TẠP CHÍ KHOA HỌC TÍNH TOÁN VÀ ĐIỀU KHIỂN, Tập III (1987), Số 2 (1–3; 8)

PHÉP TÁCH SUY BIẾN CỦA MỘT QUAN HỆ TRONG CƠ SỞ DỮ LIỆU

LÊ TIẾN VƯƠNG

1. Nhập đề.

Hai vấn đề trung tâm của lý thuyết cơ sở dữ liệu quan hệ (CSDL) là tối ưu hóa việc xử lý các câu hỏi và vấn đề tách. Để tối ưu hóa cần sử dụng các phép tính của đại số quan hệ, trong khi tách chỉ cần phép chiếu. Gần đây/BFMY83/ đã đưa ra một lớp tách phi chu trình và khảo sát các tính chất của nó.

Trong bài này chúng ta quan tâm tới một lớp các phép tách suy biến nhằm thu được nhiều thông tin về ngữ nghĩa của dữ liệu. Trong trường hợp đó, thay vì một quan hệ được thiết lập từ các quan hệ cơ sở nhờ phép kết nối tự nhiên, có thể sử dụng phép hợp của các quan hệ. Những CSDL như vậy sẽ bảo đảm việc cập nhật dữ liệu (xóa, bổ sung, sửa đổi) là duy nhất và không gây nên các dị thường.

Giả sử rằng độc giả đã làm quen với CSDL quan hệ. Ở đây chỉ nhắc lại vài khái niệm cơ bản.

Gọi U là một tập các thuộc tính. A là một thuộc tính của U. Với mỗi $A \in U$ có một miền giá trị tương ứng là dom (A). Ánh xạ gán mỗi thành phần A của U một giá trị trong dom (A) gọi là một bộ trên U. Một tập các bộ R được gọi là một quan hệ trên U. Nếu X là một tập con của U và r là một bộ trên U thì hạn chế của ánh xạ r đối với X gọi là X–giá trị của r (ký hiệu là $r[X]$) Nếu R là một quan hệ trên U, $X \subseteq U$ thì tập tất cả các X–giá trị của các bộ thuộc R là một quan hệ trên X và gọi là hình chiếu của R trên X (ký hiệu là $R[X]$).

Gọi X, Y là hai tập con của U. Phụ thuộc hàm (viết tắt là FD) $X \to Y$ thỏa trong R nếu với hai bộ bất kỳ $r_1, r_2 \in R$ sao cho $r_1[X] = r_2[X]$ thì $r_1[Y] = r_2[Y]$.

X, Y, Z là ba tập con của U với $Z = U \setminus XY$ (1). Phụ thuộc đa trị (viết tắt là MVD) $X \to\to Y \mid Z$ thỏa trong R nếu hai bộ bất kỳ $r_1, r_2 \in R$ với $r_1[X] = r_2[X]$ thì tồn tại một bộ $r \in R$ (không nhất thiết phải khác r_1, r_2) sao cho $r[X] = r_1[X] = r_2[X]$, $r[Y] = r_1[Y]$ và $r[Z] = r_2[Z]$.

2. Phép tách – 3 suy biến của một quan hệ.

Một lớp các phép tách đáng quan tâm đóng vai trò quan trọng trong quá trình thiết kế một CSDL là phép tách phi chu trình. Trong/LE83/, /LE84/ đã khảo sát các tính chất chung cho lớp các phép tách–3, tách–n của một quan hệ. Trong bài này chỉ khảo sát một lớp đặc biệt của lớp tách–3.

Định nghĩa 1. Gọi Z, Y, X là ba tập con của U sao cho $XYZ = U$. Quan hệ R xác định trên U gọi là tách–3 được (ternarydecomposable) không mất mát thông tin nếu với ba bộ bất kỳ $r_1, r_2, r_3 \in R$ sao cho

$$r_1[X \cap Y] = r_2[X \cap Y],\ r_2[Y \cap Z] = r_3[Y \cap Z],\ r_3[Z \cap X] = r_1[Z \cap X]$$

thì tồn tại một bộ $r \in R$ sao cho $r[X] = r_1[X]$, $r[Y] = r_2[Y]$ và $r[Z] = r_3[Z]$.

Nếu X, Y, Z khác rỗng và $X \cap Y \neq \phi$, $Y \cap Z \neq \phi$, $X \cap Z = \phi$ thì phép tách–3 này được gọi là phi chu trình.

Từ đây các phép tách của một quan hệ được hiểu là phép tách không mất mát thông tin. Gọi D là họ tất cả các phép tách–3 phi chu trình (ký hiệu là (X, Y, Z)) của quan hệ R trên U. Ta có kết quả sau đây.

(1) Ký hiệu hợp của hai tập hợp $X \cup Y$ viết gọn là XY.

Định lý 1 R là một quan hệ xác định trên U. X, Y, Z là ba tập con bất kỳ khác rỗng của U với $XYZ = U$, $X \cap Y \neq \phi$, $Y \cap Z \neq \phi$ và $X \cap Z = \phi$. $(X, Y, Z) \in D$ khi và chỉ khi tập các phụ thuộc đa trị sau đây thỏa trong R:

$$X \cap Y \rightarrow\rightarrow X,\ (X \cap Y)(X \cap Z) \rightarrow\rightarrow Y \text{ và } Y \cap Z \rightarrow\rightarrow Z.$$

Chứng minh: Dễ dàng kiểm tra bằng định nghĩa 1 và định nghĩa của phụ thuộc đa trị.

Từ định nghĩa 1 và định lý 1 trực tiếp suy ra

Bổ đề 2. Nếu (X, Y, Z) là một phép tách−3 phi chu trình thì (XY, Z), (X, YZ) (XY, YX) là những phép tách−2 của R và (X, Y), (Y, Z) là những phép tách−2 của các hình chiếu tương ứng R[XY] và R[YZ].

Chú ý: Ký hiệu (X, Y) là phép tách−2 của một quan hệ xác định trên tập thuộc tính XY. Đó là trường hợp đặc biệt của lớp tách−3 khi một thành phần của nó (chẳng hạn Z) là tập rỗng (xem /ARDE 80/).

Định nghĩa 2. /ARDE 80/. Phụ thuộc đa trị $X \rightarrow\rightarrow Y|Z$, $z = U \setminus XY$ trong quan hệ R được gọi là suy biến nếu cho mỗi cặp bộ $r_1, r_2 \in R$ sao cho $r_1[X] = r_2[X]$ thì $r_1[Y] = r_2[Y]$ hoặc $r_1[Z] = r_2[Z]$.

Định nghĩa phụ thuộc đa trị suy biến có thể biểu diễn tương đương dưới dạng phép tách như sau:

Định nghĩa 3. R là một quan hệ xác định trên U. (X, Y) là phép tách−2 của R. (X, Y) được gọi là phép tách−2 suy biến nếu tồn tại hai quan hệ R_1, R_2 sao cho $R = R_1 \cup R_2$ và $X \cap Y \rightarrow X$ thỏa trong R_1, $X \cap Y \rightarrow Y$ thỏa trong R_2 với $R_1[X \cap Y] \cap R_2[X \cap Y] = \phi$.

Trong /ARDE 80/ các tác giả đã nghiên cứu phép tách−2 suy biến cho vấn đề cập nhật dữ liệu để không gây mâu thuẫn trong CSDL (tức là đảm bảo tính duy nhất, không gây ra các dị thường khi thực hiện phép xóa, phép bổ sung vào các quan hệ của CSDL). Để tổng quát hóa kết quả trên, ở đây đưa ra khái niệm phép tách−3 suy biến như sau:

Định nghĩa 4. Phép tách −3 phi chu trình $(X, Y, Z) \in D$ của quan hệ R trên U được gọi là suy biến nếu tồn tại hai quan hệ R_1, R_2 sao cho $R = R_1 \cup R_2$ và $X \cap Y \rightarrow YZ$ thỏa trong R_1, $Y \cap Z \rightarrow XY$ thỏa trong R_2 với $R_1[X \cap Y] \cap R_2[X \cap Y] = \phi$ hoặc $R_1[Y \cap Z] \cap R_2[Y \cap Z] = \phi$.

Sau đây luôn giả thiết rằng $(X, Z, Y) \in D$, tức là một phép tách−3 phi chu trình của quan hệ R trên U. Phép tách này có những tính chất sau đây:

Bổ đề 3. Nếu (X, Y, Z) là phép tách−3 suy biến của R thì (X, YZ), (XY, Z), (XY, YZ) là những phép tách−2 suy biến của R và (X, Y), (Y, Z) là những phép tách−2 suy biến trên hình chiếu tương ứng R[XY] và R[YZ].

Chứng minh: Vì (X, Y, Z) là suy biến cho nên tồn tại hai quan hệ R_1 và R_2 để $R = R_1 \cup R_2$ và $X \cap Y \rightarrow YZ$ thỏa trên R_1, $Y \cap Z \rightarrow XY$ thỏa trên R_2. Không làm mất tính tổng quát giả sử rằng $R_1[X \cap Y] \cap R_2[X \cap Y] = \phi$.

− Trước hết chứng minh rằng (X, YZ) là tách−2 suy biến. Trên R_1 đã thỏa $X \cap Y \rightarrow YZ$. Chỉ cần chỉ ra rằng trên R_2 thỏa $X \cap Y \rightarrow X$. Thật vậy, vì (X, YZ) là phép tách − 2 và $R_1[X \cap Y] \cap R_2[X \cap Y] = \phi$ nên các phụ thuộc dữ liệu sau đây cũng thỏa trên R_2:

$$X \cap Y \rightarrow\rightarrow X,\ Y \cap Z \rightarrow XY, \text{ do đó } Y \cap Z \rightarrow X.$$

Áp dụng luật suy dẫn kết hợp FD−MVD2 (2) suy ra $X \cap Y \rightarrow X$. Do đó (X, YZ) là suy biến.

− Để chứng minh cho (XY, Z) là suy biến theo giả thiết trên, ta xây dựng hai quan hệ R'_1 và R'_2 như sau:

Vì (XY, Z) là phép tách−2 cho nên luôn thỏa: Với mọi cặp $r_1, r_2 \in R$ sao cho $r[X \cap Y] = r_2[Y \cap Z]$ thì luôn tồn tại một bộ $r \in R$ sao cho $r[XY] = r_1[XY]$ và $r[Z] = r_2[Z]$. Ta có các trường hợp sau đây:

a) Nếu $r_1, r_2 \in R_1$ dễ dàng suy ra $r \in R_1$

Nếu $r_1, r_2 \in R_2$ dễ dàng suy ra $r \in R_2$.

(2) Luật FD − MVD2: nếu $X \rightarrow\rightarrow Z$, $Y \rightarrow Z'$, ; $Z' \subseteq Z$, $Y \cap Z = \phi$ thì $X \rightarrow Z'$

b) Không làm mất tính tổng quát, giả sử rằng $r_1 \in R_1$ và $r_2 \in R_2$. Khi đó nếu $r \in R$ ta có $r[XY] = r_1[XY]$, $r[Z] = r_2[Z]$, tức là $r[X \cap Y] = r_1[X \cap Y]$. Điều đó mâu thuẫn với giả thiết $R_1[X \cap Y] \cap R_2[X \cap Y] = \phi$. Vậy chỉ còn $r \in R_1$. Vì rằng $r[XY] = r_1[XY]$ và $X \cap Y \to YZ$ thỏa trên R_1 cho nên $r \equiv r_1$ trên R_1 và ta có $r_1[Z] = r_2[Z]$. Như vậy nếu có hai bộ $r_1 \in R_1$ và $r_2 \in R_2$ với $r_1[Y \cap Z] = r_2[Y \cap Z]$ thì luôn thỏa $r_1[Z] = r_2[Z]$. Nhóm tất cả các bộ có tính chất như trên của R_2 thành tập T. Thiết lập hai quan hệ R'_1 và R'_2 như sau:

$$R'_1 = R_1 \cup T \text{ và } R'_2 = R_2 \setminus T.$$

Rõ ràng $R'_1[Y \cap Z] \cap R'_2[Y \cap Z] = \phi$ và trên R'_1 thỏa $Y \cap Z \to Z$, trên R'_2 thỏa $Y \cap Z \to XY$. Do vậy (XY, Z) là suy biến.

– Các trường hợp còn lại chứng minh dễ dàng. Bổ đề được chứng minh hoàn toàn.

Bổ đề 4. Nếu (XY, Z), (X, YZ) là những phép tách −2 suy biến của R thì (X, Y), (Y, Z) là những phép tách −2 suy biến của các hình chiếu tương ứng $R[XY]$ và $R[YZ]$.

Chứng minh: dễ dàng.

Bổ đề 5. (XY, Z) là phép tách −2 suy biến của R, (X, Y) là phép tách −2 suy biến của hình chiếu $R[XY]$ khi và chỉ khi (X, YZ) là phép tách −2 suy biến của R và (Y, Z) là phép tách −2 suy biến của hình chiếu $R[YZ]$.

Chứng minh: Từ (XY, Z) là suy biến, theo bổ đề 4 ta có (Y, Z) là suy biến. Cần chứng minh rằng (X, YZ) cũng là suy biến.

Vì (XY, Z) suy biến của R nên tồn tại hai quan hệ R_1 và R_2 sao cho $R = R_1 \cup R_2$ và $Y \cap Z \to XY$ thỏa trong R_1, $Y \cap Z \to Z$ thỏa trong R_2 với $R_1[Y \cap Z] \cap R_2[Y \cap Z] = \phi$. Cần xây dựng hai quan hệ R'_1, R'_2 từ R_1 và R_2 để thỏa các điều kiện của phép tách −2 suy biến (X, YZ).

Vì (X, XZ) là phép tách −2 của R do đó với mọi r_1, $r_2 \in R$ sao cho $r_1[X \cap Y] = r_2[X \cap Y]$ thì luôn tồn tại một bộ $r \in R$ sao cho $r[X] = r_1[X]$ và $r[YZ] = r_2[YZ]$. Ta có các trường hợp sau đây:

– Trường hợp 1: Nếu $r_1, r_2 \in R_1$ thì dễ dàng suy ra $r \in R_1$

– Trường hợp 2: Tương tự nếu $r_1, r_2 \in R_2$ thì $r \in R_2$

– Trường hợp 3: Không làm mất tính tổng quát, giả sử rằng $r_1 \in R_1$, $r_2 \in R_2$. Nếu $\in R_1$ thì $r[Y \cap Z] = r_2[Z \cap Z]$. Điều đó trái với giả thiết là $R_1[Y \cap Z] \cap R_2[Y \cap Z] = \phi$. Do đó, trong trường hợp này chỉ còn $r \in R_2$. Vì (X, Y) là phép tách −2 suy biến, do đó r_1, r_2 phải thỏa: nếu $r_1[X \cap Y] = r_2[X \cap Y]$ thì $r_1[X] = r_2[X]$ hoặc $r_1[Y] = r_2[Y]$. Nhưng do (XY, YZ) cũng là suy biến cho nên $r_1[Y] \neq r_2[Y]$, vậy $r_1[X] = r_2[X]$. Điều đó chứng tỏ rằng $r \equiv r_2$ trong R_2. Nếu tồn tại một bộ $r \in R_2$ sao cho $r[X \cap Y] = r'[X \cap Y]$, $r' \in R_1$ thì luôn có $r[X] = r'[X]$.

Nhóm tất cả những bộ có tính chất trên của R_2 và những bộ thuộc R_2 mà bằng nhau trên X nhưng khác nhau trên Y tạo thành tập T. Ta thiết lập hai quan hệ R'_1 và R'_2 như sau:

$$R'_1 = R_1 \cup T \quad \text{và} \quad R'_2 = R_2 \setminus T.$$

Khi đó $R = R'_1 \cup R'_2$.

Từ cách xây dựng R'_1 và R'_2 rõ ràng $R'_1[X \cap Y] \cap R'_2[X \cap Y] = \phi$ và $X \cap Y \to X$ thỏa trong R'_1, $X \cap Y \to YZ$ thỏa trong R'_2. Do đó (X, YZ) là suy biến.

Chiều ngược lại chứng minh hoàn toàn tương tự. Bổ đề được chứng minh xong.

Nhận ngày 1-10-1986

(xem tiếp trang 8)