TD 9 : Normalisation et dépendances

Dépendances fonctionnelles

2024-11-22

Avec solutions

- L3 MIASHS/Ingémath
- Université Paris Cité
- Année 2024-2025
- Course Homepage
- Moodle



Définitions

Une dépendance fonctionnelle est une expression de la forme

$$A_1, A_2, \dots, A_k \to A_{k+1}, \dots, A_n$$

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,\dots,A_k doivent avoir la même valeur sur chaque colonnes A_{k+1},\dots,A_n (en français : A_1,\dots,A_k déterminent A_{k+1},\dots,A_n . On dit que les attributs A_{k+1}, \dots, A_n dépendent fonctionnellement de A_1, A_2, \dots, A_k .

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

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

La clôture transitive des attributs A_1, \dots, A_k pour un ensemble de dépendances fonctionnelles $\mathcal F$ est l'ensemble des attributs B_1, \dots, B_ℓ qui dépendent fonctionnellement de A_1, \dots, A_k .

On la note

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

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

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

Un schéma est en :

- FN₃ si pour toute dépendance fonctionnelle non triviale, le membre de gauche contient une clef ou tout attribut du membre de droit 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, Numéro de téléphone
- Commune, Département → Code postal
- Code postal → Département

Exercice

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

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

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

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

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

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

$$\begin{array}{c|c}
\hline
A & B \\
\hline
1 & 2 \\
4 & 2
\end{array}$$

satisfait $A \to B$ mais pas $B \to A$.

• La relation

$$\begin{array}{c|cccc}
\hline
A & B & C \\
\hline
1 & 2 & 3 \\
4 & 2 & 4
\end{array}$$

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

• La relation

A	В	С
1	2	3
4	2	4
1	3	1

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

Exercice

• 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 $\Sigma \sim ?$

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 $\{B, G\}^+ = \{A, B, C, D, E, G, H\}.$

Solution

Pour que \mathcal{F} soit minimale, il y a 3 conditions à remplir :

- F est sous forme canonique, un seul attribut à droite.
- Aucune DF redondante, i.e. ne pouvant être déduite des autres.
- Aucune DF redondante à gauche.

n'est pas minimale : $C, E \longrightarrow H$ est redondante à gauche.

On la remplace par $C \longrightarrow H$.

De même $A \longrightarrow H$ est redondante.

Une version minimale est :

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

• 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 $(\Sigma_1 \models \Sigma_2)$.

 $A \to B$ 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 :

$$\begin{array}{c|cccc} \hline A & B & C \\ \hline 1 & 2 & 3 \\ 4 & 2 & 5 \\ \hline \end{array}$$

•

Solution

Cette relation satisfait $AB \to C$. 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.

• 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	В	С	D	Е
a	b_1	c_1	d	$\overline{e_1}$
a	b	c_2	d_2	e_2
a_3	b	c_3	d_3	e
a_4	b_4	$^{\mathrm{c}}$	d	e
\mathbf{a}	b_5	c_5	d_5	e

Par la dépendance $A \to C$, 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	В	С	D	Е
a	b_1	c_1	d	$\overline{e_1}$
a	b	c_1	d_2	e_2
a_3	b	c_3	d_3	e
a_4	b_4	$^{\mathrm{c}}$	d	e
a	b_5	c_1	d_5	e

Solution (suite)

On traite maintenant $B \longrightarrow C$.

A	В	С	D	Ε
a	b_1	c_1	d	e_1
\mathbf{a}	b	c_1	d_2	e_2
a_3	b	c_1	d_3	e
a_4	b_4	\mathbf{c}	d	e
\mathbf{a}	b_5	c_1	d_5	e

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

A	В	С	D	Е
a	b_1	c_1	d	$\overline{e_1}$
\mathbf{a}	b	c_1	d	e_2
a_3	b	c_1	d	e
a_4	b_4	\mathbf{c}	d	e
\mathbf{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	В	С	D	Е
a	b_1	c	d	e_1
\mathbf{a}	b	\mathbf{c}	d	e_2
a_3	b	$^{\mathrm{c}}$	d	e
a_4	b_4	\mathbf{c}	d	e
a	b_5	\mathbf{c}	d	e

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

Ā	В	С	D	Е
a	b_1	c	d	e_1
a	b	\mathbf{c}	d	e_2
a	b	\mathbf{c}	d	e
a	b_4	\mathbf{c}	d	e
a	b_5	$^{\mathrm{c}}$	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	В	С	D	Е
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	\mathbf{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	В	С	D	Е
a	b_1	c_1	d	$\overline{e_1}$
a	b	c_1	d_2	e_2
a_3	b	c_3	d_3	e
a_4	b_4	$^{\mathrm{c}}$	$^{\mathrm{d}}$	e_4
a_5	b_5	c_5	$^{\mathrm{d}}$	e
a	b_6	c_1	d_6	e

Solution (suite)

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

A	В	С	D	\mathbf{E}
a	b_1	c_1	d	$\overline{e_1}$
a	b	c_1	d_2	e_2
a_3	b	c_1	d_3	e
a_4	b_4	$^{\mathrm{c}}$	d	e_4
a_5	b_5	c_5	d	\mathbf{e}
a	b_6	c_1	d_6	e

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

A	В	С	D	Е
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	\mathbf{c}	d	e_4
a_5	b_5	c_5	d	e
\mathbf{a}	b_6	c_1	d	e

Solution (suite)

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

_				
A	В	С	D	Е
a	b_1	c_1	d	e_1
\mathbf{a}	b	c_1	d	e_2
a_3	b	c_1	d	e
a_4	b_4	\mathbf{c}	d	e_4
a_5	b_5	c_1	d	e
a	b_6	c_1	d	e

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

A	В	С	D	Е
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	\mathbf{c}	d	e_4
\mathbf{a}	b_5	c_1	d	e
\mathbf{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 $1 \le i \le 6$, $a_i \ne a$, $b_i \ne b$, $c_i \ne c$, $d_i \ne d$, $e_i \ne e$, 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

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

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

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

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

Solution

	A	В	С	D
$\overline{\{A,B,C\}}$		b	c	$\overline{d_1}$
{B,C,D}	a_2	b	c	d
$\{A,C,E\}$	a	b_3	$^{\mathrm{c}}$	d_3

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

	A	В	С	D
$\overline{\{A,B,C\}}$	a	b	c	$\overline{d_1}$
$\{B,C,D\}$	\mathbf{a}	b	$^{\mathrm{c}}$	d
$\{A,C,E\}$	a	b_3	\mathbf{c}	d_3

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

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

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

Solution (suite)

	A	В	С	D
$\overline{\{A,B,C\}}$	a	b	\mathbf{c}	$\overline{d_1}$
$\{B,C,D\}$	a_2	b	$^{\mathrm{c}}$	d
$\{A,C,E\}$	a	b_3	\mathbf{c}	d_3

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

	A	В	С	D
$\overline{\{A,B,C\}}$	a	b	\mathbf{c}	d
$\{B,C,D\}$	a_2	b	$^{\mathrm{c}}$	d
$\{A,C,E\}$	\mathbf{a}	b_3	\mathbf{c}	d_3

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

• Solution (suite)

•

	A	В	\mathbf{C}	D
$ \overline{\{A,B,C\}} $ $ \{B,C,D\} $	a a_2	b b	c c	$\frac{\overline{d_1}}{\mathrm{d}}$
$\{A,C,E\}$	a_2	b_3	c	d_3

 $A \to D \text{ donc } d_3 = d_1.$

	A	В	С	D
$\overline{\{A,B,C\}}$	a	b	c	$\overline{d_1}$
$\{B,C,D\}$	a_2	b	$^{\mathrm{c}}$	$^{\mathrm{d}}$
$\{A,C,E\}$	a	b_3	$^{\mathrm{c}}$	d_1

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

	A	В	С	D
$\overline{\{A,B,C\}}$	a	b	c	$\overline{d_1}$
$\{B,C,D\}$	a_2	b	$^{\mathrm{c}}$	d
$\{A,C,E\}$	a	b_3	$^{\mathrm{c}}$	d_1

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

	A	В	С	D
$\overline{\{A,B,C\}}$	a	b	c	d
$\{B,C,D\}$	a_2	b	\mathbf{c}	d
$\{A,C,E\}$	a	b_3	$^{\mathrm{c}}$	d_1

La décomposition est donc SPI.

Solution (suite)

•

	A	В	С	D
$\overline{\{A,B,C\}}$	a	b	c	$\overline{d_1}$
$\{B,C,D\}$	a_2	b	\mathbf{c}	d
$\{A,C,E\}$	\mathbf{a}	b_3	\mathbf{c}	d_3

 $A \to D \text{ donc } d_3 = d_1.$

	A	В	С	D
$\overline{\{A,B,C\}}$	a	b	c	$\overline{d_1}$
$\{B,C,D\}$	a_2	b	\mathbf{c}	d
$\{A,C,E\}$	\mathbf{a}	b_3	\mathbf{c}	d_1

 $CD \to E \text{ donc } e_1 = e.$

	A	В	\mathbf{C}	D
$\overline{\{A,B,C\}}$	a	b	c	$\overline{d_1}$
$\{B,C,D\}$	a_2	b	$^{\mathrm{c}}$	d
$\{A,C,E\}$	\mathbf{a}	b_3	\mathbf{c}	d_1

 $E \to D$ 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:

 $C \rightarrow T \text{ donne T1(CT) et T2(CHSEN)}.$

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) ...