Chương 7

Phụ thuộc hàm
và
Chuẩn hóa cơ sở dữ liệu

Nội dung trình bày

- Phụ thuộc hàm.
- Các dạng chuẩn.
- Một số thuật toán chuẩn hóa.

Phụ thuộc hàm (1)

- R(U), U={A_1, A_2,…, A_n}, phụ thuộc hàm giữa hai tập thuộc tính X, Y $\subseteq$ U
  - Ký hiệu: X $\rightarrow$ Y là

    $\forall r \in R,\ \forall\ t_1, t_2 \in r$ nếu $t_1[X] = t_2[X]$ thì $t_1[Y] = t_2[Y]$
  - Có nghĩa là: nếu 2 bộ có cùng giá trị X thì có cùng giá trị Y
  - X là vế phải của pth, Y là vế trái của pth.
- Phụ thuộc hàm được suy ra từ những quy tắc dữ liệu khi ta khảo sát yêu cầu của bài toán
  - Ví dụ: Từ MaNV, ta có thể suy ra được tên của nhân viên (MaNV$\rightarrow$ Ho). Từ MaDA, ta có thể suy ra tên và vị trí của dự án (MaDA$\rightarrow${TenDA, DiaDiem})
- Nếu K là khóa của R thì K xác định hàm tất cả các tập thuộc tính của R.

Phụ thuộc hàm (2)

NHANVIEN_PHONGBAN

TenNV	MaNV	NgSinh	Diachi	MaPB	TenPB	TrPhong

$MaNV \rightarrow TenNV$ $MaNV \rightarrow MaPB$ $MaPB \rightarrow \{TenPB, TrPhong\}$

Luật dẫn – Hệ tiên đề Armstrong

- Cho R(U), U là tập thuộc tính, F là tập các phụ thuộc hàm được định nghĩa trên quan hệ R. Ta nói phụ thuộc hàm $A \rightarrow B$ **được suy diễn logic từ F** nếu mọi trạng thái $r \in R$ thỏa các phụ thuộc hàm trong F thì cũng thỏa phụ thuộc hàm $A \rightarrow B$.
  - Ví dụ: F = {MaNV → TenNV, MaPB → {TenPB, TrPhong}, MaNV → MaPB}. Ta có phụ thuộc hàm MaNV → {TenPB, TrPhong} là phụ thuộc hàm được suy diễn từ F.

Luật dẫn – Hệ tiên đề Armstrong

- Hệ tiên đề Armstrong được sử dụng để tìm ra các phụ thuộc hàm được suy diễn từ F
  - Phản xạ: $Y \subseteq X \Rightarrow X \to Y$.
  - Tăng trưởng: $X \to Y \Rightarrow XZ \to YZ$, với $XZ = X \cup Z$.
  - Bắc cầu: $X \to Y$, $Y \to Z \Rightarrow X \to Z$.
- Các luật khác:
  - Phân rã: $X \to YZ \Rightarrow X \to Y$, $X \to Z$.
  - Hợp: $X \to Y$, $X \to Z \Rightarrow X \to YZ$.
  - Bắc cầu giả: $X \to Y$, $WY \to Z \Rightarrow WX \to Z$.
- Nhận xét
  - Hệ luật Armstrong là đầy đủ.

Bao đóng của tập PTH

- Ta gọi f là một phụ thuộc hàm được suy dẫn từ F nếu tồn tại một chuỗi phụ thuộc hàm: f_1, f_2, …, f_n sao cho f_n = f và mỗi f_i là một thành viên của F hay được suy diễn từ những phụ thuộc hàm j = 1, …, i-1 trước đó nhờ vào luật dẫn.
- Bao đóng của F, ký hiệu F^+, là tập hợp các phụ thuộc hàm được suy diễn từ F nhờ vào hệ tiên đề Armstrong.
  - $F \subseteq F^+$.
  - F gọi là đầy đủ nếu $F = F^+$.

Bao đóng của tập thuộc tính

- Bao đóng của tập thuộc tính X xác định trên tập phụ thuộc hàm F ký hiệu là X_F^+ là tập hợp tất cả các thuộc tính có thể suy ra từ X.
  - $X_F^+ = \{A \in U \mid X \rightarrow A \in F^+\}$
- Nhận xét
  - Bài toán xác định một phụ thuộc hàm có được suy diễn từ tập phụ thuộc hàm cho trước?
    - $X \rightarrow Y \in F^+ \Leftrightarrow Y \subseteq X_F^+$.

Thuật toán tìm bao đóng của tập thuộc tính

- Nhập: U, F và $X \subseteq U$
- Xuất: X_F^+
- Thuật toán 7.1
  - $X^+ = X$;
  - Repeat

    $oldX^+ = X^+$;

    For (mỗi phụ thuộc hàm $Y \to Z \in F$) do

    if ($Y \subseteq X^+$ và $Z \not\subset X^+$) then $X^+ := X^+ \cup Z$;
  - Until ($oldX^+ = X^+$).

Ví dụ tìm bao đóng X^+

- Cho:
  - $F = \{AB \rightarrow C, BC \rightarrow D, D \rightarrow EG\}$.
  - $X = BD$.
- Tính X^+:
  - $X^+ = BD$.
  - Lặp 1:
    - oldX$^+$ = X^+ = BD
    - $D \rightarrow EG$, thêm EG vào X^+ ta được X^+ = BDEG.
  - Lặp 2:
    - oldX$^+$ = X^+ = BDEG
    - Không thể mở rộng thêm X^+.
  - Vậy X^+ = BDEG.

Bài toán xác định phụ thuộc hàm suy diễn

- Cho F = {AB $\rightarrow$ C, A $\rightarrow$ D, D $\rightarrow$ E, AC $\rightarrow$ B}
- Hai phụ thuộc hàm AB $\rightarrow$ E và D $\rightarrow$ C có được suy diễn từ F hay không?

X	X_F^+
AB	ABCDE
D	DE

AB $\rightarrow$ E Được suy diễn từ F

Tập phụ thuộc hàm tương đương

- Tập phụ thuộc hàm F được nói là phủ tập phụ thuộc hàm G nếu tất cả các phụ thuộc hàm trong G có thể được suy diễn từ F.
  - F phủ G $\Leftrightarrow$ G $\subset$ F^+.
  - $\forall X \rightarrow Y \in G$, nếu $Y \subseteq X_F^+$ thì F phủ G.
- Hai tập phụ thuộc hàm F và G là tương đương nếu tất cả các phụ thuộc hàm trong F có thể được suy diễn từ G, và tất cả các phụ thuộc hàm trong G có thể được suy diễn từ F.
  - F và G tương đương $\Leftrightarrow$ $F^+ = G^+$.
  - F và G là tương đương $\Leftrightarrow$ F phủ G và G phủ F.

Tập phụ thuộc hàm tối thiểu

- Vấn đề thừa phụ thuộc hàm
  - $\{A \rightarrow B, B \rightarrow C, A \rightarrow C\}$, vì $A \rightarrow C$ được suy diễn từ $\{A \rightarrow B, B \rightarrow C\}$

    $A \rightarrow B, B \rightarrow C \Rightarrow A \rightarrow C$ (luật bắc cầu).
- Vấn đề thừa thuộc tính
  - $\{A \rightarrow B, B \rightarrow C, A \rightarrow CD\}$, vì $A \rightarrow CD$ được suy diễn từ $\{A \rightarrow B, B \rightarrow C, A \rightarrow D\}$

    $A \rightarrow B, B \rightarrow C \Rightarrow A \rightarrow C$ (luật bắc cầu)

    $A \rightarrow C, A \rightarrow D \Rightarrow A \rightarrow CD$ (luật hợp).
  - $\{A \rightarrow B, B \rightarrow C, AC \rightarrow D\}$, vì $AC \rightarrow D$ được suy diễn từ $\{A \rightarrow B, B \rightarrow C, A \rightarrow D\}$

    $A \rightarrow B, A \rightarrow D \Rightarrow A \rightarrow BD$ (luật hợp)

    $A \rightarrow BD \Rightarrow AC \rightarrow BCD$ (luật tăng trưởng)

    $AC \rightarrow BCD \Rightarrow AC \rightarrow D$ (luật phân rã).

Tập phụ thuộc hàm tối thiểu

- Tập phụ thuộc hàm F là tối thiểu nếu nó thỏa các điều kiện sau:
  - Mọi phụ thuộc hàm của F chỉ có một thuộc tính ở vế phải.
  - Không thể thay thế bất kỳ phụ thuộc hàm $X \rightarrow A$ nào trong F bằng phụ thuộc hàm $Y \rightarrow A$, với $Y \subset X$ mà vẫn có được một tập phụ thuộc hàm tương đương với F (tức là, không có thuộc tính dư thừa trong phụ thuộc hàm)
  - Không thể bỏ đi bất kỳ một phụ thuộc hàm nào trong F mà vẫn có được một tập phụ thuộc hàm tương đương với F (tức là, không có phụ thuộc hàm dư thừa).
- Phủ tối thiểu của tập phụ thuộc hàm F là tập phụ thuộc hàm tối thiểu tương đương với F.
  - Tất cả các tập phụ thuộc hàm đều có ít nhất một phủ tối thiếu.

Thuật toán tìm phủ tối thiểu

- Nhập: tập phụ thuộc hàm F.
- Xuất: phủ tối thiểu F.
- Thuật toán 7.2
  - *B1*: G := $\emptyset$.
  - *B2*: Với mỗi $X \rightarrow Y \in F$, $Y = \{A_1, \ldots, A_k\}$, $A_i \in U$

    $G := G \cup \{X \rightarrow A_i\}$.
  - *B3*: (Loại bỏ thuộc tính thừa trong các phụ thuộc hàm)

    Với mỗi $X \rightarrow A \in G$, với mỗi thuộc tính $B \in X$ nếu $A \in (X - B)_G^+$ thì $G := (G - \{X \rightarrow A\}) \cup \{(X - B) \rightarrow A\}$.
  - *B4*: (Loại bỏ phụ thuộc hàm thừa)

    Với mỗi $X \rightarrow A \in G$

    $G' := G - \{X \rightarrow A\}$

    Nếu $A \in X_{G'}^+$ thì $G := G - \{X \rightarrow \{A\}\}$.

Ví dụ tìm phủ tối thiểu

- Tìm phủ tối thiểu của F = {$A \rightarrow BC$, $A \rightarrow B$, $B \rightarrow C$, $AB \rightarrow C$}
  - *B1*: G = $\varnothing$.
  - *B2*: G = {$A \rightarrow B$, $A \rightarrow C$, $B \rightarrow C$, $AB \rightarrow C$}.
  - *B3*: Xét $AB \rightarrow C$

    $(B)_G^+ = BC$

    G = {$A \rightarrow B$, $A \rightarrow C$, $B \rightarrow C$}.
  - *B4*: $A \rightarrow C$ thừa.

    G = {$A \rightarrow B$, $B \rightarrow C$}.

Siêu khóa và khóa

- Cho R(U) và tập phụ thuộc hàm F xác định trên R
  - Tập thuộc tính $S \subseteq U$ là siêu khóa nếu $\forall r \in R$, $\forall t_1, t_2 \in r, t_1 \neq t_2$ thì $t_1[S] \neq t_2[S]$.
  - Tập thuộc tính $K \subseteq U$ là khóa nếu K là siêu khóa và nếu bỏ đi một thuộc tính bất kỳ trong K thì tập còn lại không còn là siêu khóa.
- Nhận xét
  - S là siêu khóa của R $\Leftrightarrow S_F^+ = U$.
  - R có thể có nhiều khóa.

Thuật toán tìm khóa

- Nhập: R(U) và tập phụ thuộc hàm F.
- Xuất: khóa K của R.
- Thuật toán 7.3.1
  - *B1*:

    $K = U = \{A_1, \ldots, A_n\}$;

    $i = 1$;
  - *B2*:

    Nếu $U \subseteq (K - A_i)_F^+$ thì $K = K - A_i$.

    $i = i + 1$;

    Nếu $i > n$ thì sang *B3*. Ngược lại, tiếp tục *B2*.
  - *B3*:

    Xuất K.

Ví dụ tìm khóa

- Cho R(U), U = {A, B, C, D, E, F, G}.
  - F = {B → A, D → C, D → BE, DF → G}.
- Tìm khóa của R
  - B1:

    K = ABCDEFG.
  - B2:
    - Lặp 1: $(BCDEFG)_F{}^+$ = BCDEFGA ⇒ K = BCDEFG.
    - Lặp 2: $(CDEFG)_F{}^+$ = CDEFGBA ⇒ K = CDEFG.
    - Lặp 3: $(DEFG)_F{}^+$ = DEFGCBA ⇒ K = DEFG.
    - Lặp 4: $(EFG)_F{}^+$ = EFG.
    - Lặp 5: $(DFG)_F{}^+$ = DFGCBEA ⇒ K = DFG.
    - Lặp 6: $(DG)_F{}^+$ = DGCBEA.
    - Lặp 7: $(DF)_F{}^+$ = DFCBEAG ⇒ K = DF.
  - B3:

    Khóa là K = DF.

Thuật toán tìm tất cả khóa

- Nhập: tập PTH F xác định trên lược đồ R(U).
- Xuất: tất cả khóa của R.
- Thuật toán 7.3.2
  - *B1*:

    Xây dựng $2^n - 1$ tập con khác rỗng của $U = \{A_1, \ldots, A_n\}$;

    $S = \{\}$;
  - *B2*:

    Với mỗi tập con $\emptyset \neq X \subseteq U$

    Tính bao đóng X_F^+;

    Nếu $U \subseteq X_F^+$ thì $S = S \cup \{X\}$.
  - *B3*:

    $\forall K_1, K_2 \in S$, nếu $K_1 \subset K_2$ thì $S = S - K_2$.
  - *B4*:

    S là tập các khóa của R.

Ví dụ tìm tất cả khóa

- Cho R(U), U = {A, B, C, D, E, F}.
  - F = {AE → C, CF → A, BD → F, AF → E}.
- Tìm tất cả khóa của R

X	**X_F^+**	**Siêu khóa**	**Khóa**
A	A		
B	B		
…			
AE	AEC		
…			
ABD	ABDFEC	ABD	ABD
…			
BCD	BCDFAE	BCD	BCD
…			
ABCD	ABCDEF	ABCD	
…			

Tập siêu khóa S = {ABD, BCD, ABCD, ABDE, BCDE, ABCDE, ABDF, BCDF, ABCDF, ABDEF, BCDEF, ABCDEF}

Thuật toán tìm tất cả các khóa cải tiến

- Tập hợp các thuộc tính ở vế trái các PTH của F : VT_F
- Tập hợp các thuộc tính ở vế phải các PTH của F : VP_F
- Tập nguồn là tập các thuộc tính không xuất hiện ở vế phải các PTH của F : $TN = U - VP_F$
- Tập đích là tập các thuộc tính có xuất hiện ở vế phải và không xuất hiện ở vế trái các PTH của F : $TD = U - (TN \cup VT_F)$
- Tập trung gian là tập các thuộc tính vừa xuất hiện trong vế trái, vừa xuất hiện trong vế phải của các PTH của F : $L = U - (TN \cup TD)$
- **Nếu K là khóa của R thì tồn tại $L_i \subseteq L$ sao cho $K = TN \cup L_i$.**

Ví dụ tìm tất cả khóa cải tiến

- VT_F = {A, B, C, D, E, F}
- VP_F = {A, C, E, F}
- TN = U – VP_F = {B, D}
- TD = U - (TN ∪ VT_F) = ∅
- L = U – (TN ∪ TD) = {A, C, E, F}

L_i	Li ∪ TN	$(Li \cup TN)_F^+$	Siêu khóa	Khóa
∅	BD	BDF		
A	ABD	ABDFEC	ABD	ABD
C	CBD	CBDFAE	CBD	CBD
E	EBD	EBDF		
F	FBD	FBD		
AC	ACBD	ABCDEF	ABCD	
…				