Algoritmo di decomposizione

Introduction

Qui mostreremo che dato uno schema di relazione R e un insieme di dipendenze funzionali F su R esiste sempre una decomposizione $\rho = \{R_1, R_2, \ldots, R_k\}$ di R tale che:

- per ogni i, $i = 1, \ldots, k$, R_i è in 3NF
- ρ preserva F
- ρ ha un join senza perdita
- tale decomposizione può essere calcolata in tempo polinomiale

Come si fa?

Il seguente algoritmo, dato uno schema di relazione R e un insieme di dipendenze funzionali F su R, che è una copertura minimale, permette di calcolare in tempo polinomiale una decomposizione $\rho\{R_1, R_2, \ldots, R_k\}$ di R che rispetta le condizioni sopraelencate

Ci interessa una qualunque copertura minimale dell'insieme di dipendenze funzionali definite sullo schema R. Se ce ne fosse più di una, con eventualmente cardinalità diversa, potremmo scegliere ad esempio quella con meno dipendenza, ma questo non è tra i nostri scopi.
Quindi per fornire l'input all'algoritmo di decomposizione è sufficiente trovarne una tra quelle possibili. Poi vedremo perché ci occorre che sia una copertura minimale

---

Algoritmo per la decomposizione di uno schema

Input uno schema di relazione R e un insieme F di dipendenze funzionali su R, che è una copertura minimale
Output una decomposizione ρ di R che preserva F e tale che per ogni schema di relazione in ρ è in 3NF

begin
$S := \varnothing$
for every $A \in R$ tale che A non è coinvolto in nessuna dipendenza funzionale in F
  do
  $S := S \cup \{A\}$
if $S \neq \varnothing$ **then**
  begin
  $R := R - S$
  $\rho := \rho \cup \{S\}$
  end
if esiste una dipendenza funzionale in F che coinvolge tutti gli attributi in R
  then $\rho := \rho \cup \{R\}$
else
  for every $X \to A$
    do
    $\rho := \rho \cup \{XA\}$
end

> **Hint**
>
> **if** esiste una dipendenza funzionale in F che coinvolge tutti gli attributi in R
>
> - R residuo dopo aver eventualmente eliminato gli attributi inseriti prima in S
>
> **then** $\rho := \rho \cup \{R\}$
>
> - in questo caso ci fermiamo anche se la copertura minimale contiene anche altre dipendenze; in altre parole la copertura minimale potrebbe contenere anche altre dipendenze

---

Teorema

Sia R uno schema di relazione ed F un insieme di dipendenze funzionali su R, che è una copertura minimale. L'algoritmo di decomposizione permette di calcolare in tempo polinomiale una decomposizione ρ di R tale che:

- ogni schema di relazione ρ è in 3NF
- ρ preserva F

> **Dimostrazione**

Dimostriamo separatamente le due proprietà della decomposizione

ρ preserva F

Sia $G = \cup_{i=1}^{k} \pi_{R_i}(F)$, ovvero l'insieme delle dipendenze di F^+ tali che il determinante e il determinato appartengono al sottoschema.
Poiché per ogni dipendenza funzionale $X \to A \in F$ si ha che $XA \in \rho$ (è proprio uno dei sottoschemi), si ha che questa dipendenza di F sarà sicuramente in G, quindi $F \subseteq G$ e, quindi $F^+ \subseteq G^+$. L'inclusione $G^+ \subseteq F^+$ è banalmente verificata in quanto per definizione, $G \subseteq F^+$

Ogni schema di relazione in ρ è in 3NF

Analizziamo i diversi casi che si possono presentare

1. Se $S \in \rho$, ogni attributo in S (elementi non coinvolti nelle dipendenze, e siccome la chiave deve determinare tutto lo schema, dovranno necessariamente essere nella chiave che li determinerà per riflessività) fa parte della chiave e quindi, banalmente, S è in 3NF
2. Se $R \in \rho$ esiste una dipendenza funzionale in F che coinvolge tutti gli attributi in R. Poiché F è una copertura minimale tale dipendenza avrà la forma $R - A \to A$. Ma se fosse esistito $Y \to A$ con $Y \subset R - A$ allora nella copertura non ci sarebbe stato $R - A \to A$
3. Se $XA \in \rho$, poiché F è una copertura minimale, non ci possono essere una dipendenza funzionale $X' \to A \in F^+$ tale che $X' \subset X$ e, quindi, X è chiave in XA. Sia $Y \to B$ una qualsiasi dipendenza in F tale che $YB \subseteq XA$; se $B = A$ allora, poiché F è una copertura minimale, $Y = X$ (cioè, Y è superchiave); se $B \neq A$ allora $B \in X$ e quindi B è primo

---

Teorema

Sia R uno schema di relazione, F un insieme di dipendenze funzionali su R, che è una copertura minimale e ρ la decomposizione di R prodotta dall'algoritmo di decomposizione. La decomposizione $\sigma = \rho \cup \{K\}$, dove K è una chiave per R, è tale che:

- ogni schema di relazione in σ è in 3NF
- σ preserva F
- σ ha un join senza perdita

Dimostrazione

σ preserva F

Poiché ρ preserva F anche σ preserva F
Stiamo aggiungendo un nuovo sottoschema, quindi alla nuova G' dobbiamo aggiungere una proiezione di F, cioè $G' = G \cup \pi_K(F)$ quindi $G' \supseteq G \supseteq F$ e quindi $G'^+ \supseteq G^+ \supseteq F^+$
L'inclusione $G'^+ \subseteq F^+$ è di nuovo banalmente verificata in quanto, per definizione, $G \subseteq F^+$

Ogni schema di relazione in σ è in 3NF

Poiché $\sigma = \rho \cup \{K\}$, è sufficiente verificare che anche lo schema di relazione K è in 3NF. Mostriamo che K è chiave anche per lo schema K.
Supponiamo per assurdo che K non sia chiave per lo schema K; allora esiste un sottoinsieme proprio K' di K che determina tutto lo schema K, cioè tale che $Kì \to K \in F^+$ (più precisamente alla chiusura di $\pi_K(F)$, ma poiché $\pi_K(F) \subset F^+$ allora $(\pi_K(F))^+ \subset F^+$).
Poiché K è chiave per lo schema R, $K \to R \in F^+$, pertanto per transitività $K' \to R \in F^+$, che contraddice il fatto che K è chiave per lo schema R (verrebbe violato il requisito di minimalità).
Pertanto K è chiave per lo schema K e quindi per ogni dipendenza funzionale $X \to A \in F^+$ con $XA \subseteq K$, A è primo

---

Esempi

Esempio 1

$$R = (A, B, C, D, E, H)$$

$$F = \{AB \to CD, C \to E, AB \to E, ABC \to D\}$$

Rispondere ai seguenti quesiti:

- Verificare che ABH è una chiave per R
- Sapendo che ABH è l'unica chiave per R, verificare che R non è in 3NF
- Trovare una copertura minimale G di F
- Trovare una decomposizione ρ di R tale che preserva G e ogni schema in ρ è in 3NF
- Trovare una decomposizione σ di R tale che preserva G, ha un join senza perdita e ogni schema σ è in 3NF

Verificare che ABH è una chiave per R

Vuol dire verificare due condizioni:

- ABH deve determinare funzionalmente l'intero schema
- Nessun sottoinsieme di ABH deve determinare funzionalmente l'intero schema

Per la prima condizione si controlla se la chiusura dell'insieme di attributi ABH contiene tutti gli attributi dello schema. Notiamo infatti che
$(ABH)^+ = \{A, B, C, D, E, H\}$
Per la seconda condizione, dobbiamo che la chiusura di nessun sottoinsieme di ABH contenga tutti gli attributi dello schema. A tal proposito notiamo che H deve comparire in ogni caso in una chiave dello schema. Quindi ci restano da controllare le chiusure di AH e BH, ma è banale

Possiamo quindi concludere che ABH è chiave dello schema dato

Verificare che R non è in 3NF

Per verificare che lo schema non è in terza forma normale, basta osservare la presenza delle dipendenze parziali $AB \to CD$ e $AB \to E$

Trovare una copertura minimale G di F

1. Riduciamo le parti destre a singleton
   $F = \{AB \to C, AB \to D, C \to E, AB \to E, ABC \to D\}$
2. Verifichiamo se nelle dipendenze ci sono ridondanze nelle parti sinistre
   Cominciamo dalla dipendenze $AB \to C$ controlliamo se C appartiene a $(A)^+_F$ o $(B)^+_F$, ma $(A)^+_F = \{A\}$ e $(B)^+_F = \{B\}$
   Proviamo quindi a ridurre $ABC \to D$. Abbiamo $(AB)^+_F = ABCDE$, dunque

D è contenuto nella chiusura di AB che ci permette di rimuovere C, ma così facendo notiamo che $AB \to D$ è ridondante quindi la rimuovo
Ho infine $F = \{AB \to C, AB \to D, C \to E, AB \to E\}$

3. Verifichiamo che l'insieme non contiene dipendenze ridondanti
Notiamo prima di tutto che C e D sono determinate da una sola dipendenza quindi $AB \to C$ e $AB \to D$ non possono essere rimosse. Però notiamo che provando ad eliminare $AB \to E$ avremmo che questa dipendenza risulta comunque rispettata grazie alla transitività $AB \to C \wedge C \to E$

Abbiamo quindi $G = \{AB \to C, AB \to D, C \to E\}$

Applicare algoritmo di decomposizione

Applichiamo l'algoritmo per la decomposizione dello schema di relazione che non è in 3NF dato l'insieme G che è una copertura minimale

Al primo passo dobbiamo inserire in un elemento della decomposizione gli attributi che non compaiono nelle dipendenze di G. Vale per l'attributo H, quindi avremo $\rho = \{H\}$

Passiamo poi a verificare che non ci sono dipendenze che non ci sono dipendenze che coinvolgono tutti gli attributi dello schema, per cui eseguiamo il passo alternativo. Abbiamo alla fine $\rho = \{H, ABC, ABD, CE\}$

Per avere una decomposizione con join senza perdita, aggiungiamo alla decomposizione precedente un sottoschema che contenga la chiave ABH (**è sufficiente una chiave qualsiasi**), che non è contenuta in alcuno degli schemi ottenuti

Si ha dunque che $\sigma = \{H, ABC, ABD, CE, ABH\}$

Esempio 2

$$R = (A, B, C, D, E)$$

$$F = \{AB \to C, B \to D, D \to C\}$$

Rispondere ai seguenti quesiti:

- Verificare che R non è in 3NF
- Fornire una decomposizione di R tale che:
  - ogni schema della decomposizione è in 3NF

- la decomposizione preserva F
- la decomposizione ha un join senza perdite

Verificare che R non è in 3NF

Notiamo che l'attributo E deve far parte della chiave perché non è determinato da nessuno e inoltre che C non compare mai a sinistra (non determina nessun attributo) quindi non farà parte della chiave

Calcoliamo $(AB)^+_F = \{A, B, C, D\}$. Manca solo la E infatti $(ABE)^+_F = \{A, B, C, E\} = R$

In F ci sono delle dipendenze parziali $AB \to C$, $B \to D$ quindi F non è in 3NF

Copertura minimale di F

1. Tutte le parti destre sono già singleton
2. Verifichiamo se nelle dipendenze ci sono ridondanze nelle parti sinistre
   Proviamo a ridurre $AB \to C$: $(A)^+_F = \{A\}$, $(B)^+_F = \{B, D, C\}$ che contiene C quindi la dipendenza si può ridurre in $B \to C$
   Abbiamo quindi $F = \{B \to C, B \to D, D \to C\}$
3. Verifichiamo se ci sono dipendenze ridondanti
   Cominciamo rimuovendo $B \to C$ e vediamo se viene comunque rispettata; ciò avviene per transitività $B \to D \wedge D \to C$ quindi è possibile rimuoverla

Abbiamo quindi $G = \{B \to D, D \to C\}$

Decomposizione

Al primo passo inseriamo in un elemento della decomposizione gli attributi che non compaiono nelle dipendenze di G quindi avremo $\rho = \{AE\}$. Passiamo poi a verificare che non ci sono dipendenze che coinvolgono tutti gli attributi dello scehma, per cui eseguiamo il passo alternativo. Abbiamo alla fine
$\rho = \{AE, BD, DC\}$
Per avere una decomposizione con join senza perdita, aggiungiamo alla decomposizione precedente un sottoschema che contenga la chiave (che non è già contenuta in alcuno degli schemi ottenuti)
Abbiamo quindi $\sigma = \{AE, BD, DC, ABE\}$

Esempio 3

$$R = (A, B, C, D, E, H)$$

$$F = \{D \to H, B \to AC, CD \to H, C \to AD\}$$

Bisogna:

- Determinare l'unica chiave di R
- Dire perché R con l'insieme di dipendenze funzionali F non è in 3NF
- Trovare una decomposizione ρ di R tale che:
  - ogni schema in ρ è in 3NF
  - ρ preserva F
  - ρ ha un join senza perdita

Verificare che lo schema non è in 3NF

Notiamo che l'attributo E deve far parte della chiave perché non è determinato da nessuno e inoltre che A non compare mai a sinistra (non determina nessun attributo) quindi non farà parte della chiave
Calcoliamo $(B)^+_F = \{B, A, C, D, H\}$ aggiungendo E abbiamo lo schema R quindi BE è chiave

Copertura minimale di F

1. Parti destre diventano singleton
   $F = \{D \to H, B \to A, B \to C, CD \to H, C \to A, C \to D\}$
2. Verifichiamo se nelle dipendenze ci sono ridondanze nelle parti sinistre
   Proviamo a ridurre $CD \to H$. Notiamo che in $(D)^+_F$ compare H quindi posso eliminare C ma così facendo ho una dipendenza duplicata, quindi la elimino del tutto
   Abbiamo quindi $F = \{D \to H, B \to A, B \to C, C \to A, C \to D\}$
3. Verifichiamo se ci sono dipendenze ridondanti
   In questo caso prendo in considerazione solamente $B \to A$ e $C \to A$ in quanto solo le uniche che hanno un determinante duplicato. Noto che togliendo $B \to A$ la dipendenza rimane rispettata per transitività
   $B \to C \land C \to A$

La copertura minimale è quindi $G = \{D \to H, B \to C, C \to A, C \to D\}$

Decomposizione

Al primo passo inseriamo in un elemento della decomposizione gli attributi che non compaiono nelle dipendenze di G quindi avremo $\rho = \{E\}$. Passiamo poi a verificare che non ci sono dipendenze che coinvolgono tutti gli attributi dello scehma, per cui eseguiamo il passo alternativo. Abbiamo alla fine
$\rho = \{E, DH, BC, CA, CD\}$
Per avere una decomposizione con join senza perdita, aggiungiamo alla decomposizione precedente un sottoschema che contenga la chiave (che non è già contenuta in alcuno degli schemi ottenuti)
Abbiamo quindi $\sigma = \{E, DH, BC, CA, CD, BE\}$

Esempio 4

$$R = (A, B, C, D, E, H, I)$$

$$F = \{A \to C, C \to D, BI \to H, H \to I\}$$

Bisogna:

- Trovare le due chiavi dello schema R e spiegare perché sono le chiavi
- Dire perché R con l'insieme di dipendenze funzionali F non è in 3NF
- Trovare una decomposizione ρ di R tale che:
  - ogni schema in ρ è in 3NF
  - ρ preserva F
  - ρ ha un join senza perdita

Verificare che lo schema non è in 3NF

Notiamo che l'attributo E deve far parte della chiave perché non è determinato da nessuno, inoltre che A e B non compaiono mai a destra (non sono determinate da nessun attributo) quindi devono far parte della chiave e infine D non viene determinato da nessuno quindi non farà parte della chiave
Calcoliamo $(ABEI)^+_F = \{A, B, C, D, E, H, I\}$ quindi $ABEI$ è chiave.
Calcoliamo $(ABEH)^+_F = \{A, B, C, D, E, H, I\}$ quindi $ABEH$ è chiave
Non è in 3NF infatti ad esempio $A \to C$ è parziale

Copertura minimale

1. Parti destre diventano singleton
   Parti destre sono già singleton
2. Verifichiamo se nelle dipendenze ci sono ridondanze nelle parti sinistre
   Proviamo a ridurre $BI \to H$. Notiamo che in $(B)^+_F$ compare solo B, calcolo quindi $(I)^+_F = \{I\}$ dunque neanche questa è eliminabile.
3. Verifichiamo se ci sono dipendenze ridondanti
   Notiamo che è inutile provare ad eliminare per tutte vale che la parte destra compare una sola volta

Dunque F è già minimale

Decomposizione

Al primo passo inseriamo in un elemento della decomposizione gli attributi che non compaiono nelle dipendenze di G quindi avremo $\rho = \{E\}$. Passiamo poi a verificare che non ci sono dipendenze che coinvolgono tutti gli attributi dello scehma, per cui eseguiamo il passo alternativo. Abbiamo alla fine
$\rho = \{E, AC, CD, BIH, HI\}$
Per avere una decomposizione con join senza perdita, aggiungiamo alla decomposizione precedente un sottoschema che contenga una delle chiavi ($ABEI$ o $ABEH$).
Abbiamo quindi due possibili decomposizioni
$\sigma_1 = \{E, AC, CD, BIH, HI, ABEH\}$ e $\sigma_1 = \{E, AC, CD, BIH, HI, ABEI\}$