Equivalenza di Espressioni Algebriche

Angelo Montanari

Dipartimento di Matematica e Informatica
Università di Udine

Introduzione

L'equivalenza di espressioni dell'algebra relazionale viene sfruttata nella fase di **ottimizzazione algebrica** delle interrogazioni:

- le interrogazioni formulate nel linguaggio SQL vengono tradotte in espressioni (equivalenti) dell'algebra relazionale
- il **costo dell'esecuzione** di un'interrogazione viene valutato in termini delle dimensioni dei risultati intermedi (della valutazione dell'espressione dell'algebra relazionale);

  in presenza di varie espressioni alternative equivalenti, viene scelta quella con costo minore

Vengono utilizzate delle **trasformazioni di equivalenza**: operazioni che sostituiscono un'espressione con un'altra a essa equivalente

Equivalenze assolute e dipendenti dallo schema

Diversi **tipi di equivalenza**:

(1) Equivalenza **dipendente dallo schema**.

$E_1 \equiv_{\mathbf{R}} E_2$ se $E_1(\mathbf{r}) = E_2(\mathbf{r})$ per ogni $\mathbf{r} \in \mathbf{R}$

(2) Equivalenza **assoluta** (non dipendente dallo schema).

$E_1 \equiv E_2$ se $E_1 \equiv_{\mathbf{R}} E_2$ per ogni schema $\mathbf{R}$

Un esempio di equivalenza assoluta:

$$\Pi_{AB}(\sigma_{A>0}(R_1)) \equiv \sigma_{A>0}(\Pi_{AB}(R_1))$$

Un esempio di equivalenza dipendente dallo schema:

$$\Pi_{AB}(R_1) \bowtie \Pi_{AC}(R_2) \equiv_{\mathbf{R}} \Pi_{ABC}(R_1 \bowtie R_2)$$

che vale se e solo se in $\mathbf{R}$ l'intersezione tra gli insiemi di attributi di R_1 e di R_2 è pari ad A.

Un primo insieme di trasformazioni - 1

Atomizzazione delle selezioni: una congiunzione di selezioni può essere sostituita da una sequenza di selezioni atomiche

$$\sigma_{F_1 \wedge F_2}(E) \equiv \sigma_{F_2}(\sigma_{F_1}(E))$$

con E espressione qualsiasi.

Idempotenza delle proiezioni: una proiezione può essere trasformata in una sequenza di proiezioni che eliminano i vari attributi in varie fasi

$$\pi_X(E) \equiv \pi_X(\pi_{XY}(E))$$

con E espressione definita su un insieme di attributi che contiene X e Y.

Un primo insieme di trasformazioni - 2

Anticipazione della selezione rispetto al join (**pushing selections down**):

$$\sigma_F(E_1 \bowtie E_2) \equiv E_1 \bowtie \sigma_F(E_2)$$

se la condizione F coinvolge solo attributi della sottoespressione E_2.

Anticipazione della proiezione rispetto al join (**pushing projections down**):

$$\pi_{X_1 Y_2}(E_1 \bowtie E_2) \equiv E_1 \bowtie \pi_{Y_2}(E_2)$$

con E_1 definita su X_1, E_2 definita su X_2, $Y_2 \subseteq X_2$ e $(X_1 \cap X_2) \subseteq Y_2$ (gli attributi di $X_2 \setminus Y_2$ non sono coinvolti nel join).

Una trasformazione derivata

Combinando la regola di anticipazione della proiezione rispetto al join con la regola di idempotenza delle proiezioni, si ottiene la seguente **regola derivata** (che elimina da subito gli attributi che non sono coinvolti nel join e non compaiono nel risultato finale):

$$\pi_Y(E_1 \bowtie_F E_2) \equiv \pi_Y(\pi_{Y_1}(E_1) \bowtie_F \pi_{Y_2}(E_2))$$

dove, assumendo che E_1 sia definita su X_1, E_2 sia definita su X_2 e J_1 (rispettivamente J_2) sia il sottoinsieme di X_1 (rispettivamente X_2) coinvolto nell'operazione di join, si ha che:

$Y_1 = (X_1 \cap Y) \cup J_1$

$Y_2 = (X_2 \cap Y) \cup J_2$

Un esempio - 1

Interrogazione: trovare i nomi dei dipartimenti il cui manager guadagna più di 80000 euro.

Soluzione non ottimizzata:

$$\pi_{DNOME}(\sigma_{CF=MANAGER \wedge STIPENDIO>80000}$$
$$(IMPIEGATO \bowtie DIPARTIMENTO))$$

che si può riscrivere come:

$$\pi_{DNOME}(\sigma_{CF=MANAGER}(\sigma_{STIPENDIO>80000}$$
$$(IMPIEGATO \bowtie DIPARTIMENTO)))$$

Un esempio - 2

La precedente espressione può a sua volta essere riscritta nel modo seguente:

$\pi_{DNOME}(\sigma_{STIPENDIO>80000}(IMPIEGATO)$

$\bowtie_{CF=MANAGER}\ DIPARTIMENTO)$

che può a sua volta essere riscritta come:

$\pi_{DNOME}(\pi_{CF}(\sigma_{STIPENDIO>80000}(IMPIEGATO))$

$\bowtie_{CF=MANAGER}\ DIPARTIMENTO)$

P.S. La regola derivata descritta in precedenza potrebbe essere utilizzata per ridurre da subito le dimensioni delle due relazioni..

Un secondo insieme di trasformazioni - 1

Distributività della selezione rispetto all'unione:

$$\sigma_F(E_1 \cup E_2) \equiv \sigma_F(E_1) \cup \sigma_F(E_2)$$

Distributività della selezione rispetto alla differenza:

$$\sigma_F(E_1 \setminus E_2) \equiv \sigma_F(E_1) \setminus \sigma_F(E_2)$$

Distributività della proiezione rispetto all'unione:

$$\pi_X(E_1 \cup E_2) \equiv \pi_X(E_1) \cup \pi_X(E_2)$$

Non vale la distributività della proiezione rispetto alla differenza. Ad esempio, date $R_1(A, B)$ e $R_2(A, B)$

$$\pi_A(R_1 \setminus R_2) \not\equiv \pi_A(R_1) \setminus \pi_A(R_2)$$

non vale se R_1 e R_2 contengono tuple uguali su A e diverse su B

Un secondo insieme di trasformazioni - 2

Trasformazioni basate sulla corrispondenza tra operatori insiemistici e selezioni complesse:

$$\sigma_{F_1 \vee F_2}(R) \equiv \sigma_{F_1}(R) \cup \sigma_{F_2}(R)$$

$$\sigma_{F_1 \wedge F_2}(R) \equiv \sigma_{F_1}(R) \cap \sigma_{F_2}(R) \equiv \sigma_{F_1}(R) \bowtie \sigma_{F_2}(R)$$

$$\sigma_{F_1 \wedge \neg(F_2)}(R) \equiv \sigma_{F_1}(R) \setminus \sigma_{F_2}(R)$$

Proprietà distributiva del join rispetto all'unione:

$$E_1 \bowtie (E_2 \cup E_3) \equiv (E_1 \bowtie E_2) \cup (E_1 \bowtie E_3)$$

N.B. Risultati intermedi vuoti (relazioni prive di tuple) permettono di semplificare le espressioni (join/prodotto cartesiano con la relazione vuota come uno degli operandi producono quale risultato la relazione vuota).