Fondamenti di Teoria delle Basi di Dati

Riccardo Torlone

Parte 7: Tableaux e chase

Tableaux e chase

- La teoria delle dipendenze (e altre proprietà molto interessanti) si possono studiare mediante alcuni interessanti strumenti:
  - Tableau: relazione con variabili che denotano valori nulli
    - può rappresentare una relazione
    - può rappresentare una query
  - Chase: procedura che trasforma un tableau sulla base di un insieme di dipendenze

Tableaux

- Siano D un insieme di costanti e V un insieme di variabili
- Dato un insieme di attributi X, un riga su X è una funzione che associa ad ogni attributo A di X una costante oppure una variabile
- Un tableau su X è un insieme di righe su X

T

Impiegato	Dipartimento	Manager
Bob	v_1	v_2
John	v_1	Jones
Mike	v_3	v_4
v_5	EE	v_6
Tom	EE	v_7

Containment mapping

- Dati due tableaux T_1 e T_2, un containing mapping da T_1 a T_2 è una funzione $\Psi : D \cup V \rightarrow D \cup V$ tale che:
  - Ψ è l'identita sulle costanti
  - estesa Ψ a righe e tableau risulta $\Psi(T_1) \subseteq T_2$
- Due tableaux sono equivalenti se esiste un containing mapping da T_1 e T_2 e un containing mapping da T_2 e T_1

T_1

Impiegato	Dip.	Manager
Bob	v_4	Robinson
John	v_5	Jones
Bob	v_4	Robinson

T_2

Impiegato	Dip.	Manager
Bob	v_4	Robinson
John	v_5	Jones

Relazioni e tableaux

- Un tableau T rappresenta una relazione r se esiste un containing mapping da T a r

r T

Impiegato	Dipartimento	Manager
Bob	CS	Jones
John	CS	Jones
Mike	Math	Smith
Ed	EE	Robinson
Tom	EE	Robinson

Chase

- Una procedura che trasforma un tableau sulla base di un insieme di FD F applicando esaustivamente il seguente passo:
  - Se esistono due righe t_1 e t_2 ed una FD $X \rightarrow A$ in F tale che $t_1[X]=t_2[X]$ e $t_1[A] \neq t_2[A]$ allora uguaglia $t_1[A]$ e $t_2[A]$
- Se il chase cerca di uguagliare due costanti diciamo che ha incontrato una contraddizione
- Il chase termina quando:
  - o non ci sono più modifiche da fare (successo)
  - o quando incontra una contraddizione (fallimento)

Applicazione del chase

T

Impiegato	Dipartimento	Manager
Bob	EE	Robinson
John	v_3	Jones
Bob	EE	Robinson

Impiegato → Dipartimento
Dipartimento → Manager

Altro esempio

T

Impiegato	Dipartimento	Manager
Bob	EE	Robinson
John	v_3	Jones
Bob	EE	Robinson

Impiegato → Dipartimento
Dipartimento → Manager

Risultati di base sul chase

- Il chase è deterministico e indipendente dall’ordine di applicazione del passo base:
  - o si genera una inconsistenza
  - o si generano tableaux equivalenti
- Il chase può essere esteso ad altri tipi di vincoli
  - equality generating dependencies (egd)
    - include le FDs
  - tuple generating dependencies (tgd)

Chase e implicazione

- Data una FD $f{:}X \rightarrow Y$ su un insieme di attributi U e sia T_f un tableau con due righe che hanno:
  - Entrambe la stessa variabile v_A, per ogni A in X
  - Variabili distinte per ogni attributo in $U-X$
- **Teorema** Un insieme di FD F implica una FD $f : X \rightarrow Y$ sse il chase basato su F applicato a T_f produce due righe con i medesimi valori su Y

Esempio di applicazione del teorema

$$F = \{BC \rightarrow DE,\ A \rightarrow B,\ AEG \rightarrow H\}$$
$$F = AC \rightarrow DE$$

T

A	B	C	D	E	G	H
v_A	v_1	v_C	v_2	v_3	v_4	v_5
v_A	v_1	v_C	v_2	v_3	v_9	v_{10}

Dimostrazione del teorema

- **Teorema** Un insieme di FD F implica una FD $f : X \rightarrow Y$ sse il chase basato su F applicato a T_f produce due righe con i medesimi valori su Y.

  Proof Note that $CHASE(T_f)$ satifies F.

  (if) Let r satisfy F and t_1, t_2 be tuples in r such that $t_1[X] = t_2[X]$. Let ψ be a function that maps the two rows in T_f to t_1 and t_2, respectively: ψ is a containing mapping from T_f to $\{t_1, t_2\}$ and also from $CHASE(T_f)$ to $\{t_1, t_2\}$. It follows that $t_1[Y] = t_2[Y]$ and so r satisfy also f.

  (only if) Assume that two rows in $CHASE(T_f)$ do not coincide on Y. Then, any relation that represented by $CHASE(T_f)$ violates f — a contradiction.

Altra applicazione del chase

- Si dice che $r(U)$ ha una decomposizione lossless rispetto a $X_1,\ldots,X_n = U$ se $\bowtie_{j=1}^{n} (\pi_{X_j}(r)) = r$
- Sia F un insieme di FD su U e sia T un tableau su U con n righe $s_1,\ldots,s_n$ tale che, per ogni $1 \le i \le n$ e per ogni $A \in U$,
  - se $A \in X_i$ allora $s_i[A] = v_A$ (variabile distinguibile) altrimenti
  - $s_i[A]$ contiene una variabile distinta da tutte le altre.
- **Teorema** $r(U)$ ha una decomposizione lossless rispetto a $X_1,\ldots,X_n$ sse il chase rispetto a F applicato a T produce una riga con tutte variabili distinguibili

Esempio di applicazione del teorema

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

$$X_1 = AB, X_2 = ACD, X_3 = BCD$$

T

A	B	C	D
v_A	v_B	v_C	v_D
v_A	v_3	v_C	v_D
v_4	v_B	v_C	v_D

Tableaux queries

- Si costruiscono a partire da query congiuntive

$$Q = \{A : x, B : y \mid R(A : x, B : y_1), R(A : x_1, B : y_1), R(A : x_1, B : y)\}$$

$$T_Q: \quad \begin{array}{c|cc} R & A & B \\ \hline & x & y_1 \\ & x_1 & y_1 \\ & x_1 & y \\ \hline & & \end{array}$$

$$Q = (T_Q, \langle x, y \rangle)$$

Altra applicazione del chase

- **Teorema** una query Q_1 è contenuta in una query Q_2 sse c'è un containing mapping tra T_{Q_2} e T_{Q_1}

$$Q = \{A : x, B : y \mid R(A : x, B : y_1), R(A : x_1, B : y_1), R(A : x_1, B : y), R(A : x, B : y)\}$$

T_Q:

R	A	B
	x	y_1
	x_1	y_1
	x_1	y
	x	y

$$Q = (T_Q, \langle x, y \rangle)$$

$T_{Q'}$:

R	A	B
	x	y

$$Q' = (T_{Q'}, \langle x, y \rangle)$$

$$Q' = \{A : x, B : y \mid R(A : x, B : y)\}$$

- Q' è contenuta in Q e viceversa: le query sono equivalenti