và Luật phân lớp

VD6: 1.1. Rút gọn các thuộc tính điều kiện

B1. Tìm ma trận phân biệt

U	H	M	T	F
x1	Yes	Yes	Normal	No
x2	Yes	Yes	High	Yes
x3	Yes	Yes	Very-high	Yes
x4	No	Yes	Normal	No
x5	No	No	High	No
x6	No	Yes	Very-high	Yes



	x1	x2	x3	x4	x5
x2	T				
x3	T	λ			
x4	λ	H,T	H,T		
x5	λ	H,M	H,M,T	λ	
x6	H,T	λ	λ	T	M,T

Rút gọn thuộc tính và Luật phân lớp

VD6: 1.1. Rút gọn các thuộc tính điều kiện

B2. Tìm hàm phân biệt

	x1	x2	x3	x4	x5
x2	T				
x3	T	λ			
x4	λ	H,T	H,T		
x5	λ	H,M	H,M,T	λ	
x6	H,T	λ	λ	T	M,T

Dòng x2: **(T)**

Dòng x3: **(T) ∧ (True)**

Dòng x4: **(True) ∧ (H ∨ T) ∧ (H ∨ T)**

Dòng x5: **(True) ∧ (H ∨ M) ∧ (H ∨ M ∨ T) ∧ (True)**

Dòng x6: **(H ∨ T) ∧ (True) ∧ (True) ∧ (T) ∧ (M ∨ T)**

Rút gọn thuộc tính và Luật phân lớp

VD6: 1.1. Rút gọn các thuộc tính điều kiện

B2. Tìm hàm phân biệt

$$f(H,M,T) = (T)$$

// dòng x2

$$\wedge (T) \wedge (\text{True})$$

// dòng x3

$$\wedge (\text{True}) \wedge (H \vee T) \wedge (H \vee T)$$

// dòng x4

$$\wedge (\text{True}) \wedge (H \vee M) \wedge (H \vee M \vee T) \wedge (\text{True})$$

// dòng x5

$$\wedge (H \vee T) \wedge (\text{True}) \wedge (\text{True}) \wedge (T) \wedge (M \vee T)$$

// dòng x6

Rút gọn thuộc tính và Luật phân lớp

VD6: 1.1. Rút gọn các thuộc tính điều kiện

B3. Tìm rút gọn và lỗi

$f(H,M,T) = (T)$

// dòng x2

$\wedge (T) \wedge (\text{True})$

// dòng x3

$\wedge (\text{True}) \wedge (H \vee T) \wedge (H \vee T)$

// dòng x4

$\wedge (\text{True}) \wedge (H \vee M) \wedge (H \vee M \vee T) \wedge (\text{True})$

// dòng x5

$\wedge (H \vee T) \wedge (\text{True}) \wedge (\text{True}) \wedge (T) \wedge (M \vee T)$

// dòng x6

// áp dụng a): bỏ True

Rút gọn thuộc tính và Luật phân lớp

VD6: 1.1. Rút gọn các thuộc tính điều kiện

B3. Tìm rút gọn và lỗi

$$\begin{aligned}f(H,M,T) &= (T) \wedge (T) \wedge (H \vee T) \wedge (H \vee T) \wedge (H \vee M) \wedge (H \vee M \vee T) \wedge (H \vee T) \\&\quad \wedge (T) \wedge (M \vee T) \\&= (T) \wedge (H \vee T) \wedge (H \vee M) \wedge (H \vee M \vee T) \wedge (M \vee T) \quad // \text{ áp dụng b)} \\&= (T) \wedge (H \vee M) \wedge (M \vee T) \quad // \text{ áp dụng b)} \\&= (T) \wedge (H \vee M) \\&= (T \wedge H) \vee (T \wedge M) \quad // \text{ áp dụng chuyển dạng hàm}\end{aligned}$$

Rút gọn thuộc tính và Luật phân lớp

VD6: 1.1. Rút gọn các thuộc tính điều kiện

B3. Tìm rút gọn và lõi

Ta có $f(H, M, T) = (T \wedge H) \vee (T \wedge M)$

\Rightarrow Có 2 tập rút gọn: $R_1 = \{T, H\}$ và $R_2 = \{T, M\}$

\Rightarrow Lõi (core) = $R_1 \cap R_2 = \{T\}$

Rút gọn thuộc tính và Luật phân lớp

VD6: 1.2. Tìm các luật phân lớp có độ chính xác 100%

- B1. Phân lớp U theo $D = \{F\}$:

$$U/D = \{X_1, X_2\}$$

$$X_1 = \{x \in U \mid D(x) = \text{"Yes"}\} = \{x_2, x_3, x_6\}$$

$$X_2 = \{x \in U \mid D(x) = \text{"No"}\} = \{x_1, x_4, x_5\}$$

U	H	M	T	F
x1	Yes	Yes	Normal	No
x2	Yes	Yes	High	Yes
x3	Yes	Yes	Very-high	Yes
x4	No	Yes	Normal	No
x5	No	No	High	No
x6	No	Yes	Very-high	Yes

Rút gọn thuộc tính và Luật phân lớp

VD6: 1.2. Tìm các luật phân lớp có độ chính xác 100%

- **B2. Phân lớp U theo $R_1 = \{T, H\}$**

$$U/R_1 = \{Z_1, Z_2, Z_3, Z_4, Z_5, Z_6\}$$

$$Z_1 = \{x1\}, \quad Z_2 = \{x2\},$$

$$Z_3 = \{x3\}, \quad Z_4 = \{x4\},$$

$$Z_5 = \{x5\}, \quad Z_6 = \{x6\}$$

U	H	M	T	F
x1	Yes	Yes	Normal	No
x2	Yes	Yes	High	Yes
x3	Yes	Yes	Very-high	Yes
x4	No	Yes	Normal	No
x5	No	No	High	No
x6	No	Yes	Very-high	Yes

Rút gọn thuộc tính và Luật phân lớp

VD6: 1.2. Tìm các luật phân lớp có độ chính xác 100%

- **B2. Phân lớp U theo $R_2 = \{T, M\}$**

$$U/R_2 = \{Z_7, Z_8, Z_9, Z_{10}\}$$

$$Z_7 = \{x1, x4\}, \quad Z_8 = \{x2\},$$

$$Z_9 = \{x3, x6\}, \quad Z_{10} = \{x5\}$$

U	H	M	T	F
x1	Yes	Yes	Normal	No
x2	Yes	Yes	High	Yes
x3	Yes	Yes	Very-high	Yes
x4	No	Yes	Normal	No
x5	No	No	High	No
x6	No	Yes	Very-high	Yes

Rút gọn thuộc tính và Luật phân lớp

VD6: 1.2. Tìm một số luật phân lớp có độ chính xác 100%

- B3. Tìm luật phân lớp

Ta có: $X_1 = \{x \in U / D(x) = \text{"Yes"}\} = \{x_2, x_3, x_6\}$

$X_2 = \{x \in U / D(x) = \text{"No"}\} = \{x_1, x_4, x_5\}$

$Z_1 = \{x_1\}, Z_2 = \{x_2\}, Z_3 = \{x_3\}, Z_4 = \{x_4\},$

$Z_5 = \{x_5\}, Z_6 = \{x_6\}, Z_7 = \{x_1, x_4\}, Z_8 = \{x_2\},$

$Z_9 = \{x_3, x_6\}, Z_{10} = \{x_5\}$

U	H	M	T	F
x1	Yes	Yes	Normal	No
x2	Yes	Yes	High	Yes
x3	Yes	Yes	Very-high	Yes
x4	No	Yes	Normal	No
x5	No	No	High	No
x6	No	Yes	Very-high	Yes

Giả sử xét: Z_1, Z_7 và Z_9

Rút gọn thuộc tính và Luật phân lớp

VD6: 1.2. Tìm một số luật phân lớp có độ chính xác 100%

- B3. Tìm luật phân lớp

- Xét $Z_1 = \{x1\}$ và $X_2 = \{x1, x4, x5\}$:

Ta thấy: $Z_1 \subset X_2$ và $Z_1 \in U/R_1$, $R_1 = \{T, H\}$

Nên ta có luật phân lớp $r1$:

Nếu $H = \text{"Yes"}$ và $T = \text{"Normal"}$ thì $F = \text{"No"}$

U	H	M	T	F
x1	Yes	Yes	Normal	No
x2	Yes	Yes	High	Yes
x3	Yes	Yes	Very-high	Yes
x4	No	Yes	Normal	No
x5	No	No	High	No
x6	No	Yes	Very-high	Yes

Rút gọn thuộc tính và Luật phân lớp

VD6: 1.2. Tìm một số luật phân lớp có độ chính xác 100%

- B3. Tìm luật phân lớp

- Xét $Z_7 = \{x1, x4\}$ và $X_2 = \{x1, x4, x5\}$:

Ta thấy: $Z_7 \subset X_2$ và $Z_7 \in U/R_2$, $R_2 = \{T, M\}$

Nên ta có luật phân lớp r_2 :

Nếu $M = \text{"Yes"}$ và $T = \text{"Normal"}$ thì $F = \text{"No"}$

U	H	M	T	F
x1	Yes	Yes	Normal	No
x2	Yes	Yes	High	Yes
x3	Yes	Yes	Very-high	Yes
x4	No	Yes	Normal	No
x5	No	No	High	No
x6	No	Yes	Very-high	Yes

Rút gọn thuộc tính và Luật phân lớp

VD6: 1.2. Tìm một số luật phân lớp có độ chính xác 100%

- B3. Tìm luật phân lớp

- Xét $Z_9 = \{x_3, x_6\}$ và $X_1 = \{x_2, x_3, x_6\}$:

Ta thấy: $Z_9 \subset X_1$ và $Z_9 \in U/R_2$, $R_2 = \{T, M\}$

Nên ta có luật phân lớp r_3 :

Nếu $M = \text{"Yes"}$ và $T = \text{"Very-high"}$ thì $F = \text{"Yes"}$

U	H	M	T	F
x1	Yes	Yes	Normal	No
x2	Yes	Yes	High	Yes
x3	Yes	Yes	Very-high	Yes
x4	No	Yes	Normal	No
x5	No	No	High	No
x6	No	Yes	Very-high	Yes

Rút gọn thuộc tính và Luật phân lớp

VD6: 2. Khảo sát sự phụ thuộc thuộc tính của $D = \{F\}$ vào $B = \{M, T\}$

- $U/B = \{\{x1, x4\}, \{x2\}, \{x3, x6\}, \{x5\}\}$
- Hệ số phụ thuộc:

$$k = \gamma(B, D) = \sum_{X \in U/D} \frac{|BX|}{|U|} = \sum_{X \in \{X_1, X_2\}} \frac{|BX|}{|U|}$$
$$= \frac{|BX_1| + |BX_2|}{|U|} = \frac{|\{x_2, x_3, x_6\}| + |\{x_1, x_4, x_5\}|}{6} = \frac{3+3}{6} = 1$$

$\Rightarrow D = \{F\}$ phụ thuộc hoàn toàn vào $B = \{M, T\}$.

U	H	M	T	F
x1	Yes	Yes	Normal	No
x2	Yes	Yes	High	Yes
x3	Yes	Yes	Very-high	Yes
x4	No	Yes	Normal	No
x5	No	No	High	No
x6	No	Yes	Very-high	Yes

Rút gọn và Luật phân lớp: Bài tập

BT09

Cho Hệ quyết định $(U, C \cup D)$. $C = \{N, K, A, L\}$ là các điều kiện và $D = \{T\}$ là quyết định

1. Tính xấp xỉ $X = \{o1, o2, o4, o5, o6\}$ qua tập thuộc tính $B = \{K, A\}$
2. Tính các thu gọn của hệ quyết định trên
3. Hãy nêu 2 luật có độ chính xác phân lớp là 100%
4. Khảo sát sự phụ thuộc của thuộc tính T vào L, N

	N	K	A	L	T
o1	1	C	X	2	C
o2	2	C	B	3	C
o3	2	C	B	2	K
o4	2	C	X	3	C
o5	1	K	B	3	C
o6	3	C	T	1	K
o7	3	K	X	1	K
o8	1	C	B	1	K
o9	1	K	B	3	C

Tổng kết chương



Các khái niệm cơ bản

1. Hệ thống tin
2. Hệ quyết định
3. Quan hệ tương đương
4. Quan hệ bất khả phân
5. Xấp xỉ tập hợp
6. Phụ thuộc thuộc tính



Rút gọn thuộc tính và luật phân lớp

1. Rút gọn thuộc tính
2. Luật phân lớp



Bài tập chương 5

5.1. Cho hệ quyết định sau:

	Bằng cấp d	Kinh nghiệm e	Tiếng Anh f	Giới thiệu r	Tuyển dụng t
x1	MBA	Trung bình	Tốt	Xuất sắc	Chấp nhận
x2	MBA	Ít	Tốt	Trung bình	Từ chối
x3	MCE	Ít	Tốt	Tốt	Từ chối
x4	MSC	Nhiều	Tốt	Trung bình	Chấp nhận
x5	MSC	Trung bình	Tốt	Trung bình	Từ chối
x6	MSC	Nhiều	Tốt	Xuất sắc	Chấp nhận
x7	MBA	Nhiều	Không	Tốt	Chấp nhận
x8	MCE	Ít	Không	Xuất sắc	Từ chối

Bài tập chương 5

5.1. Các yêu cầu:

1. Khảo sát sự phụ thuộc thuộc tính giữa $B = \{\text{Bằng cấp, Giới thiệu}\}$ và $C = \{\text{Tuyển dụng}\}$ và đề xuất một số phân lớp chính xác 100%.
2. Tính xấp xỉ tập $X = \{x_1, x_2, x_3\}$ qua thuộc tính $B = \{\text{Bằng cấp, Kinh nghiệm}\}$
3. Hãy tìm ma trận phân biệt và rút gọn thuộc tính điều kiện

Bài tập chương 5

5.2. Cho hệ quyết định sau. Hãy tìm tập rút gọn thuộc tính và xác định tất cả các loại có độ chính xác 100%.

	Tên	Màu tóc	Chiều cao	Cân nặng	Dùng thuốc	Kết quả
x1	Hoa	Đen	Tầm thước	Nhẹ	Không	Bị râm
x2	Lan	Đen	Cao	Vừa	Có	Không
x3	Xuân	Râm	Thấp	Vừa	Có	Không
x4	Hạ	Đen	Thấp	Vừa	Không	Bị râm
x5	Thu	Bạc	Tầm thước	Nặng	Không	Bị râm
x6	Đông	Râm	Cao	Nặng	Không	Không
x7	Mơ	Râm	Tầm thước	Nặng	Không	Không
x8	Đào	Đen	Thấp	Nhẹ	Có	Không

THANKS!

Any questions?

