Riconoscere e formalizzare le dipendenze funzionali

Giorgio Ghelli

25 ottobre 2007

1 Riconoscere e formalizzare le dipendenze funzionali

Non sempre è facile indiduare le dipendenze funzionali espresse in linguaggio naturale, perché esistono diversi modi di esprimere la stessa dipendenza, ed esistono espressioni simili tra loro che esprimono dipendenze diverse. In questa nota esploriamo questa varietà di espressioni in modo sistematico, in modo da facilitarne il riconoscimento e la traduzione nella dipendenza corrispondente.

Affrontiamo per prima cosa le dipendenze in cui il determinante è un solo attributo, per poi passare al caso più generale delle dipendenze con più attributi nel determinante, ed al caso particolare dei vincoli di superchiave.

1.1 Dipendenze funzionali $\mathsf{A} \rightarrow \mathsf{B}$

Una dipendenza $\mathsf{X} \rightarrow \mathsf{Y}$ su di una relazione R rappresenta il seguente vincolo:

$\forall r$ istanza valida di R, $\forall t_1, t_2 \in r.\ t_1[X] = t_2[X] \Rightarrow t_1[Y] = t_2[Y]$

Quindi, tale dipendenza rappresenta un'implicazione di *uguaglianze*, che noi indicheremo sinteticamente come:

$\mathsf{X}_= \Rightarrow \mathsf{Y}_=$

Ad esempio, si consideri una relazione che rappresenta l'allocazione settimanale delle aule di una struttura didattica a dei corsi, con schema:

Allocazioni(Aula, Giorno, Ora, Corso, Docente)

Il vincolo:

Corso $\rightarrow$ Docente

significa che ogni volta che due allocazioni riguardano lo stesso corso, esse riguardano anche lo stesso docente, ovvero che per ogni corso c'è un unico docente, ovvero l'implicazione $\mathsf{Corso}_{=} \Rightarrow \mathsf{Docente}_{=}$.

La stessa sentenza può essere espressa in almeno altri due modi con opportune manipolazioni logiche sfruttando le seguenti equivalenze della logica proposizionale (dove $\wedge$ lega più di $\Rightarrow$):

$\mathsf{A} \Rightarrow \mathsf{B}$	(A implica B: implicazione di uguaglianze o implicazione diretta)	$\Leftrightarrow$
$\mathsf{A} \wedge \neg\mathsf{B} \Rightarrow \mathsf{False}$	(A e non B è impossibile: impossibilità o per assurdo)	$\Leftrightarrow$
$\neg\mathsf{B} \Rightarrow \neg\mathsf{A}$	(non B implica non A: implicazione di disuguaglianze o implicazione contrapposta)	

In altre parole, una proposizione si può spostare da un lato all'altro dell'implicazione, purché ne venga cambiato il segno. Per essere precisi, in una formula

$$\mathsf{A}_1 \wedge \ldots \wedge \mathsf{A}_{n-1} \wedge \mathsf{A}_n \wedge \mathsf{True} \Rightarrow \mathsf{B}_1 \vee \ldots \vee \mathsf{B}_m \vee \mathsf{False}$$

è possibile spostare a destra, negandolo, ciascun fattore A_i e spostare a sinistra, negandolo, ciascun addendo B_j.

Se indichiamo con $\mathsf{A}_{\neq}$ la negazione di $\mathsf{A}_{=}$, abbiamo quindi tre modi diversi di esprimere una dipendenza funzionale $\mathsf{Corso} \rightarrow \mathsf{Docente}$:

1. $\mathsf{Corso}_{=} \Rightarrow \mathsf{Docente}_{=}$ (implicazione diretta): "se in due allocazioni il corso è uguale, allora in docente è uguale".

2. $\mathsf{Corso}_{=} \wedge \mathsf{Docente}_{\neq} \Rightarrow \mathsf{False}$ (impossibilità): "non possono esserci due allocazioni con lo stesso corso e docente diverso".

3. $\mathsf{Docente}_{\neq} \Rightarrow \mathsf{Corso}_{\neq}$ (implicazione contrapposta): "se in due allocazioni il docente è diverso, allora in corso è diverso".

Sulla base di questa osservazione, proponiamo il seguente metodo per riconoscere le dipendenze funzionali di questo tipo:

1. riconoscere se l'affermazione da formalizzare è un'implicazione diretta, un'espressione di impossibilità, o un'implicazione contrapposta, e formalizzarla usando la notazione indicata in predecenza;

2. nel caso non si tratti di un'implicazione diretta, spostare le disuguaglianze $\mathsf{Attributo}_{\neq}$ nella parte opposta dell'implicazione, in modo da arrivare ad un'implicazione diretta.

Esercizio 1 Date le seguenti sentenze, (a) stabilite in che forma di trovano, (b) rappresentatele in forma $\mathsf{A} \Rightarrow \mathsf{B}$ oppure $\mathsf{A} \wedge \mathsf{B}_{\neq} \Rightarrow \mathsf{False}$ oppure $\mathsf{B}_{\neq} \Rightarrow \mathsf{A}_{\neq}$, (c) portatele nella forma $\mathsf{A} \Rightarrow \mathsf{B}$, ed infine (d) scrivete la dipendenza funzionale corrispondente.

1. Non ci sono mai due corsi diversi con lo stesso docente.

2. In una tabella Esami(Voto, Matricola, Cognome), se la Matricola è diversa, allora anche il Cognome è diverso.

3. In una tabella Esami(Voto, Matricola, Cognome), ad ogni Matricola corrisponde un unico Cognome.

1.2 Dipendenze funzionali $\mathsf{X} \rightarrow \mathsf{B}$

Nel caso di dipendenze con più attributi nel determinante, le manipolazioni logiche si generalizzano come segue:

$A_1 \wedge \ldots \wedge A_{n-1} \wedge A_n \Rightarrow B$	(implicazione diretta)	$\Leftrightarrow$
$A_1 \wedge \ldots \wedge A_{n-1} \wedge A_n \wedge \neg B \Rightarrow \text{False}$	(impossibilità)	$\Leftrightarrow$
$A_1 \wedge \ldots \wedge A_{n-1} \wedge \neg B \Rightarrow \neg A$	(implicazione contrapposta)	

Quindi, l'espressione di una dipendenza di questo tipo in modo diretto o tramite impossibilità non cambia rispetto al caso precedente. Tuttavia, per quanto riguarda l'implicazione contrapposta, osserviamo come, tra le tante uguaglianze a sinistra della dipendenza, solo una è negata e va quindi spostata. Facciamo un esempio.

Il vincolo

Aula, Giorno, Ora $\rightarrow$ Corso

significa che se due righe coincidono su Aula, Giorno, Ora devono coincidere anche su Corso ($\text{Aula}_= \wedge \text{Giorno}_= \wedge \text{Ora}_= \Rightarrow \text{Corso}_=$). Anche questo vincolo si può esprimere nei tre stili indicati.

1. Implicazione diretta ("se in due allocazioni aula, giorno ed ora coincidono, il corso è uguale" oppure "aula, giorno ed ora determinano il corso"):

   $\text{Aula}_= \wedge \text{Giorno}_= \wedge \text{Ora}_= \Rightarrow \text{Corso}_=$

2. Impossibilità ("non possono esserci due corsi diversi contemporaneamente nella stessa aula"):

   $\text{Aula}_= \wedge \text{Giorno}_= \wedge \text{Ora}_= \wedge \text{Corso}_{\neq} \Rightarrow \text{False}$

3. Implicazione contrapposta

   (a) ("in ogni giorno ed ora, due corsi diversi stanno in due aule diverse"):
   $\text{Giorno}_= \wedge \text{Ora}_= \wedge \text{Corso}_{\neq} \Rightarrow \text{Aula}_{\neq}$

(b) ("in ogni giorno ed in ogni aula, due corsi diversi devono fare lezione in due ore diverse"):

$\mathsf{Aula}_= \wedge \mathsf{Giorno}_= \wedge \mathsf{Corso}_{\neq} \Rightarrow \mathsf{Ora}_{\neq}$

(c) ("in ogni ora ed in ogni aula, due corsi diversi devono fare lezione in due giorni diversi"):

$\mathsf{Aula}_= \wedge \mathsf{Ora}_= \wedge \mathsf{Corso}_{\neq} \Rightarrow \mathsf{Giorno}_{\neq}$

(d) ecc.

La forma (2) è probabilmente la più comune per esprimere dipendenze complesse come queste. Anche per queste dipendenze si consiglia il metodo indicato in precedenza: riconoscerne la forma, formalizzarle, spostare le disuguaglianze sul lato opposto dell'implicazione.

Esercizio 2 Date le seguenti sentenze, (a) stabilite in che forma di trovano, (b) rappresentatele in forma

$\mathsf{A}_1 \wedge \ldots \wedge \mathsf{A}_{n-1} \wedge \mathsf{A}_n \Rightarrow \mathsf{B}$ oppure
$\mathsf{A}_1 \wedge \ldots \wedge \mathsf{A}_{n-1} \wedge \mathsf{A}_n \wedge \neg\mathsf{B} \Rightarrow \mathsf{False}$ oppure
$\mathsf{A}_1 \wedge \ldots \wedge \mathsf{A}_{n-1} \wedge \neg\mathsf{B} \Rightarrow \neg\mathsf{A}$

(c) portatele nella forma $\mathsf{A}_1 \wedge \ldots \wedge \mathsf{A}_n \Rightarrow \mathsf{B}$, ed infine (d) scrivete la dipendenza funzionale corrispondente.

1. Lo stesso docente non può essere in due aule contemporaneamente.
2. In una tabella Esami(Voto, Materia, Matricola), se in due esami c'è la stessa Matricola ma il voto è diverso, allora anche la Materia è diversa.
3. In una tabella Esami(Voto, Materia, Matricola), non posso trovare due esami diversi che hanno la stessa Materia e Matricola.

1.3 Dipendenze funzionali $\mathsf{X} \rightarrow \mathsf{T}$, ovvero vincoli di superchiave

Si consideri l'affermazione: "Aula, giorno ed ora non possono coincidere tutte e tre". Si potrebbe essere tentati di formalizzarla come segue:

$\mathsf{Aula}_= \wedge \mathsf{Giorno}_= \wedge \mathsf{Ora}_= \Rightarrow \mathsf{False}$

Però questa implicazione non ha nessuna delle forme che noi vogliamo, e non ci porta a nessuna dipendenza funzionale (inoltre, questa implicazione è falsa: se consideriamo due volte la stessa riga, abbiamo due righe che soddisfano l'ipotesi ma non la tesi). La soluzione più semplice è osservare che questa affermazione esprime in realtà un vincolo di superchiave, e quindi si può immediatamente formalizzare come:

Aula, Giorno, Ora $\rightarrow$ tutti gli altri attributi

La discussione potrebbe fermarsi qui. Se però vogliamo trattare anche questo caso come gli altri, dobbiamo per prima cosa rendere esplicito il fatto che stiamo parlando di righe diverse, e dobbiamo riscrivere l'affermazione come segue:

Aula, giorno ed ora non possono coincidere tutte e tre *in due righe diverse*

Infatti, se consideriamo due volte la stessa riga, l'affermazione originaria era falsa. Questa affermazione corretta si formalizza come segue:

$\mathsf{Aula}_= \wedge \mathsf{Giorno}_= \wedge \mathsf{Ora}_= \wedge \neg(\mathsf{Aula}_= \wedge \mathsf{Giorno}_= \wedge \mathsf{Ora}_= \wedge \mathsf{Docente}_=) \Rightarrow \mathsf{False}$

Oppure (equivalentemente):

$\mathsf{Aula}_= \wedge \mathsf{Giorno}_= \wedge \mathsf{Ora}_= \wedge \neg(\mathsf{Corso}_= \wedge \mathsf{Docente}_=) \Rightarrow \mathsf{False}$

Passando dalla forma (2) alla forma (1), ovvero portando a destra la sottoformula negata, otteniamo:

$\mathsf{Aula}_= \wedge \mathsf{Giorno}_= \wedge \mathsf{Ora}_= \Rightarrow (\mathsf{Aula}_= \wedge \mathsf{Giorno}_= \wedge \mathsf{Ora}_= \wedge \mathsf{Docente}_=)$

Oppure:

$\mathsf{Aula}_= \wedge \mathsf{Giorno}_= \wedge \mathsf{Ora}_= \Rightarrow (\mathsf{Corso}_= \wedge \mathsf{Docente}_=)$

A cui corrispondono le dipendenze equivalenti:

Aula, Giorno, Ora $\rightarrow$ Aula, Giorno, Ora, Corso, Docente
Aula, Giorno, Ora $\rightarrow$ Corso, Docente

L'osservazione di fondo è che quando usiamo la forma (2) per descrivere un vincolo, in genere esprimiamo esplicitamente un attributo che deve essere diverso (il Docente o il Corso negli esempi dei Capitoli 1.1 e 1.2). Nei vincoli di superchiave basta che sia diverso un attributo qualunque, e questa condizione viene spesso espressa in modo ambiguo ("*due* righe non possono coincidere su X Y Z": la diversità su almeno un attributo è espressa dal numerale *due*) o addirittura non viene espressa per nulla ("X Y e Z non possono mai coincidere tutti e tre"). Per usare il nostro metodo, questa ipotesi di diversità deve diventare esplicita.

2 Soluzione agli esercizi

Esercizio 1 Date le seguenti sentenze, (a) stabilite in che forma di trovano, (b) rappresentatele in forma $\mathsf{A} \Rightarrow \mathsf{B}$ oppure $\mathsf{A} \wedge \mathsf{B}_{\neq} \Rightarrow \mathsf{False}$ oppure $\mathsf{B}_{\neq} \Rightarrow \mathsf{A}_{\neq}$, (c) portatele nella forma $\mathsf{A} \Rightarrow \mathsf{B}$, ed infine (d) scrivete la dipendenza funzionale corrispondente.

1. Non ci sono mai due corsi diversi con lo stesso docente.
   (a) forma (2) (impossibilità),
   (b) $\mathsf{Corso}_{\neq} \wedge \mathsf{Docente}_{=} \Rightarrow \mathsf{False}$
   (c) $\mathsf{Docente}_{=} \Rightarrow \mathsf{Corso}_{=}$
   (d) $\mathsf{Docente} \longrightarrow \mathsf{Corso}$
2. In una tabella Esami(Voto, Matricola, Cognome), se la Matricola è diversa, allora anche il Cognome è diverso.
   (a) forma (3) (implicazione contrapposta),
   (b) $\mathsf{Matricola}_{\neq} \Rightarrow \mathsf{Cognome}_{\neq}$
   (c) $\mathsf{Cognome}_{=} \Rightarrow \mathsf{Matricola}_{=}$
   (d) $\mathsf{Cognome} \longrightarrow \mathsf{Matricola}$
3. In una tabella Esami(Voto, Matricola, Cognome), ad ogni Matricola corrisponde un unico Cognome.
   (a) forma (1) (implicazione diretta),
   (b) $\mathsf{Matricola}_{=} \Rightarrow \mathsf{Cognome}_{=}$
   (c) $\mathsf{Matricola}_{=} \Rightarrow \mathsf{Cognome}_{=}$
   (d) $\mathsf{Matricola} \longrightarrow \mathsf{Cognome}$

Esercizio 2 Date le seguenti sentenze, (a) stabilite in che forma di trovano, (b) rappresentatele in forma
$\mathsf{A}_1 \wedge \ldots \wedge \mathsf{A}_{n-1} \wedge \mathsf{A}_n \Rightarrow \mathsf{B}$ oppure
$\mathsf{A}_1 \wedge \ldots \wedge \mathsf{A}_{n-1} \wedge \mathsf{A}_n \wedge \neg\mathsf{B} \Rightarrow \mathsf{False}$ oppure
$\mathsf{A}_1 \wedge \ldots \wedge \mathsf{A}_{n-1} \wedge \neg\mathsf{B} \Rightarrow \neg\mathsf{A}$
(c) portatele nella forma $\mathsf{A}_1 \wedge \ldots \wedge \mathsf{A}_n \Rightarrow \mathsf{B}$, ed infine (d) scrivete la dipendenza funzionale corrispondente.

1. Lo stesso docente non può essere in due aule contemporaneamente.
   (a) forma (2) (impossibilità),
   (b) $\mathsf{Aula}_{\neq} \wedge \mathsf{Giorno}_{=} \wedge \mathsf{Ora}_{=} \wedge \mathsf{Docente}_{=} \Rightarrow \mathsf{False}$
   (c) $\mathsf{Giorno}_{=} \wedge \mathsf{Ora}_{=} \wedge \mathsf{Docente}_{=} \Rightarrow \mathsf{Aula}_{=}$
   (d) $\mathsf{Giorno}, \mathsf{Ora}, \mathsf{Docente} \longrightarrow \mathsf{Aula}$

2. In una tabella Esami(Voto, Materia, Matricola), se in due esami c'è la stessa Matricola ma il voto è diverso, allora anche la Materia è diversa.

   (a) forma (3) (implicazione contrapposta),
   (b) $\text{Matricola}_= \wedge \text{Voto}_{\neq} \Rightarrow \text{Materia}_{\neq}$
   (c) $\text{Matricola}_= \wedge \text{Materia}_= \Rightarrow \text{Voto}_=$
   (d) $\text{Matricola}, \text{Materia} \longrightarrow \text{Voto}$

3. In una tabella Esami(Voto, Materia, Matricola), non posso trovare due esami diversi che hanno la stessa Materia e Matricola.

   (a) forma (2) (impossibilità),
   (b) $\text{Matricola}_= \wedge \text{Materia}_= \wedge \text{Voto}_{\neq} \Rightarrow \text{False}$
   (c) $\text{Matricola}_= \wedge \text{Materia}_= \Rightarrow \text{Voto}_=$
   (d) $\text{Matricola}, \text{Materia} \longrightarrow \text{Voto}$