

Appunti Basi di Dati

Luca Seggiani

3 Maggio 2024

1 Dipendenze funzionali

Definiamo un metodo per valutare formalmente la qualità della progettazione degli schemi relazionali, ovvero misurare se un raggruppamento di attributi è migliore o peggiore di un altro, senza doversi affidare al buon senso del progettista dello schema o del modello ER. Seguiamo l'approccio top-down: iniziamo dall'individuare raggruppamenti più generali di attributi, ed effettuiamo poi decomposizioni successive per raffinare tali raggruppamenti generali. Gli obiettivi che abbiamo durante lo sviluppo del processo logico sono:

- **Conservazione dell'informazione**, ovvero il mantenimento delle informazioni che ci interessano della realtà in analisi.
- **Minimizzazione della ridondanza**, ovvero la riduzione al minimo della memorizzazione ripetuta della stessa informazione.

Possiamo ricavare da questo alcune linee guida:

- **Linea guida 1: semplice è bello**

Uno schema di relazione deve essere progettato in modo da essere semplice da capire. Non si devono raggruppare attributi da più tipi di entità in un'unica relazione. Intuitivamente, anzi, dovremmo far corrispondere ad ogni schema di relazione una sola entità o una sola relationship, in modo da evitare **ambiguità semantiche**.

- **Linea guida 2: no alle anomalie**

Gli schemi di basi di dati vanno progettati in modo che non si possano presentare anomalie in fase di definizione dei dati. La mancanza di anomalie va certificata attraverso una descrizione fondamentale della semantica dei fatti descritti della realtà di interesse. Se si possono presentare anomalie, esse vanno rilevate e si devi assicurare che i programmi che aggiornano la base operino correttamente e in sicurezza.

- **Linea guida 3: evitare valori NULL**

Conviene evitare il più possibile i valori NULL, in quanto riempiono lo schema di informazioni inutili. I valori NULL possono rivelarsi necessari solamente nei casi eccezionali rispetto alla cardinalità di una relazione.

Dipendenza Funzionale

Una dipendenza funzionale (*functional dependency, FD*) esprime un legame semantico tra due gruppi di attributi di uno schema di relazione R . Una FD è una proprietà di R , non un suo particolare stato valido r di R . Una FD non può quindi essere dedotta da uno stato valido r , ma deve essere definita esplicitamente sulla base della semantica degli attributi di R .

Forma normale

Una forma normale è una proprietà di una base di dati relazionale che ne garantisce la qualità, ovvero l'assenza di determinati difetti (quelli di cui si parla nelle linee guida). Una relazione non normalizzata presenta:

- **Ridondanze;**
- **Comportamenti anomali** in fase di aggiornamento.

La normalizzazione è solitamente definita sui modelli relazionali, ma ha senso in altri contesti, ad esempio il modello ER.

Normalizzazione

La procedura di normalizzazione (che è un vero e proprio algoritmo) serve a trasformare schemi non normalizzati in schemi che godono della forma normale. La normalizzazione va utilizzata come tecnica di verifica di una base di dati, e non costituisce una metodologia di progettazione. Cercare di progettare uno schema di relazione attraverso la normalizzazione sarebbe infatti troppo complesso per rappresentare un'opzione viabile.

Definizione (pseudo)formale di dipendenza funzionale

Data una relazione r su $R(X)$, e due sottoinsiemi non vuoti Y e Z di x . Esiste una dipendenza funzionale in r da Y a Z se, per ogni coppia di n-uple t_1 e t_2 di r con gli stessi valori di Y , risulta che t_1 e t_2 hanno gli stessi valori anche su Z . Come avevamo già visto, la notazione è:

$$Y \rightarrow Z$$

Notiamo poi che $Y \rightarrow Z$ non implica $Z \rightarrow Y$. Una dipendenza funzionale è detta **completa** quando $Y \rightarrow Z$ e per ogni $W \subset Y$, non vale $W \rightarrow Z$. Se Y è una superchiave di $R(X)$, allora Y determina ogni altro attributo della relazione, ergo $Y \rightarrow X$. A questo punto, se Y è una chiave (ovvero superchiave minimale), si ha che $Y \rightarrow X$ dalla chiave a $R(X)$ è completa. Una dipendenza funzionale si dice **banale** se è sempre soddisfatta:

- $Y \rightarrow Y$ è banale;
- $Y \rightarrow A$ se $A \notin Y$ non è banale;
- $Y \rightarrow Z$ non è banale se nessun attributo di Z appartiene a Y (è una condizione più generale della precedente).

Legami fra dipendenze funzionali e anomalie

Le dipendenze funzionali possono essere usate per verificare l'eventuale presenza di anomalie in un progetto. Tornano utili anche nella normalizzazione di schemi. Data la loro importanza, indicheremo con $R(X, F)$ uno schema di relazione $R(X)$ che rispetta un'insieme di dipendenze funzionali F .

Implicazione

Sia F un insieme di dipendenze funzionali definite su $R(Z)$, e sia $X \rightarrow Y$. Si dice che F implica (logicamente) $X \rightarrow Y$ se, per ogni possibile istanza r di R che verifica tutte le dipendenze funzionali di F , risulta verificata anche la dipendenza funzionale $X \rightarrow Y$. In simboli, $F \models X \rightarrow Y$. La definizione di implicazione non è direttamente utilizzabile nella pratica. Essa prevede una quantificazione universale sulle istanze della base di dati, e non disponiamo di un'algoritmo per calcolare tutte le dipendenze funzionali implicate da un'insieme F .

Regole di inferenza di Armstrong

Armstrong (1974) fornisce delle regole di inferenza che permettono di derivare costruttivamente tutte le dipendenze funzionali che sono implicate da un insieme di dipendenze iniziale. Esse sono:

- **Riflessività:**

$$Y \subseteq X \models X \rightarrow Y$$

- **Additività** (o arricchimento):

$$X \rightarrow Y \models XZ \rightarrow YZ \forall Z$$

- **Transitività:**

$$X \rightarrow Y \wedge Y \rightarrow Z \models X \rightarrow Z$$

Derivazione

Le regole di inferenza di Armstrong permettono di definire derivazioni. Dato un insieme di regole di inferenza RI , un insieme di dipendenze funzionali F , e una dipendenza funzionale f , una derivazione di f da F è una sequenza finita f_1, \dots, f_n tale che:

- $f_n = f$
- ogni f_i è un elemento di F o è ottenuta dalle precedenti dipendenze f_1, \dots, f_{n-1} attraverso una regola di inferenza RI .

Indichiamo con $F \vdash X \rightarrow Y$ il fatto che la dipendenza funzionale $X \rightarrow Y$ sia derivabile da F usando RI . Le regole di derivazione comunemente usate sono, a partire dalle regole di inferenza di Armstrong:

- **Unione:**

$$\{X \rightarrow Y, X \rightarrow Z\} \vdash X \rightarrow YZ$$

Dimostrazione:

1. $X \rightarrow Z$ da ipotesi;
2. $X(X) \rightarrow ZX$ da additività di (1);
3. $X \rightarrow Y$ da ipotesi;
4. $ZX \rightarrow YZ$ da additività di (3);
5. $X \rightarrow ZX, ZX \rightarrow YZ \Rightarrow X \rightarrow YZ$ dalla transitività di (2) e (4).

- **Decomposizione:**

$$\{X \rightarrow YZ\} \vdash X \rightarrow Y$$

Dimostrazione:

1. $X \rightarrow YZ$ da ipotesi;
2. $YZ \rightarrow Y$ da riflessività;
3. $X \rightarrow YZ, YZ \rightarrow Y \Rightarrow X \rightarrow Y$ dalla transitività di (1) e (2).

- **Indebolimento:**

$$\{X \rightarrow Y\} \vdash XZ \rightarrow Y$$

1. $XZ \rightarrow X$ per riflessività;
2. $X \rightarrow Y$ da ipotesi;
3. $XZ \rightarrow X, X \rightarrow Y \Rightarrow XZ \rightarrow Y$ dalla transitività di (1) e (2).

- **Identità:**

$$\{\} \vdash X \rightarrow X$$

Chiusura di un insieme di attributi

Dato uno schema $R(T, F)$ con $X \subseteq T$, la chiusura di X rispetto a F , indicata col simbolo R_F^+ , è definita come:

$$X_F^* = \{A \subset T | F \vdash X \rightarrow A\}$$

Se non vi sono ambiguità si può semplicemente scrivere X^+ .

Teorema di chiusura degli attributi

Se è possibile, usando le regole di inferenza, scrivere, partendo da F , che $X \rightarrow Y$, allora Y è contenuto nella chiusura di X e viceversa. In simboli:

$$F \vdash X \rightarrow Y \Leftrightarrow Y \subseteq X^+$$

Correttezza e completezza

Dato un qualche insieme di regole di inferenza RI e un insieme di dipendenze funzionali F , RI è corretto se:

$$F \vdash X \rightarrow Y \Rightarrow F \models X \rightarrow Y$$

Ovvero applicando RI ad un insieme F di dipendenze funzionali, si ottengono solo dipendenze logicamente implilcate da F . Inoltre RI è completo se:

$$F \models X \rightarrow Y \Rightarrow F \vdash X \rightarrow Y$$

Ovvero applicando RI a un insieme F di dipendenze funzionali si ottengono tutte le dipendenze logicamente implicate da F . A questo punto si può enunciare:

Teorema fondamentale della correttezza e completezza

Le regole di inferenza di Armstrong sono corrette e complete. Questo teorema ci permette di scambiare i simboli \models e \vdash . In particolare questo si applica alla definizione di chiusura di attributi, cioè:

$$X_F^+ = \{A \subset T | F \models X \rightarrow A\}$$

Si può dimostrare che le regole di inferenza di Armstrong sono minimali, ovvero che non se ne può ignorare anche soltanto una senza privarle della completezza. Esistono però altri insieme di regole di inferenza completi che non coincidono con le regole di inferenza di Armstrong.

Chiusura di un'insieme di dipendenze funzionali

Sia F un insieme di dipendenze funzionali definite su $R(Z)$. Allora la chiusura di F è l'insieme F^+ di tutte e sole le dipendenze funzionali implicate da F :

$$F^+ = \{X \rightarrow Y | F \models X \rightarrow Y\}$$

Dato un'insieme di dipendenze funzionali F definite su $R(Z)$, un'istanza r di R che soddisfa F soddisfa anche le dipendenze funzionali di F^+ .

Algoritmo per il calcolo di F^+

Riportiamo adesso un'algoritmo per il calcolo della chiusura di un insieme di dipendenze funzionali F^+ a partire da F , usando le regole di inferenza di Armstrong:

- Input: $R(T, F)$;
- Output: $F^+ \text{ chiusura}$.

```
1 metti F in F+
2 while (F+ non cambia) do
3     foreach f in F+ do
4         #applica riflessivita' e addittivita' a f e aggiungi le
        dipendenze ottenute a F+
5     foreach f_1, f_2 in F+ do
6         #se possibile, applica la transitivita' a f_1 e f_2 e
        aggiungi a F+ la dipendenza ottenuta
7 return F+
```