Fondements des bases de données

Inférence de dépendances fonctionnelles : Système d'Armstrong

Équipe pédagogique BD

https://perso.liris.cnrs.fr/marc.plantevit/doku/doku.php?id=lifbdw2_2018a

Version du 19 septembre 2018

Inférence de dépendances

Notion d'inférence

- Concept fondamental pour la théorie des bases de données.
- C'est la déduction des contraintes qui sont sémantiquement impliquées par d'autres.
- On va donner un algorithme (un calcul symbolique) qui permet de décider *effectivement* de cette implication logique.
- L'algorithme est un processus qui manipule des symboles, il s'appuie uniquement sur la *syntaxe* des dépendances. On vérifiera que cet algorithme est :
  - correct : toutes les déductions syntaxiques sont bien sémantiquement valides ;
  - complet : toutes les déductions sémantiquement valides peuvent bien être syntaxiquement déduites.

Implication logique des dépendances fonctionnelles

Définition

Soit F un ensemble de DF sur un schéma de relation R et f une DF sur R. On surcharge $\models$ pour un ensemble de DFs :

$$r \models F \text{ ssi } \forall f \in F . r \models f$$

F implique logiquement (sémantiquement) f, noté

$$F \models f \text{ ssi } \forall r . r \models F \Rightarrow r \models f$$

Exemple

Soit le schéma $Etudiant(Num, Nom, Ville, Region, CP)$ et l'ensemble de dépendances $F = \{f_1 = Num \rightarrow Nom, Ville, Region$
$f_2 = Ville, Region \rightarrow CP\}$. On peut déduire les DFs suivantes :

- $f_1' = Num \rightarrow Nom, Ville$
- $f_2' = Num \rightarrow CP$

Implication logique des dépendances fonctionnelles

En utilisant la sémantique des dépendances fonctionnelles, prouver qu'effectivement $F \models f_1'$ et $F \models f_2'$

- Difficile de prouver que chaque intuition est correcte.
- Comment s'assurer que l'énumération $\{f \mid F \models f\}$ est exhaustive ?
- On va donner un algorithme qui le fait.

Règles et système d'inférences

- Une règle d'inférence est une expression de la forme

$$\frac{f_0 \qquad \dots \qquad f_n}{f}$$

- Un système d'inférence est un ensemble de règles d'inférence.
- Une preuve de f à partir de F notée $F \vdash f$ est une *séquence* $(f_0, \dots, f_n)$ de DFs telle que $f_n = f$ et $\forall i \in [0..n]$, :
  - soit $f_i \in F$;
  - soit f_i est la conséquence d'une règle dont toutes les prémisses $f_0 \dots f_p$ apparaissent avant f_i dans la séquence.

Liens avec le cours de logique classique

Ce sont bien les même notions que celles des systèmes de déduction $\mathcal{G}$ et $\mathcal{LK}$. Pour les preuves formelles, on pourra utiliser le même formalise des arbres de preuves.

Système d'Armstrong

Système d'inférence d'Armstrong

Soit F un ensemble de DF sur un schéma de relation R. Les règles d'inférence suivantes sont appelées *système d'Armstrong*

- Réflexivité

$$\frac{Y \subseteq X}{X \rightarrow Y}$$

- Augmentation

$$\frac{X \rightarrow Y}{WX \rightarrow WY}$$

- Transitivité

$$\frac{X \rightarrow Y \qquad Y \rightarrow Z}{X \rightarrow Z}$$

Propriétés du système d'Armstrong

Correction et complétude

- Il est correct : $F \vdash f \Rightarrow F \models f$
- Il est complet : $F \models f \Rightarrow F \vdash f$

$$F \models \alpha \Leftrightarrow F \vdash \alpha$$

Exemple

Considérons l'ensemble $\Sigma = \{A \to B, B \to C, CD \to E\}$ de DFs sur $\{A, B, C, D, E\}$. Montrons que $\Sigma \vdash AD \to E$

$$\dfrac{\dfrac{\dfrac{A \to B \qquad B \to C}{A \to C}}{AD \to CD} \qquad CD \to E}{AD \to E}$$

On a montré que $\Sigma \models AD \to E$.

Propriétés du système d'Armstrong

Preuve de la correction

Le principle est de montrer qu'une instance qui satisfait l'hypothèse de la règle satisfait aussi sa conclusion (voir TD3).

Exemple : la transitivité

Soit r une base de donnée sur R telle que $r \models X \rightarrow Y$ et $r \models Y \rightarrow Z$. Supposons deux tuples $t_1, t_2 \in r$ tels que $t_1[X] = t_2[X]$, il faut montrer que $t_1[Z] = t_2[Z]$. C'est immédiat, car en utilisant le fait que $r \models X \rightarrow Y$ on déduit que $t_1[Y] = t_2[Y]$, puis en utilisant $r \models Y \rightarrow Z$ on déduit que $t_1[Z] = t_2[Z]$.

Preuve de la complétude

Dans le cours suivant. Voir son application aux bases d'Armstrong.

Autres règles

- Décomposition

$$\frac{X \to YZ}{X \to Y}$$

- Composition

$$\frac{X \to Y \qquad X \to Z}{X \to YZ}$$

- Pseudo-transitivité

$$\frac{X \to Y \qquad WY \to Z}{WX \to Z}$$

Prouver que ces règles sont valides et qu'on peut les ajouter au système d'Armstrong

Exercice

Soit l'ensemble F de DF suivant sur le schéma $R = ABCDEFG$:

- $A \rightarrow B$
- $A \rightarrow C$
- $A \rightarrow D$
- $CD \rightarrow E$
- $BE \rightarrow F$
- $ABE \rightarrow G$
- $EG \rightarrow ABD$
- $FG \rightarrow AE$

- Démontrer que DF $F \models A \rightarrow F$.
- Démontrer que DF $F \models A \rightarrow G$.

Fin du cours.