Chương 5

Các loại phụ thuộc dữ liệu

Phạm Thị Ngọc Diễm
Bộ môn HTTT - ĐHCT

Nội dung

- Giới thiệu vấn đề
- Phụ thuộc hàm

Nội dung

- Giới thiệu vấn đề
- Phụ thuộc hàm

Dư thừa dữ liệu

- Cho quan hệ :
  - `DULIEU(MSSV,hoten,diachi,MM,tenmon,diem)`
- Cho các quan hệ:
  - `SINHVIEN(MSSV,hoten,diachi)`
  - `MONHOC(MM,tenmon)`
  - `DANGKY(MSSV,MM,diem)`

Dư thừa dữ liệu

CÁ_NHÂN (id, hoten, diachi)
THICH (id,sothich)

- Xét quan hệ CÁ_NHÂN (id, hoten, diachi, sothich) với các thể hiện:

id	hoten	diachi	sothich
10110100	John Doe	123 Lý Tự Trọng	Bơi lội
10110100	John Doe	123 Ly Tự Trọng	Bida
21345622	Huỳnh Huy	2 Võ Thị Sáu	Cầu Lông
21345622	Huỳnh Huy	2 Võ Thị Sáu	Bóng chuyền
55555555	Lê Văn Tám	411 30/4	Leo núi

→ (21345622, Huỳnh Huy, 1 Ly tu trong, Bơi lội)

- Xét 4 bộ đầu tiên:
  - Nhiều thông tin lặp lại (id, hoten, diachi) => lưu trữ **dư thừa** cho cùng thông tin => Đây không là vấn đề chính
  - Vấn đề chính là giữ cho các bản sao **dư thừa** luôn nhất quán trong CSDL và điều này phải được thực hiện một cách hiệu quả.

=> Dư thừa có thể dẫn đến bất thường (anomaly) dữ liệu

Dị thường dữ liệu

- Dị thường dữ liệu là
  - Những mâu thuẫn trong các dữ liệu được lưu trữ trong cơ sở dữ liệu,
  - Kết quả của một thao tác như cập nhật, thêm, và / hoặc xóa.
- Sự mâu thuẫn như vậy có thể phát sinh khi
  - Có một bộ đặc biệt được lưu trữ tại nhiều địa điểm (các bản sao);
  - Nhưng không phải tất cả các bản sao đều được cập nhật.

Dị thường dữ liệu - ví dụ

- *Dị thường do cập nhật*
  - Thay đổi địa chỉ cho "Huỳnh Huy" → Phải thay đổi đ/c tất cả các bộ mô tả "Huỳnh Huy"
- *Dị thường do thêm*
  - Thêm bộ mới mà không có thông tin (hoặc NULL) về ***sothich*** => không thể thêm vì ***sothich*** là một phần của Khóa
- *Dị thường do xoá :*
  - Giả sử rằng "Lê Văn Tám" không thích leo núi nữa và ta muốn xoá sở thích này khỏi quan hệ:
    - Không có cách nào để chỉ xoá sở thích
    - Hoặc xóa tất cả thông tin mô tả "Lê Văn Tám" → Mất thông tin do xóa
    - Hoặc thay thế ***sothich*** bởi NULL → vấn đề NULL trong khóa chính

Dị thường dữ liệu

- *Nhận xét: Nếu chỉ có một bộ có thể mô tả một cá nhân*

  → **sothich** *sẽ không là một phần của khoá*

→ *Các vấn đề nêu trong các ví dụ sẽ không xảy ra*

Các tiêu chí đánh giá thiết kế LĐQH

- Đảm bảo rằng ngữ nghĩa của các thuộc tính là rõ ràng trong lược đồ
- Giảm thông tin dư thừa trong các bộ. Dư thừa dữ liệu gây:
  - Dị thường dữ liệu khi thêm hoặc sửa
  - Mất thông tin khi xoá
- Giảm giá trị NULL trong các bộ. Các giá trị NULL làm:
  - Lãng phí không gian lưu trữ
  - Khó thực hiện việc chọn, các hàm kết tập và nối kết
- Không chấp nhận khả năng tạo ra các bộ giả (spurious tuples)
  - Sinh ra do kết nối các quan hệ không dựa trên khoá chính và khoá ngoài.

Các tiêu chí đánh giá thiết kế LĐQH

- Giảm giá trị NULL trong các bộ. Các giá trị NULL làm:

  TAIKHOAN(STK, ngay, sodu,MAKH, MADN)

  TAIKHOAN(STK, ngay, sodu,MAKH,loai, ghichu)

Nội dung

- Giới thiệu vấn đề
- Phụ thuộc hàm

Giới thiệu PTH

- Khái niệm quan trọng nhất trong lý thuyết thiết kế lược đồ quan hệ là phụ thuộc hàm (PTH)
- PTH là công cụ hình thức để phân tích các lược đồ quan hệ :
  - Cho phép phát hiện và
  - Mô tả một số các vấn đề vừa nêu trên
- Một PTH là một ràng buộc giữa hai tập thuộc tính từ một cơ sở dữ liệu.
- PTH được sử dụng để xác định các dạng chuẩn (Normal Form).

Định nghĩa

- Cho lược đồ quan hệ R(U) với:
  - U={ A1, A2, ..., An},
  - $X \neq \emptyset$, $Y \neq \emptyset$, $X \subseteq U$, $Y \subseteq U$

Định nghĩa

X xác định Y hay Y phụ thuộc hàm vào X , nếu và chỉ nếu với mỗi giá trị của X xác định duy nhất một giá trị của Y, hay:

$\forall r \in R,\ \forall t1, t2 \in r$, *t1[X]=t2[X] thì t1[Y]=t2[Y]*

  - Ký hiệu **X → Y**
  - X là vế trái và Y là vế phải của PTH

Ví dụ

R(U)	A	B
	1	4
	1	5
	3	7

- A → B ?
- B → A ?
- MSSV → hoten
- Hoten → MSSV

Ví dụ

- Cho quan hệ R với tập phụ thuộc hàm F :

F= { $A \rightarrow B$
$B,C \rightarrow D$
$D \rightarrow E$
$A,C \rightarrow D$
$A,C \rightarrow E$ }

$AB \rightarrow E$???

R	A	B	C	D	E
	a1	b1	c1	d3	e2
	a1	b1	c3	d4	e3
	a2	b2	c4	d2	e1
	a3	b1	c1	d3	e2
	a2	b2	c4	d2	e1

Ví dụ

- PTH

  MASV → hoten

  MASV → diachi

  MASV, MM, hk, nk → diem

- Chú ý: **không** tồn tại các PTH sau
  - MASV → MAMON
  - MCB → ngaybay
- **Chú ý**: Sự tồn tại của một số PTH trong một LĐQH có thể dẫn đến dị thường (khi thêm/sửa/xóa).
  - Ví dụ: id, sothich → hoten, diachi

    => Chuẩn hoá và tách LĐQH

Các bước thiết kế CSDL theo dạng chuẩn

R(U)

↓

F

↓

F tối tiểu → ***Xác định NF của R*** → ***Tách lược đồ R(U)=R1 * R2 *....*Rn***

Luật suy diễn - Hệ tiên đề Armstrong

- Cho lược đồ quan hệ R(U), U={ A1, A2, ..., An}, $X \neq \emptyset$, $Y \neq \emptyset$, X, Y, Z, $W \subseteq U$
- Hệ tiên đề Armstrong gồm các luật sau:

  - *Phản xạ:* Nếu $Y \subseteq X$ Thì $X \rightarrow Y$
  - *Tăng trưởng:* Nếu $X \rightarrow Y$ Thì $XZ \rightarrow YZ$
  - *Bắc cầu:* Nếu $X \rightarrow Y$ và $Y \rightarrow Z$ Thì $X \rightarrow Z$

- 3 luật trên có thể suy diễn ra các luật sau:

  - *Hợp:* Nếu $X \rightarrow Y$ và $X \rightarrow Z$, Thì $X \rightarrow YZ$
  - *Giả bắc cầu:* Nếu $X \rightarrow Y$ và $YZ \rightarrow W$, Thì $XZ \rightarrow W$
  - *Phân rã:* Nếu $X \rightarrow YZ$, Thì $X \rightarrow Y$ và $X \rightarrow Z$

Sử dụng hệ tiên đề Armstrong

- Sử dụng hệ tiên đề Armstrong để suy diễn một phụ thuộc hàm mới từ một tập các phụ thuộc hàm cho trước
- Ví dụ: Cho quan hệ R với tập PTH F như sau:

$$F = \{ \quad A \rightarrow C,\quad B \rightarrow H,\quad C,G \rightarrow H,I,\quad A,B \rightarrow I \}$$

- Chứng minh rằng AG → I được suy diễn từ F
  - Ta có: A → C
    C,G → H,I
    => AG → H, I (tựa bắc cầu)
    => AG → I (phân rã)

  => Vậy AG → I được suy diễn từ F

Các tính chất của PTH

F1:	Phản xạ:	Nếu $Y \subseteq X$ Thì $X \rightarrow Y$
F2 :	Bắc cầu:	Nếu $X \rightarrow Y$ và $Y \rightarrow Z$ Thì $X \rightarrow Z$
F3:	Tăng trưởng:	Nếu $X \rightarrow Y$ Thì $XZ \rightarrow YZ$
F4:	Giả bắc cầu:	Nếu $X \rightarrow Y$ và $YZ \rightarrow W$, Thì $XZ \rightarrow W$
F5:	Phản xạ chặt:	$X \rightarrow X$
F6:	Mở rộng vế trái, thu hẹp vế phải	
		Nếu $X \rightarrow Y$ thì $XZ \rightarrow Y \backslash W$ với $Z, W \subseteq U$
F7 :	Cộng tính đầy đủ:	Nếu $X \rightarrow Y$ và $Z \rightarrow W$ thì $XZ \rightarrow YW$
F8 :	Mở rộng vế trái:	Nếu $X \rightarrow Y$ thì $XZ \rightarrow Y$
F9:	Hợp:	Nếu $X \rightarrow Y$ và $X \rightarrow Z$, Thì $X \rightarrow YZ$
F10 :	Phân rã :	Nếu $X \rightarrow YZ$, Thì $X \rightarrow Y$ và $X \rightarrow Z$
F11 :	Tích lũy	Nếu $X \rightarrow YZ$, $Z \rightarrow AW$ thì $X \rightarrow YAW$

Bao đóng (Closure)

- Bao đóng của tập các PTH
- Bao đóng của tập các thuộc tính

Bao đóng của tập các PTH

- Cho lược đồ quan hệ R(U), U={ A1, A2, ..., An}
- F là tập PTH trên R
- Bao đóng của F, ký hiệu F^+ bao gồm:
  - F
  - Và các PTH được suy diễn từ F
- F gọi là đầy đủ nếu $F = F^+$
- Trên thực tế, việc tính F^+ khó thực hiện vì có thể dẫn đến sự bùng nổ tổ hợp
  => Thay vào đó, ta sẽ xét xem một PTH dạng $X \rightarrow Y$ có thuộc F^+ hay không, nghĩa là $X \rightarrow Y$ được suy diễn từ F không ?

Bao đóng của tập các PTH

- Ví dụ:
- Cho F= { AB → C
  BC → D
  D → E, G }
  - Chứng minh rằng AB → E ∈ F^+

- Ta có: AB → C
  BC → D
  => AB → D (tựa bắc cầu)
  và ta có D → E,G => D → E (phân rã)
  => AB → E (bắc cầu)

=> Vậy AB → E ∈ F^+ hay AB → E được suy diễn từ F

Bao đóng của tập thuộc tính

- Cho lược đồ quan hệ R(U), U = {A1, A2, ..., An}
- X là tập các thuộc tính trên U, F là tập các PTH trên R

- Bao đóng của tập thuộc tính X đối với F, ký hiệu X^+ bao gồm tập các thuộc tính PTH vào X, nghĩa là:
  - $X^+ = \{ A \in U \mid X \rightarrow A \in F^+ \}$

- Nhận xét:
  - Làm thế nào biết được một PTH X → Y có được suy diễn từ F không ?
    - $X \rightarrow Y \in F^+ \iff X^+ \supseteq Y$
  - Nếu $X^+ = U$ thì X là siêu khóa của R

Thuật toán tìm X^+

- Dữ liệu vào: lược đồ quan hệ R với tập thuộc tính U, tập PTH F và $X \subseteq U$
- Dữ liệu ra : X^+
- Giải thuật:
  - Bước 1: $X^+ = X$
  - Bước 2: Nếu tồn tại $A \rightarrow B \in F$ và $A \subseteq X^+$ thì :

    $$X^+ = X^+ \cup B$$
  - Lặp lại bước 2 cho đến khi không thể thêm thuộc tính cho X^+ hoặc tất cả các PTH đã được xét.
  - Bước 3 : Kết quả là X^+

Thuật toán tìm X^+

- Ví dụ: cho tập PTH

  F= { A → C

  B → H

  C,G → H,I

  A,B → I }
  - Tìm $(A)^+$ $(AB)^+$
  - PTH A → H có suy diễn được từ F không ?

    Hay A → H có thuộc F^+ không ?
  - PTH AB → CH có suy diễn được từ F không ?

    Hay AB → CH có thuộc F^+ không ?

- Tìm $(A)^+$
  - Bước 1: khởi tao

    $A^+ = A$
  - Bước 2: lặp
    - $A^+ = A + C = AC$ $(A \rightarrow C)$
    - $A^+ = AC + \varnothing$
  - B3: $A^+ = AC$

- Ta có $A^+ = AC$ không chứa H

  => $A \rightarrow H \notin F^+$

Thuật toán tìm X^+

- Ví dụ: cho tập PTH

  F= { A → C
  B → H
  C,G → H,I
  A,B → I }
  - Tìm $(A)^+$ $(AB)^+$
  - PTH A → H có suy diễn được từ F không ?
    Hay A → H có thuộc F^+ không ?
  - PTH AB → CH có suy diễn được từ F không ?
    Hay AB → CH có thuộc F^+ không ?

- Tìm $(AB)^+$
  - Bước 1
    $(AB)^+$= AB
  - Bước 2 lặp
    - $(AB)^+$= AB + C + H + I
      = ABCHI (A → C, B → H, AB → I)
    - AB^+= ABCHI + ∅
  - B3: AB^+= ABCHI
- Tương tự ta có: $(AB)^+$ = ABCHI

Ta có: $(AB)^+$= ABCHI ⊇ CH => AB → CH ∈ F^+

Phủ tối tiểu

- Tập PTH tương đương
  - Hai tập PTH F và G trên cùng lược đồ quan hệ là tương đương nếu và chỉ nếu $F^+ = G^+$

Phủ tối tiểu

- Tập PTH tối tiểu: F được gọi là tối tiểu nếu F thỏa các điều kiện sau:
  - Mọi PTH trong F chỉ có một thuộc tính ở vế phải
  - Không tồn tại PTH thừa
  - Không tồn tại PTH mà vế trái của nó có thuộc tính thừa

  => Phủ tối tiểu của tập PTH F là tập F' tương đương với F

  => Mọi tập PTH đều có ít nhất một tập PTH tối tiểu

Phủ tối tiểu

- Phụ thuộc hàm thừa trong một tập các PTH
  - Một PTH $X \rightarrow Y$ được gọi là PTH thừa trong tập PTH F nếu và chỉ nếu
    - F tương đương $F \backslash \{X \rightarrow Y\}$
    - Hay $\{X \rightarrow Y\}$ được suy diễn từ $F \backslash \{X \rightarrow Y\}$
  - Ví dụ: $F=\{ A \rightarrow B, B \rightarrow C, D \rightarrow BE, A \rightarrow C \}$
    - PTH thừa ???

=> $A \rightarrow C$ vì PTH này được suy diễn từ hai PTH $A \rightarrow B$ và $B \rightarrow C$

Phủ tối tiểu

- Thuộc tính thừa ở vế trái
  - Xét PTH có dạng $X_iX_j \rightarrow Y$
  - Thuộc tính X_i ở vế trái của PTH $X_iX_j \rightarrow Y$ được gọi là thuộc tính thừa nếu trong F thay $X_iX_j \rightarrow Y$ bằng $X_j \rightarrow Y$ thì F^+ vẫn không thay đổi
    - Hay $F^+ = (F \setminus \{X_iX_j \rightarrow Y\} \cup \{X_j \rightarrow Y\})^+$
  - Ví dụ: $F=\{ A \rightarrow B, AB \rightarrow C, D \rightarrow BE\}$
    - Thuộc tính thừa ???

=> B trong PTH $AB \rightarrow C$ vì $A \rightarrow B$

PTH rút gọn tự nhiên

- Tập PTH F được gọi là rút gọn tự nhiên nếu
  - Vế phải và vế trái của mọi PTH không có thuộc tính chung
    - Nếu có thuộc tính chung thì bỏ thuộc tính chung đó ở vế phải
  - Hai PTH khác nhau có vế trái khác nhau
    - Nếu có hai PTH cùng vế trái dạng $L_1 \rightarrow R_1$ và $L_1 \rightarrow R_2$ thì gom hai PTH đó lại thành $L_1 \rightarrow R_1R_2$
- Ví dụ: Tìm tập PTH rút gọn tự nhiên của F

F= {AD → CD

B → H

C,G → H

C,G → I

A,B → I }

=> F_{TN}= {AD → C

B → H

C,G → H,I

A,B → I }

Bài toán tìm phủ tối tiểu

- Các bước tìm tập PTH tối tiểu của F:
  - Tách các PTH sao cho VP có 1 thuộc tính (dùng phân rã)
  - Loại bỏ các PTH thừa
  - Loại bỏ các thuộc tính thừa ở VT
  - Tìm tập PTH rút gọn tự nhiên của F

Ví dụ Phủ tối tiểu

- Cho F={ A → B, B → C, D → BE, A → C}, tìm F tối tiểu ?

1) Tách các PTH sao cho VP có 1 thuộc tính
   - F={ A → B,
     B → C,
     D → B,
     D → E,
     A → C}
2) Loại bỏ PHT thừa
   - Ta có: A → B và B → C => A → C (thừa => loại PTH này)
3) Không có thuộc tính thừa ở VT
4) F rút gọn tự nhiên
   - F={ A → B,
     B → C,
     D → BE }

Ví dụ

- Tìm tập PTH tối tiểu của:

  $F=\{A \rightarrow BC, B \rightarrow CE, A \rightarrow E, AC \rightarrow H, D \rightarrow B\}$

Ví dụ

- Tìm tập PTH tối tiểu của:

  $F=\{A \rightarrow BC,\ B \rightarrow CE,\ A \rightarrow E,\ AC \rightarrow H,\ D \rightarrow B\}$

- B1: $F=\{A \rightarrow B,$

  $A \rightarrow C,$

  $B \rightarrow C,$

  $B \rightarrow E,$

  $A \rightarrow E,$

  $AC \rightarrow H,$

  $D \rightarrow B\}$

B2

$F=\{A \rightarrow B,$

$B \rightarrow C,$

$B \rightarrow E,$

$AC \rightarrow H,$

$D \rightarrow B\}$

Ta có: $A \rightarrow B,\ B \rightarrow C \Rightarrow A \rightarrow C$ (thừa)

$A \rightarrow B,\ B \rightarrow E \Rightarrow A \rightarrow E$ (thừa)

Ví dụ

- $F=\{A \to BC, B \to CE, A \to E, AC \to H, D \to B\}$
- B3: $F=\{A \to B,$

$B \to C,$

$B \to E,$

$AC \to H,$

$D \to B\}$

$F=\{A \to B,$

$B \to C,$

$B \to E,$

$A \to H,$

$D \to B\}$

B4

$F=\{A \to BH,$

$B \to CE,$

$D \to B\}$

là tập tối tiểu

Ta có $AC \to H$ có vế trái có 2 thuộc tính

Do $A \to B$ và $B \to C$ => $A \to C$

=> thuộc tính C trong $AC \to H$ thừa

Bài tập

- **Bài 1**: Cho quan hệ sau KHACHHANG(id, ten, tpho) như sau
  - Các phát biểu sau đúng hay sai:

001	Albert	Bruxelles
002	Francois	Liege
003	Brabo	Anvers
004	Albert	Anvers
005	Leon	Liege
006	Philippe	Bruxelles
007	Brabo	Anvers

1)Id → ten

2)id → tpho

3)ten → id

4)ten → tpho

5)tpho → id

6)id → tpho, ten

Bài tập

- **Bài 2**: Cho R(A, B, C, D, E) và F= {AB → C, B → D, CD → E}
  - Tính $(BC)^+$
  - Chứng minh rằng F^+ chứa AB → E
  - Chứng minh rằng AB là khóa

Bài tập

- **Bài 2**: Cho R(A, B, C, D, E) và F= {AB → C, B → D, CD → E}
  - Tính $(BC)^+$ = BC
    - $(BC)^+$ = BC +D=BCD (B → D)
    - $(BC)^+$ =BCD + E = BCDE (CD → E)
    - $(BC)^+$ =BCDE + $\varnothing$ = BCDE
  - Chứng minh rằng F^+ chứa AB → E
    - (AB)+= ABCDE $\supseteq$ E => AB → E $\in F^+$
  - Chứng minh rằng AB là khóa

    Ta có (AB)+= ABCDE =U => AB là siêu khoá
    - Mặt khác: A+=A => A không là SK

      B+=BD => B không là SK

AB là Sk nhỏ nhất
=> AB là khoá

Bài tập

- **Bài 3** :Cho quan hệ

  KhamBenh(idBN, tenBN, idBsi, tenBsi, ngaykham, loaibenh)

  Một bác sĩ có thể khám nhiều bệnh liên quan nhiều bệnh nhân

  – Xác định tất cả các PTH của quan hệ KhamBenh biét rằng mỗi bệnh nhân có thể khám nhiều lần trong ngày nhưng không quá 1 lần với cùng bác sĩ.

Bài tập

- **Bài 4**:

  Cho F = {$AB \rightarrow C$, $A \rightarrow D$, $D \rightarrow E$, $AC \rightarrow B$}

  Hai PTH $AB \rightarrow E$ và $D \rightarrow C$ có được suy diễn từ F hay không?