Inférence, fermetures et couvertures de contraintes

Marc Plantevit

marc.plantevit@liris.cnrs.fr

Contraintes et Inférence

Classe de contraintes d'intégrité

- Correspond à un type particulier de contraintes d'intégrité sur un schéma de relation ou de base de données.

Ex. : Ensemble de toutes les dépendances fonctionnelles sur un schéma de relation R

Contraintes et Inférence

Classe de contraintes d'intégrité

- Correspond à un type particulier de contraintes d'intégrité sur un schéma de relation ou de base de données.

Ex. : Ensemble de toutes les dépendances fonctionnelles sur un schéma de relation R

Inférence de contraintes

- Concept fondamental pour la théorie des bases de données.
- Consiste à considérer les contraintes étant impliquées par d'autres contraintes.
- Les contraintes impliquées sont tout aussi valides que les contraintes de départ, mais souvent de façon implicite : elles peuvent échapper à la connaissance du concepteur

Plan

Implication de DF (Définition)

Implication de DF

Soit F un ensemble de DF sur un schéma de relation R, et f une DF sur R. On dit que F implique f, noté $F \models f$ si pour toute relation r sur R telle que $r \models F$, on a $r \models f$.

Implication de DF (Définition)

Implication de DF

Soit F un ensemble de DF sur un schéma de relation R, et f une DF sur R. On dit que F implique f, noté $F \models f$ si pour toute relation r sur R telle que $r \models F$, on a $r \models f$.

Exemple

Soit le schéma $ETUDIANT(NUMETUD, NOMETUD, NOMVILLE, REGION, CP)$ et l'ensemble de DF sur ce schéma :

$F = \{NUMETUD \rightarrow NOMETUD, NOMVILLE, REGION, CP; NOMVILLE, REGION \rightarrow CP\}$

On comprend intuitivement que les deux DF suivantes seront également valides, par implication :

$F \models \{NUMETUD \rightarrow NOMVILLE, REGION; NUMETUD \rightarrow CP\}$ TODO : montrer le!!

Difficile de prouver que chaque intuition est correcte et d'être complet.

$\Rightarrow$ Traitement automatique nécéssaire!!

Règles d'inférences

Règle et système d'inférence

- Une **règle d'inférence** est une expression de la forme $F \Rightarrow f$.

Règles d'inférences

Règle et système d'inférence

- Une **règle d'inférence** est une expression de la forme $F \Rightarrow f$.
- Un **système d'inférence** est un ensemble de règles d'inférence.

Règles d'inférences

Règle et système d'inférence

- Une **règle d'inférence** est une expression de la forme $F \Rightarrow f$.
- Un **système d'inférence** est un ensemble de règles d'inférence.
- Si S est un système d'inférence, on note $F \vdash_S f$ le fait qu'on puisse dériver f à partir de F par des applications successives de règles de S.

Système d'Armstrong

Le système d'inférence des DF est le *système d'Armstrong*.

Definition (Système d'inférence d'Armstrong)

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

- **Réflexivité** : si $Y \subseteq X \subseteq R$ alors $F \vdash X \rightarrow Y$.

TODO : montrer que ces axiomes sont justes (TD3).

Système d'Armstrong

Le système d'inférence des DF est le *système d'Armstrong*.

Definition (Système d'inférence d'Armstrong)

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

- **Réflexivité** : si $Y \subseteq X \subseteq R$ alors $F \vdash X \rightarrow Y$.
- **Augmentation** : si $F \vdash X \rightarrow Y$ et $F \vdash W \subseteq R$ alors $F \vdash WX \rightarrow WY$.

TODO : montrer que ces axiomes sont justes (TD3).

Système d'Armstrong

Le système d'inférence des DF est le *système d'Armstrong*.

Definition (Système d'inférence d'Armstrong)

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

- **Réflexivité** : si $Y \subseteq X \subseteq R$ alors $F \vdash X \rightarrow Y$.
- **Augmentation** : si $F \vdash X \rightarrow Y$ et $F \vdash W \subseteq R$ alors $F \vdash WX \rightarrow WY$.
- **Transitivité** : si $F \vdash X \rightarrow Y$ et $F \vdash Y \rightarrow Z$ alors $F \vdash X \rightarrow Z$

TODO : montrer que ces axiomes sont justes (TD3).

Propriétés du Système d'Inférence d'Armstrong

- Il est juste, soit $F \vdash f \Longrightarrow F \models f$;
- Il est complet, soit $F \models f \Longrightarrow F \vdash f$;
- Il est minimal en nombre de règles d'inférence.

Propriétés du Système d'Inférence d'Armstrong

- Il est juste, soit $F \vdash f \Longrightarrow F \models f$;
- Il est complet, soit $F \models f \Longrightarrow F \vdash f$;
- Il est minimal en nombre de règles d'inférence.

Conséquence : $F \models \alpha \Leftrightarrow F \vdash \alpha$ **; nous utiliserons dans la suite la notation** $F \models \alpha$**.**

Exemple

Considérons l'ensemble $\Sigma = \{A \rightarrow C, B \rightarrow C, CD \rightarrow E\}$ de df sur $\{A, B, C, D, E\}$. Montrons que $\Sigma \models AD \rightarrow E$:

σ_1 :	$A \rightarrow C$	$\in \Sigma$;
σ_2 :	$AD \rightarrow CD$	à partir de **augmentation** et σ_1 ;
σ_3 :	$CD \rightarrow E$	$\in \Sigma$;
σ_4 :	$AD \rightarrow E$	à partir de σ_2 et σ_3 en utilisant la **transitivité**.

D'autres règles utiles mais non nécessaires.

Decomposition : Si $X \rightarrow YZ$ alors $X \rightarrow Y$ et $X \rightarrow Z$.

D'autres règles utiles mais non nécessaires.

Decomposition : Si $X \rightarrow YZ$ alors $X \rightarrow Y$ et $X \rightarrow Z$.

Union (composition) : .Si $X \rightarrow Y$ et $X \rightarrow Z$ alors $X \rightarrow YZ$.

D'autres règles utiles mais non nécessaires.

Decomposition : Si $X \rightarrow YZ$ alors $X \rightarrow Y$ et $X \rightarrow Z$.

Union (composition) : .Si $X \rightarrow Y$ et $X \rightarrow Z$ alors $X \rightarrow YZ$.

Pseudo-transitivité : Si $X \rightarrow Y$ et $WY \rightarrow Z$ alors $WX \rightarrow Z$.

TODO

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

$$
\begin{array}{lll}
A \rightarrow B & A \rightarrow C & A \rightarrow D \\
CD \rightarrow E & BE \rightarrow F & \\
ABE \rightarrow G & EG \rightarrow ABD & FG \rightarrow AE
\end{array}
$$

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

Problème d'inférence de DF.

- Le système d'Armstrong permet de systématiser la recherche de toutes les DF impliquées par d'autres DF.
- **Problème d'inférence des DF** : Soit F un ensemble de DF, et f une DF, a-t-on $F \models f$?
- La résolution de ce problème est "facile" pour les DF,
- linéaire en la taille de F et de f (c'est à dire l'espace nécessaire pour les écrire).

Problème d'inférence de DF.

- Le système d'Armstrong permet de systématiser la recherche de toutes les DF impliquées par d'autres DF.
- **Problème d'inférence des DF** : Soit F un ensemble de DF, et f une DF, a-t-on $F \models f$?
- La résolution de ce problème est "facile" pour les DF,
- linéaire en la taille de F et de f (c'est à dire l'espace nécessaire pour les écrire).
- Le concept fondamental utilisé dans cet algorithme est celui de *fermeture d'un ensemble d'attributs*.

Problème d'inférence de DF.

- Le système d'Armstrong permet de systématiser la recherche de toutes les DF impliquées par d'autres DF.
- **Problème d'inférence des DF** : Soit F un ensemble de DF, et f une DF, a-t-on $F \models f$?
- La résolution de ce problème est "facile" pour les DF,
- linéaire en la taille de F et de f (c'est à dire l'espace nécessaire pour les écrire).
- Le concept fondamental utilisé dans cet algorithme est celui de *fermeture d'un ensemble d'attributs*.

Fermeture

Soit F un ensemble de DF, on note F^+ (fermeture de F) l'ensemble de toutes les contraintes f telles que $F \models f$.

Fermeture d'un ensemble d'attributs

- X est un ensemble d'attribut et F un ensemble de DF,
- on appelle **fermeture de** X par rapport à F l'ensemble de tous les attributs qu'on peut "déduire" de X par des dépendances fonctionnelles.
- Il faut aussi tenir compte des DF qui ne sont pas dans F, mais qui sont dérivables à partir de F. Donc :

$$X^+ = \{A \mid F \models X \to A\}$$

ou encore

$$X^+ = \{A \mid X \to A \in F^+\}$$

Définition alternative de la notion de clé d'une relation

X est clé d'un schéma R si et seulement si $X^+ = R$

Définition alternative de la notion de clé d'une relation

X est clé d'un schéma R si et seulement si $X^+ = R$

Lemme

Soient F un ensemble de DF et $X \rightarrow Y$ une DF. Alors on a $F \models X \rightarrow Y$ ssi on a $Y \subseteq X^+$.

Définition alternative de la notion de clé d'une relation

X est clé d'un schéma R si et seulement si $X^+ = R$

Lemme

Soient F un ensemble de DF et $X \rightarrow Y$ une DF. Alors on a $F \models X \rightarrow Y$ ssi on a $Y \subseteq X^+$.

Ainsi, pour tester si on a $F \models X \rightarrow Y$, on calcule X^+

Algorithme : Fermeture d'un ensemble d'attributs

Data: F un ensemble de DF, et X un ensemble d'attributs.
Result: X^+, la fermeture de X par F.

$unused := F$
$closure := X$
repeat
    if $W \to Z \in unused$ *and* $W \subseteq closure$ **then**
        $unused := unused - \{W \to Z\}$
        $closure := closure \cup Z$
until jusqu'à ce qu'il n'y ait plus de changement ;
Print($closure$)

- Algorithme permet de vérifier si un ensemble de df implique une dépendance.
- Pour tester, l'implication d'un ensemble de dépendances, il suffit de tester indépendamment, l'implication de chaque dépendance de l'ensemble.
- TODO : l'algorithme précédent tourne en temps $O(n^2)$, où n est la longueur de F et de X.
- Amélioration pour obtenir un temps linéaire :
  - Pour toute df, $X \to Y$ de F non utilisée, il faut stocker le nombre d'attributs de X non encore dans *closure*.
  - Pour le faire efficacement, il faut maintenir à jour une liste pour chaque attribut A des df de F non utilisées pour lesquelles A apparait en partie gauche.

TODO

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

$$
\begin{array}{lll}
A \rightarrow B & A \rightarrow C & A \rightarrow D \\
CD \rightarrow E & BE \rightarrow F & \\
ABE \rightarrow G & EG \rightarrow ABD & FG \rightarrow AE
\end{array}
$$

- Démontrer que DF $F \models A \rightarrow F$ (avec fermeture).
- Démontrer que DF $F \models A \rightarrow G$ (avec fermeture).
- Démontrez que BEF n'est pas une clé de R
- Démontrez que la DF $CDE \rightarrow A$ n'est pas impliquée par F

Plan

Le système d'inférence pour les DI est le suivant :

Definition

Système d'inférence de Casanova et al. Soit I un ensemble de DI sur un schéma de base de données $\mathbf{R}$. Les règles d'inférence de DI suivantes sont appelées *système d'inférence de Casanova et al.* :

Le système d'inférence pour les DI est le suivant :

Definition

Système d'inférence de Casanova et al. Soit I un ensemble de DI sur un schéma de base de données **R**. Les règles d'inférence de DI suivantes sont appelées *système d'inférence de Casanova et al.* :

- **Reflexivité** : $I \vdash R[X] \subseteq R[X]$.

Le système d'inférence pour les DI est le suivant :

Definition

Système d'inférence de Casanova et al. Soit I un ensemble de DI sur un schéma de base de données **R**. Les règles d'inférence de DI suivantes sont appelées *système d'inférence de Casanova et al.* :

- **Reflexivité** : $I \vdash R[X] \subseteq R[X]$.
- **Projection et permutation** :
  si $I \vdash R[A_1 ... A_n] \subseteq S[B_1 ... B_n]$ alors
  $I \vdash R[A_{\sigma 1} ... A_{\sigma m}] \subseteq S[B_{\sigma 1} ... B_{\sigma m}]$ pour chaque séquence $\sigma 1 ... \sigma m$ d'entier distincts dans $\{1 ... n\}$

Le système d'inférence pour les DI est le suivant :

Definition

Système d'inférence de Casanova et al. Soit I un ensemble de DI sur un schéma de base de données **R**. Les règles d'inférence de DI suivantes sont appelées *système d'inférence de Casanova et al.* :

- **Reflexivité** : $I \vdash R[X] \subseteq R[X]$.
- **Projection et permutation** :
  si $I \vdash R[A_1 ... A_n] \subseteq S[B_1 ... B_n]$ alors
  $I \vdash R[A_{\sigma 1} ... A_{\sigma m}] \subseteq S[B_{\sigma 1} ... B_{\sigma m}]$ pour chaque séquence $\sigma 1 ... \sigma m$ d'entier distincts dans $\{1 ... n\}$
- **transitivité** : si $I \vdash R[X] \subseteq S[Y]$ et $I \vdash S[Y] \subseteq T[Z]$ alors
  $I \vdash R[X] \subseteq T[Z]$

Propriétés

- Ce système d'inférence est lui aussi juste, complet et minimal pour les dépendances d'inclusion.

Propriétés

- Ce système d'inférence est lui aussi juste, complet et minimal pour les dépendances d'inclusion.
- Contrairement aux DF, le problème d'inférence des DI est PSPACE-complet, donc infaisable dans le cas général.

Propriétés

- Ce système d'inférence est lui aussi juste, complet et minimal pour les dépendances d'inclusion.
- Contrairement aux DF, le problème d'inférence des DI est PSPACE-complet, donc infaisable dans le cas général.
- La notion I^+ s'applique aussi pour noter la fermeture d'un ensemble de DI.

Propriétés

- Ce système d'inférence est lui aussi juste, complet et minimal pour les dépendances d'inclusion.
- Contrairement aux DF, le problème d'inférence des DI est PSPACE-complet, donc infaisable dans le cas général.
- La notion I^+ s'applique aussi pour noter la fermeture d'un ensemble de DI.
- En revanche, la notion de fermeture d'un ensemble d'attributs par rapport à un ensemble de DI ne s'applique pas, car les DI manipulent des séquences et non des ensembles.

Plan

La notion de couverture est une relation d'équivalence entre des ensembles de contraintes.

Definition

Couverture d'un ensemble de DF Soit Σ et Γ deux ensembles de contraintes. Γ est une couverture de Σ si $\Gamma^+ = \Sigma^+$.

- Une couverture d'un ensemble de DF est donc une **représentation alternative**, avec d'autres DF, mais véhiculant une **sémantique rigoureusement équivalente**.
- C'est exactement le même ensemble de DF qui est implicite.
- Pareil pour les DI.

Trois propriétés sont importantes pour les couvertures des DF - peu d'études existent pour les DI.

- un ensemble F de DF est dit *non redondant* s'il n'existe pas de couverture G de F telle que $G \subset F$.
- un ensemble F de DF est dit *minimum* s'il n'existe pas de couverture G de F tel que $|G| \leq |F|$.
- F est dit *optimum* s'il n'existe pas de couverture G de F avec moins d'attributs que dans F.

Remarque : une couverture minimum est non redondante, une couverture optimum est minimum.

Couverture minimum pour un ensemble F de DF.

Data: F un ensemble de DF
Result: G une couverture minimum de F

$G := \emptyset$ **for** $X \rightarrow Y \in F$ **do**
  $G := G \cup \{X \rightarrow X^+\}$
for $X \rightarrow X^+ \in G$ **do**
  if $G - \{X \rightarrow X^+\} \vdash X^+$ **then**
    $G := G - \{X \rightarrow X^+\}$
return G

- Cet algorithme est polynomial dans le nombre de DF dans F et le nombre d'attributs dans F.
- La couverture minimum calculée par l'algorithme n'est pas forcément unique : d'autres couvertures peuvent avoir le même nombre de DF, mais être différentes.
- Parmi celles-ci, certaines sont optimum ; malheureusement, leur calcul est un problème "NP-Complet" dans le cas général.

Exemple

$$F = \left\{ \begin{array}{lll} AB \rightarrow C & C \rightarrow A & BC \rightarrow D \\ ACD \rightarrow B & D \rightarrow EF & ABE \rightarrow C \\ CF \rightarrow BD & CE \rightarrow AF & \end{array} \right\}$$

1. A partir des df $X \rightarrow Y$ de F, construire G l'ensemble des règles de la forme $X \rightarrow X^+$.
   - $AB \rightarrow AB^+ = ABCDEF$
   - ...

Exemple

$$F = \left\{ \begin{array}{lll} AB \rightarrow C & C \rightarrow A & BC \rightarrow D \\ ACD \rightarrow B & D \rightarrow EF & ABE \rightarrow C \\ CF \rightarrow BD & CE \rightarrow AF & \end{array} \right\}$$

1. A partir des df $X \rightarrow Y$ de F, construire G l'ensemble des règles de la forme $X \rightarrow X^+$.
   - $AB \rightarrow AB^+ = ABCDEF$
   - ...
2. Supprimer de G les règles inutiles ;

Exemple

$$F = \left\{ \begin{array}{lll} AB \rightarrow C & C \rightarrow A & BC \rightarrow D \\ ACD \rightarrow B & D \rightarrow EF & ABE \rightarrow C \\ CF \rightarrow BD & CE \rightarrow AF & \end{array} \right\}$$

1. A partir des df $X \rightarrow Y$ de F, construire G l'ensemble des règles de la forme $X \rightarrow X^+$.
   - $AB \rightarrow AB^+ = ABCDEF$
   - ...
2. Supprimer de G les règles inutiles ;
3. Retourner G.

TODO

Calculer les couvertures de :

F_1

$$AB \to D \quad C \to A \quad BC \to D$$
$$D \to EF \quad BE \to C \quad CF \to B$$
$$CE \to A \quad CE \to G$$

F_2

$$AB \to C \quad C \to A \quad BC \to D$$
$$D \to EF \quad BE \to C \quad CF \to B$$
$$CE \to F$$

Plan

DF et DI

- Jusqu'à présent, soit on inférait des DF à partir d'un ensemble de DF, soit on inférait des DI à partir d'un ensemble de DI.
- Il existe des interactions entre DF et DI.

Référence

Inclusion Dependencies and Their Interaction with Functional Dependencies
MARCO A. CASANOVA, RONALD FAGINS, et CHRISTOS H. PAPADIMITRIOU
JOURNAL OF COMPUTER AND SYSTEM SCIENCES 28, 29-59 (1984)

Proposition 1

Supposons que $|X| = |T|$. Alors,

$$\{R[XY] \subseteq S[TU], S : T \to U\} \models R : X \to Y$$

Démonstration.

TODO $\square$

Proposition 2

Supposons que $|X| = |T|$. Alors,

$$\{R[XY] \subseteq S[TU], R[XZ] \subseteq S[TV], S : T \rightarrow U\} \models R[XYZ] \subseteq S[TUV]$$

Démonstration.

TODO $\square$

Proposition 3 (cas spécial de prop 2)

Supposons que $|X| = |T|$. Soit, $\Sigma = \{R[XY] \subseteq S[TU], R[XZ] \subseteq S[TV], S : T \rightarrow U\}$, si r, s sont des instances satisfaisant Σ et si $u \in r$ alors $u[Y] = u[Z]$.

Démonstration.

Triviale à partir de preuve précédente. $\square$

Plan

Utilité et Intérêt

Représentation "par l'exemple" d'ensemble de contraintes :

- on manipule des valeurs, visualisation plus simple des éventuels conflits, incohérences, mauvaise conception ...

Deux principaux domaines d'application :

1. Conception par l'exemple ;
2. Échantillonnage, compréhension des bases de données existantes.

Idée

Représenter sur un exemple un ensemble de contraintes et que celui la, i.e. toutes les autres contraintes sont fausses.

Construction

- Calculer les fermés de F : $Cl(F) = \{X^+, X \subseteq R\}$
- Construire la relation d'Amstrong r correspondante :

Etape 0 Pour le fermé particulier R, construire le tuple $t_0 = (0, \ldots, 0)$;

Etape i Pour chaque élément $X \in Cl(F) \setminus R$

  - Ajouter un tuple t_i à r tq $ti[A] = 0$ pour tout $A \in X$ et $t_i[B] = i$ pourtout $B \in (schema(R)?X)$.

Exemple

$R = [ABC]$ et $F = \{A \rightarrow BC; B \rightarrow C;\}$

- $Cl(F) = \{ABC, BC, C\}$
- Relation r :

Fermés	A	B	C
$ABC(R)$	0	0	0
BC	1	0	0
C	2	2	0

TODO

F

$AB \longrightarrow C$ $C \longrightarrow A$ $BC \longrightarrow D$
$ACD \longrightarrow B$ $D \longrightarrow E$ $ABE \longrightarrow C$
$C \longrightarrow BD$ $CE \longrightarrow A$

1. Calculer l'ensemble des fermés de F ;
2. Construire la base à partir des fermés ;

Fin

Partie "Contraintes et dépendances du modèle relationnel"

Ce qu'il faut retenir :

- Ce qu'est une contrainte ;
- Définition DF et DI (syntaxe, sémantique) ;
- Implication de dépendances ;
- Différence $\models$ et $\vdash$;
- Fermeture ;
- Couverture ;