TD 9 : Normalisation et dépendances

Dépendances fonctionnelles

2024-11-22

Avec solutions

L3 MIASHS/Ingémath
Université Paris Cité
Année 2024
Course Homepage
Moodle

Définitions

Une *dépendance fonctionnelle* est une expression de la forme

$$A_1, A_2, \ldots, A_k \rightarrow A_{k+1}, \ldots, A_n$$

où $A_1, A_2, \ldots, A_k, A_{k+1}, \ldots, A_n$ sont des attributs (colonnes) d'une base de données.

Elle signifie que deux tuples ayant la même valeur sur $A_1, \ldots, A_k$ doivent avoir la même valeur sur chaque colonnes $A_{k+1}, \ldots, A_n$ (en français : $A_1, \ldots, A_k$ *déterminent* $A_{k+1}, \ldots, A_n$. On dit que les attributs $A_{k+1}, \ldots, A_n$ *dépendent fonctionnellement* de $A_1, A_2, \ldots, A_k$.

La notion de dépendance est transitive : si $A \rightarrow B$ et $B \rightarrow C$ alors $A \rightarrow C$.

Un ensemble de dépendances fonctionnelles $\mathcal{F}$ est *irréductible* si aucune dépendance ne peut être déduite des autres en utilisant les règles suivantes :

- trivialité : si $Y \subseteq X$ alors $X \rightarrow Y$
- augmentation : si $X \rightarrow Y$ alors $X, Z \rightarrow Y, Z$ pour toute suite d'attributs Z.
- transitivité : si $X \rightarrow Y$ et $Y \rightarrow Z$ alors $X \rightarrow Z$
- union : si $X \rightarrow Y$ et $X \rightarrow Z$ alors $X \rightarrow Y, Z$
- décomposition/séparation si $X \rightarrow Y$ et $Z \subseteq Y$ alors $X \rightarrow Z$

La *clôture transitive* des attributs $A_1, \ldots, A_k$ pour un ensemble de dépendances fonctionnelles $\mathcal{F}$ est l'ensemble des attributs $B_1, \ldots, B_\ell$ qui dépendent fonctionnellement de $A_1, \ldots, A_k$.

On la note

$$[A_1, \ldots, A_k]^+_{\mathcal{F}}$$

en oubliant $\mathcal{F}$ si le contexte est clair.

Un ensemble d'attributs $A_1, \ldots, A_k$ est une *super-clé pour une relation* $R(B_1, \ldots, B_\ell)$ si ce sont des attributs de R et si sa clôture transitive contient $B_1, \ldots, B_\ell$. C'est une clé si elle est minimale, c'est-à-dire, aucun sous-ensemble strict de cette super-clé n'est une clé.

Un schéma est en :

- FN_3 si pour toute dépendance fonctionnelle non triviale, le membre de gauche contient une clef ou tout attribut du membre de droite appartient à une clef.
- FNBC si pour toute dépendance fonctionnelle non triviale, le membre de gauche contient une clef.

Un schéma et un ensemble de dépendances fonctionnelles peut se *décomposer* en une collection de schémas, dans le sens où chaque relation R peut se décomposer en $R_1, \ldots, R_k$ tels que $R_i = \pi_i(R)$ pour une certaine projection π_i.

On dit cette décomposition *sans perte d'information* si toute relation R du schéma d'origine peut être retrouvée à partir des relations $R_1, \ldots, R_k$: $R = \pi_1(R) \bowtie \ldots \bowtie \pi_k(R)$.

On dit que cette décomposition *respecte les dépendances fonctionnelles* si celles-ci sont toujours satisfaites par la nouvelle décomposition.

Exercice

Soit une relation concernant des personnes en France avec les attributs suivants~ :

`Nom, Numéro de sécurité sociale, Commune, Département, Code postal, Numéro de téléphone`

Quelles sont les dépendances fonctionnelles censées être satisfaites ?

Solution

- `Numéro de sécurité sociale → Nom, Commune, Département, Code postal`
- `Commune, Département → Code postal` (presque : Paris est une commune et un département. Paris possède plusieurs codes postaux)
- `Code postal → Département`

Exercice

Soit un schéma d'attributs $A_1, A_2, \ldots A_n$ et un ensemble de dépendances fonctionnelles. Calculer le nombre de super-clefs (en fonction de n) dans les cas suivants~ :

- La seule clef est $\{A_1\}$.
- Les seules clefs sont $\{A_1\}$ et $\{A_2\}$.
- Les seules clefs sont $\{A_1, A_2\}$ et $\{A_3, A_4\}$.
- Les seules clefs sont $\{A_1, A_2\}$ et $\{A_1, A_3\}$.

Solution

- 2^{n-1}
- $3 \times 2^{n-2}$
- $2^{n-4} \times 3 \times 2 + 2^{n-4}$
- $3 \times 2^{n-3}$

Exercice

Soit le schéma $\mathcal{A} = \{A, B, C, D\}$ et l'ensemble de dépendances fonctionnelles

$$\Sigma = \{A \longrightarrow B, B \longrightarrow C\}$$

- Quelle est la fermeture $\{A\}^+$ de $\{A\}$?

Solution

Initialisation : $X = \{A\}$
Etape 1 : Il existe une DF dont la partie gauche est incluse dans \$ X\$: $A \longrightarrow B$. On rajoute les attributs en partie droite. D'où $X = \{A, B\}$
Etape 2 : Il existe une DF dont la partie gauche est incluse dans \$ X\$: $B \longrightarrow C$. On rajoute les attributs en partie droite. D'où $X = A, B, C$.
C'est fini, plus de DF à utiliser. Conclusion $\{A\}^+ = A, B, C$

- Quelles sont les super-clés ? Les clés ?

Solution

Une clef doit contenir $\{A, D\}$ puisque ces deux attributs ne sont à droite d'aucune DF de Σ.
De plus $\{A, D\}^+ = \{A, B, C, D\}$.
La seule clef est donc $\{A, D\}$.

Exercice

Soit le schéma $\mathcal{A} = \{A, B, C, D, E, F\}$ et l'ensemble de dépendances fonctionnelles

$$\Sigma = \Big\{\{A, B\} \to C, \{B, C\} \to \{A, D\}, D \to E, \{C, F\} \to B\Big\}$$

- Calculer la fermeture $\{A, B\}^+$ de $\{A, B\}$.
- Est-ce que Σ implique la dépendance fonctionnelle $\{A, B\} \to D$?
- Est-ce que Σ implique la dépendance fonctionnelle $D \to A$?

Solution

- On obtient $\{A, B\}^+ = \{A, B, C, D, E\}$.
- Oui car $D \in \{A, B\}^+$
- Non car $\{D\}^+ = \{D, E\}$ ne contient pas A.

Exercice

Montrer que les assertions suivantes sont fausses :

- $A \to B$ implique $B \to A$.
- Si $\{A, B\} \to C$ et $A \to C$ alors $B \to C$.
- Si $\{A, B\} \to C$ alors $A \to C$ ou $B \to C$.

Solution

- La relation

A	B
1	2
4	2

satisfait $A \to B$ mais pas $B \to A$.
- La relation

A	B	C
1	2	3
4	2	4

satisfait $\{A, B\} \to C$ et $A \to C$ mais pas $B \to C$.
- La relation

A	B	C
1	2	3
4	2	4
1	3	1

satisfait $\{A, B\} \to C$ mais ni $A \to C$ ni $B \to C$.

Exercice : déductions de DF

Soit le schéma $\mathcal{A} = \{A, B, C, D, E, F, G, H\}$ et soit

$$\Sigma = \{AB \longrightarrow C;\ \ B \longrightarrow D;\ \ CD \longrightarrow E;\ \ CE \longrightarrow GH;\ \ G \longrightarrow A\}$$

Est-ce que les dépendances

- $A, B \longrightarrow E$
- $B, G \longrightarrow C$
- $A, B \longrightarrow G$

sont déductibles de Σ ?

Solution

oui... Méthode à suivre : pour la première et la troisième, on calcule la fermeture de $\{A, B\}$.
On a $\{A, B\}^+ = \{A, B, C, D, E, G, H\}$.
Pour la seconde, on a $\{B, G\}^+ = \{A, B, C, D, E, G, H\}$.

Σ est elle irréductible/minimale ?

Rappel

Pour que Σ soit minimale, il y a 3 conditions à remplir :

- Σ est sous forme canonique, un seul attribut à droite.
- Aucune DF redondante , i.e. aucune DF ne peut être déduite des autres.
- Aucune DF redondante à gauch, .i.e. les déterminants sont minimaux

Solution

Σ n'est pas minimale/irréductible : $C, E \longrightarrow H$ est redondante à gauche.
On la remplace par $C \longrightarrow H$.
De même $A \longrightarrow H$ est redondante.
Une version minimale est :

$$\mathcal{F} = \{A \longrightarrow B; C \longrightarrow H; C \longrightarrow E; A \longrightarrow C\}$$

Exercice : équivalence d'ensembles de DFs

- Soit
$$\Sigma_1 = \{A \longrightarrow B; C, E \longrightarrow H; C \longrightarrow E; A \longrightarrow C, H\}$$
et
$$\Sigma_2 = \{A \longrightarrow B, C; C \longrightarrow E, H\}$$
Les deux ensembles de dépendances fonctionnelles Σ_1 et Σ_2 sont-ils *équivalents* ?

Solution

Montrons que Σ_1 implique Σ_2 ($\Sigma_1 \vDash \Sigma_2$).
$A \to B$ et $C \to E$ sont dans Σ_1. $A \to C$ est impliqué par $A \to CH \in \Sigma_1$.
Donc $A \to BC$ se déduit de Σ_1. De plus Σ_1 implique $C \to H$ (puisque Σ_1 contient $C \to E$ et $CE \to H$).
Donc Σ_1 implique $C \to EH$. On a montré que toutes les DF de Σ_2 sont impliquées par Σ_1.
Montrons que Σ_2 implique Σ_1.

- Σ_2 contient $A \to B$.
- Σ_2 contient $C \to EH$ qui implique $C \to H$ qui implique $CE \to H$.
- Σ_2 contient $C \to EH$ qui implique $C \to E$.
- Σ_2 implique $A \to CH$ ...

Exercice : Décomposition et perte d'information

- On considère le schéma de relation $\mathcal{A} = A, B, C$ et la dépendance fonctionnelle suivante :

$$\Sigma = \{A, B \longrightarrow C\}.$$

Déterminer si la décomposition suivante est sans perte d'information

$$\mathcal{A}_1 = \{A, B\}, \quad \mathcal{A}_2 = \{B, C\}$$

en étudiant le cas de la table suivante :

A	B	C
1	2	3
4	2	5

Solution

Cette relation satisfait $AB \rightarrow C$. De plus, la jointure naturelle des deux projections contient les deux nouveaux tuples (1,2,5) et (4,2,3). Donc il y a perte d'information.

Exercice : poursuite

- On considère le schéma de relation $\mathcal{A} = \{A, B, C, D, E\}$ et les dépendances fonctionnelles suivantes :

$$\Sigma = \{A \longrightarrow C; B \longrightarrow C; C \longrightarrow D; D, E \longrightarrow C; C, E \longrightarrow A\}.$$

Appliquer l'algorithme de poursuite pour déterminer si la décomposition suivante est sans perte d'information :

$$\mathcal{A}_1 = \{A, D\}, \mathcal{A}_2 = \{A, B\}, \mathcal{A}_3 = \{B, E\}, \mathcal{A}_4 = \{C, D, E\}, \mathcal{A}_5 = \{A, E\}$$

Même question pour la décomposition :

$$\mathcal{A}_1 = \{A, D\}, \mathcal{A}_2 = \{A, B\}, \mathcal{A}_3 = \{B, E\}, \mathcal{A}_4 = \{C, D\}, \mathcal{A}_5 = \{D, E\}, \mathcal{A}_6 = \{A, E\}$$

Solution

La première décomposition est SPI. On doit montrer que :
$R = \pi_{A,D}(R) \bowtie \pi_{A,B}(R) \bowtie \pi_{B,E}(R) \bowtie \pi_{C,D,E}(R) \bowtie \pi_{A,E}(R)$.
On voit que :

$$R \subseteq \pi_{A,D}(R) \bowtie \pi_{A,B}(R) \bowtie \pi_{B,E}(R) \bowtie \pi_{C,D,E}(R) \bowtie \pi_{A,E}(R)$$

et il faut montrer l'autre inclusion.
On considère un tuple $t = (a, b, c, d, e)$ de la jointure naturelle.
Pour $1 \leq i \leq 5$, comme $\pi_{\mathcal{A}_i}(t) \in \pi_{\mathcal{A}_i}(R)$ il existe un tuple $t_i \in R$ tel que $\pi_{\mathcal{A}_i}(t) = \pi_{\mathcal{A}_i}(t_i)$, ce que l'on représente par le tableau

A	B	C	D	E
a	b_1	c_1	d	e_1
a	b	c_2	d_2	e_2
a_3	b	c_3	d_3	e
a_4	b_4	c	d	e
a	b_5	c_5	d_5	e

Par la dépendance $A \longrightarrow C$, on sait que deux tuples ayant la même valeur sur A, ont la même sur C. On remplace dans la table les valeurs indicées par la valeur c quand c'est possible, ou on unifie simplement les valeurs indicées sinon (ici on prendra c_1).

A	B	C	D	E
a	b_1	c_1	d	e_1
a	b	c_1	d_2	e_2
a_3	b	c_3	d_3	e
a_4	b_4	c	d	e
a	b_5	c_1	d_5	e

Solution (suite)

On traite maintenant $B \longrightarrow C$.

A	B	C	D	E
a	b_1	c_1	d	e_1
a	b	c_1	d_2	e_2
a_3	b	c_1	d_3	e
a_4	b_4	c	d	e
a	b_5	c_1	d_5	e

Pour la dépendance $C \longrightarrow D$, on obtient :

A	B	C	D	E
a	b_1	c_1	d	e_1
a	b	c_1	d	e_2
a_3	b	c_1	d	e
a_4	b_4	c	d	e
a	b_5	c_1	d	e

Solution (suite)

Pour $D, E \longrightarrow C$, on a cette fois (notez qu'on remplace tous les c_1 du coup comme l'un d'entre eux devait l'être) :

A	B	C	D	E
a	b_1	c	d	e_1
a	b	c	d	e_2
a_3	b	c	d	e
a_4	b_4	c	d	e
a	b_5	c	d	e

Enfin, on termine avec $C, E \longrightarrow A$:

A	B	C	D	E
a	b_1	c	d	e_1
a	b	c	d	e_2
a	b	c	d	e
a	b_4	c	d	e
a	b_5	c	d	e

On voit que le tuple (a, b, c, d, e) apparait.
En d'autres termes :
$R \supset \pi_{A,D}(R) \bowtie \pi_{A,B}(R) \bowtie \pi_{B,E}(R) \bowtie \pi_{C,D,E}(R) \bowtie \pi_{A,E}(R)$. La décomposition est donc SPI.

Solution (suite)

La seconde décomposition n'est pas SPI : l'algorithme de poursuite échoue.
Le tableau de départ s'écrit :

A	B	C	D	E
a	b_1	c_1	d	e_1
a	b	c_2	d_2	e_2
a_3	b	c_3	d_3	e
a_4	b_4	c	d	e_4
a_5	b_5	c_5	d	e
a	b_6	c_6	d_6	e

- $A \rightarrow C$ donc $c_1 = c_2 = c_6$

A	B	C	D	E
a	b_1	c_1	d	e_1
a	b	c_1	d_2	e_2
a_3	b	c_3	d_3	e
a_4	b_4	c	d	e_4
a_5	b_5	c_5	d	e
a	b_6	c_1	d_6	e

Solution (suite)

$B \to C$ donc $c_3 = c_1$

A	B	C	D	E
a	b_1	c_1	d	e_1
a	b	c_1	d_2	e_2
a_3	b	c_1	d_3	e
a_4	b_4	c	d	e_4
a_5	b_5	c_5	d	e
a	b_6	c_1	d_6	e

$C \to D$ donc $d_1 = d_3 = d_6 = d$

A	B	C	D	E
a	b_1	c_1	d	e_1
a	b	c_1	d	e_2
a_3	b	c_1	d	e
a_4	b_4	c	d	e_4
a_5	b_5	c_5	d	e
a	b_6	c_1	d	e

Solution (suite)

$DE \to C$ donc $c_5 = c1$

A	B	C	D	E
a	b_1	c_1	d	e_1
a	b	c_1	d	e_2
a_3	b	c_1	d	e
a_4	b_4	c	d	e_4
a_5	b_5	c_1	d	e
a	b_6	c_1	d	e

$CE \to A$ donc $a_3 = a_5 = a$

A	B	C	D	E
a	b_1	c_1	d	e_1
a	b	c_1	d	e_2
a	b	c_1	d	e
a_4	b_4	c	d	e_4
a	b_5	c_1	d	e
a	b_6	c_1	d	e

Toutes les DF de Σ sont satisfaites et aucune ligne n'est égale à (a, b, c, d, e).
Donc la décomposition n'est pas SPI.
Si on suppose que pour tout $1 \leq i \leq 6$, $a_i \neq a$, $b_i \neq b$, $c_i \neq c$, $d_i \neq d$, $e_i \neq e$, ce dernier tableau fournit un exemple de relation strictement plus petite que la jointure naturelle des projections.
En effet Σ est satisfait et (a, b, c, d, e) est clairement dans la jointure naturelles des projections. Or (a, b, c, d, e) n'est pas dans la relation.

Exercice

Soit $\mathcal{A} = \{A, B, C, D, E\}$ un schéma et soit la décomposition $\{\mathcal{A}_1, \mathcal{A}_2, \mathcal{A}_3\}$ où

$$\mathcal{A}_1 = A, B, C \quad \mathcal{A}_2 = B, C, D \quad \mathcal{A}_3 = A, C, E$$

Pour chaque ensemble Σ de dépendances fonctionnelles ci-dessous, appliquer l'algorithme de poursuite pour déterminer si la décomposition est sans perte d'information. Dans le cas où il y a perte d'information, donner une relation R de schéma $\mathcal{A}$ satisfaisant Σ et telle que

$$\pi_{\mathcal{A}_1}(R) \bowtie \pi_{\mathcal{A}_2}(R) \bowtie \pi_{\mathcal{A}_3}(R) \not\subseteq R$$

- $\Sigma = \{B \to E, CE \to A\}$
- $\Sigma = \{AC \to E, BC \to D\}$
- $\Sigma = \{A \to D, D \to E, B \to D\}$
- $\Sigma = \{A \to D, CD \to E, E \to D\}$

Solution

-

	A	B	C	D	E
{A,B,C}	a	b	c	d_1	e_1
{B,C,D}	a_2	b	c	d	e_2
{A,C,E}	a	b_3	c	d_3	e

$B \to E$ donc $e_1 = e_2$. Ensuite $CE \to A$ donc $a_2 = a$. On obtient

	A	B	C	D	E
{A,B,C}	a	b	c	d_1	e_1
{B,C,D}	a	b	c	d	e_1
{A,C,E}	a	b_3	c	d_3	e

Toutes les DF de Σ sont satisfaites. Donc il y a perte d'information. Ce dernier tableau est une relation R qui satisfait Σ et telle que

$$\pi_{\mathcal{A}_1}(R) \bowtie \pi_{\mathcal{A}_2}(R) \bowtie \pi_{\mathcal{A}_3}(R) \not\subseteq R$$

puisque $(a, b, c, d, e) \notin R$.

Solution (suite)

	A	B	C	D	E
{A,B,C}	a	b	c	d_1	e_1
{B,C,D}	a_2	b	c	d	e_2
{A,C,E}	a	b_3	c	d_3	e

$AC \to E$ donc $e_1 = e$. Ensuite $BC \to D$ donc $d_1 = d$. On obtient

	A	B	C	D	E
{A,B,C}	a	b	c	d	e
{B,C,D}	a_2	b	c	d	e_2
{A,C,E}	a	b_3	c	d_3	e

Le premier tuple est (a, b, c, d, e). Donc la décomposition est SPI.

Solution (suite)

-

	A	B	C	D	E
{A,B,C}	a	b	c	d_1	e_1
{B,C,D}	a_2	b	c	d	e_2
{A,C,E}	a	b_3	c	d_3	e

$A \to D$ donc $d_3 = d_1$.

	A	B	C	D	E
{A,B,C}	a	b	c	d_1	e_1
{B,C,D}	a_2	b	c	d	e_2
{A,C,E}	a	b_3	c	d_1	e

$D \to E$ donc $e_1 = e$

	A	B	C	D	E
{A,B,C}	a	b	c	d_1	e
{B,C,D}	a_2	b	c	d	e_2
{A,C,E}	a	b_3	c	d_1	e

$B \to D$ donc $d_1 = d$

	A	B	C	D	E
{A,B,C}	a	b	c	d	e
{B,C,D}	a_2	b	c	d	e_2
{A,C,E}	a	b_3	c	d_1	e

La décomposition est donc SPI.

Solution (suite)

•

	A	B	C	D	E
{A,B,C}	a	b	c	d_1	e_1
{B,C,D}	a_2	b	c	d	e_2
{A,C,E}	a	b_3	c	d_3	e

$A \to D$ donc $d_3 = d_1$.

	A	B	C	D	E
{A,B,C}	a	b	c	d_1	e_1
{B,C,D}	a_2	b	c	d	e_2
{A,C,E}	a	b_3	c	d_1	e

$CD \to E$ donc $e_1 = e$.

	A	B	C	D	E
{A,B,C}	a	b	c	d_1	e
{B,C,D}	a_2	b	c	d	e_2
{A,C,E}	a	b_3	c	d_1	e

$E \to D$ est satisfaite ainsi que les deux premières DF. Donc la décomposition n'est pas SPI.

Exercice : Normalisation

On considère le schéma de relation `R(C,T,H,S,E,N)` :

`R(Cours, Enseignant, Horaire, Salle, Étudiant, Note)`

et les dépendances fonctionnelles suivantes :

$$\mathcal{F} = \{\mathtt{C} \to \mathtt{T};\quad \mathtt{H,S} \to \mathtt{C};\quad \mathtt{H,T} \to \mathtt{S};\quad \mathtt{C,E} \to \mathtt{N};\quad \mathtt{H,E} \to S\}.$$

- Calculer une clé.

Solution

`H,E` n'étant jamais à droite, ils font obligatoirement partis d'une clé. Or `HE+ = ALL`

- Mettre en Boyce-Codd Normal Form (BCNF), donner plusieurs résultats possibles.

Solution

1er version :
`C -> T` donne `T1(CT)` et `T2(CHSEN)`.
`CE -> N` donne `T1(CT),T2(C,E,N) T3(CHSE)`
`HE -> S` donne `(HES ; HEC)`
2eme version :
`CE -> N` donne `T1(CENT), T2(CEHS) ...`