問合せ最適化 1 — 問合せ結果の包含関係

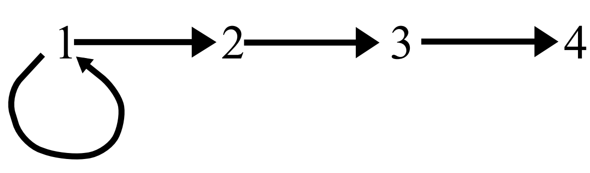
論理積問合せ

$$Q: \qquad H :- G_1 \& \ldots \& G_n$$

- ルールを使って問合せを表現.
- 本体の G_i の述語は EDB 述語.
- D を各 G_i に対する EDB 関係の集合.
  D をデータベースと呼ぶ.
- $T_Q(D) = \{H \mid H :- G_1 \& \ldots \& G_n$ は基礎代入例. $G_i \in D\}$
- $T_Q(D)$ はデータベース D に対する問合せ Q の答え.
- 論理積問合せは２０年近く研究されている「オーソドックス」な問題.
- 近年, 仮想ビュー管理やデータベース統合の観点から再び脚光を浴びている.

例

- Q:　q(X,Z) :- e(X,Y) & e(Y,Z)
- D = { e(1,1), e(1,2), e(2,3), e(3,4) }
- T_{Q}(D) = {q(1,2), q(1,3), q(2,4) }



論理積問合せの包含関係

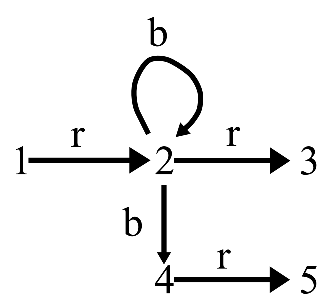
- 論理積問合せ Q_1, Q_2 に関して

$$Q_1 \subseteq Q_2$$

とは, 任意のデータベース D に対して $T_{Q_1}(D) \subseteq T_{Q_2}(D)$.

例

- Q_1 : 　p(X,Y) :- r(X,W) & b(W,W) & r(W,Y)
- Q_2 : 　p(X,Y) :- r(X,W) & b(W,Z) & r(Z,Y)
- D = {r(1,2), r(2,3), r(4,5), b(2,2), b(2,4)}
- T_{Q_1}(D) = {p(1,3)}
- T_{Q_2}(D) = {p(1,3), p(1,5)}



$Q_1 \subseteq Q_2$ の証明

- D を任意のデータベース, p(x,y) を $T_{Q_1}(D)$ の任意の元.
- p(x,y) を推論するために使った D の元を r(x,w), b(w,w), w(w,y) とする. 変数を小文字にして定数のように扱う.
- 問合せ Q_2 において代入

$$\begin{array}{cccc} \mathrm{x} & \mathrm{w} & \mathrm{w} & \mathrm{y} \\ \downarrow & \downarrow & \downarrow & \downarrow \\ \mathrm{X} & \mathrm{W} & \mathrm{Z} & \mathrm{Y} \end{array}$$

を行うと p(x,y) を推論でき, p(x,y) は $T_{Q_2}(D)$ の元.

論理積問合せの間の包含写像

$Q_1 \subseteq Q_2$ を示すためには, 以下の条件を満たす Q_2 の変数から Q_1 の変数への上への写像 (包含写像と呼ぶ) を構成すれば十分.

- Q_2 の頭部が Q_1 の頭部と一致.
- Q_2 の本体の各原子式が Q_1 の本体のある原子式と一致.

以上の条件は Q_1 での推論を Q_2 で模倣できることを保証.

例

- Q_1 : p(X,Y) :- r(X,W) & b(W,W) & r(W,Y)
- Q_2 : p(X,Y) :- r(X,W) & b(W,Z) & r(Z,Y)
- $Q_1 \subseteq Q_2$. Q_2 から Q_1 への包含写像を構成.

$$\begin{array}{cccc} \mathrm{X} & \mathrm{W} & \mathrm{Z} & \mathrm{Y} \\ \downarrow & \downarrow & \downarrow & \downarrow \\ \mathrm{X} & \mathrm{W} & \mathrm{W} & \mathrm{Y} \end{array}$$

例

- Q_1 : p(X) :- a(X,Y) & a(Y,X)
- Q_2 : p(X) :- a(X,Y) & a(Y,Z)
- $Q_1 \subseteq Q_2$: Q_2 から Q_1 への包含写像の構成.

$$\begin{array}{ccc} \mathrm{X} & \mathrm{Y} & \mathrm{Z} \\ \downarrow & \downarrow & \downarrow \\ \mathrm{X} & \mathrm{Y} & \mathrm{X} \end{array}$$

例

- $Q_1:\quad$ p(X) :- a(X,Y) & a(Y,X)
- $Q_2:\quad$ p(X) :- a(X,Y) & a(Y,Z) & a(Z,W)
- $Q_1 \subseteq Q_2$: Q_2 から Q_1 への包含写像の構成.

$$\begin{array}{cccc} \mathrm{X} & \mathrm{Y} & \mathrm{Z} & \mathrm{W} \\ \downarrow & \downarrow & \downarrow & \downarrow \\ \mathrm{X} & \mathrm{Y} & \mathrm{X} & \mathrm{Y} \end{array}$$

- Q_1 から Q_2 への包含写像は存在しない. Q_1 の X を, Q_2 の X と Z の 2 つの異なる変数へ写像する状況が生まれるため.

包含写像定理

$$Q_1 \subseteq Q_2 \Leftrightarrow Q_2 \text{ から } Q_1 \text{ への包含写像が存在.}$$

($\Leftarrow$)

- Q_1: H_1 :- G_1 & $\ldots G_n$
  Q_2: H_2 :- F_1 & $\ldots F_m$

- ν を Q_2 から Q_1 への包含写像.
  各 F_i に対して, ある G_j が存在し, $\nu(F_i) = G_j$.
  また $\nu(H_2) = H_1$.

- t を $T_{Q_1}(D)$ の任意の元.
  t を推論する際に使った定数の変数への代入を σ.
  $t = \sigma(H_1)$:- $\sigma(G_1)$ & $\ldots \sigma(G_n)$
  $\sigma(G_i) \in D \quad i = 1, \ldots, n$

- $\sigma \circ \nu(H_2)$:- $\sigma \circ \nu(F_1)$ & $\ldots$ & $\sigma \circ \nu(F_m)$
  各 F_i に対して, ある G_j が存在し, $\sigma \circ \nu(F_i) = \sigma(G_j)$.
  しかも $\sigma(G_j)$ は D の元.

- $t = \sigma(H_1) = \sigma \circ \nu(H_2)$ は $T_{Q_2}(D)$ の元.

包含写像定理　証明 (⇒)

- Q_1 の本体の原子式の変数を凍結 (frozen) し定数として扱い, 同時に原子式も EDB の元として扱う. このような EDB の集まりを正規データベース (canonical database) と呼ぶ.

  例
  Q_1　p(X) :- a(X,Y) & a(Y,X)
  正規データベース = { a(x,y), a(y,x) }

- D を Q_1 の本体を凍結してつくった正規データベースとし, Q_1 の頭部を t とする.

  例
  t = p(x) :- a(x,y) & a(y,x)

- $Q_1 \subseteq Q_2$ より $t \in T_{Q_2}(D)$.

  例
  Q_2　p(X) :- a(X,Y) & a(Y,Z) & a(Z,W)
  代入　p(x) :- a(x,y) & a(y,x) & a(x,y)

- Q_2 を使って t を推論する際, Q_2 の本体の原子式の変数には D の定数 (Q_1 の変数を凍結した定数) が代入される.

- この代入は Q_2 から Q_1 への包含写像を与える.

  例

$$\begin{array}{cccc} X & Y & Z & W \\ \downarrow & \downarrow & \downarrow & \downarrow \\ X & Y & X & Y \end{array}$$

正規データベースを使った包含関係 $Q_1 \subseteq Q_2$ の判定

- Q_1 の本体から変数を凍結し正規データベース D を生成. 凍結した際の Q_1 の頭部を t とする.
- $T_{Q_2}(D)$ を計算.
- $Q_1 \subseteq Q_2 \quad \Leftrightarrow \quad t \in T_{Q_2}(D)$
  直観的には Q_2 を使って Q_1 のすべての推論を模倣できることを調べている.
- 証明: $t \in T_{Q_2}(D)$ ならば Q_2 から Q_1 への包含写像をつくれる.
- $Q_1 \subseteq Q_2$ か否かの判定問題は NP 完全.
  Ashok K. Chandra, Philip M. Merlin: Optimal Implementation of Conjunctive Queries in Relational Data Bases. STOC 1977: 77-90.
- 現実には, 本体中の原子式の数も少なく, 変数の数も少ないので必ずしも困難な問題ではない.
- 問合せ本体が否定記号をふくむ場合, 再帰的呼出しがある場合, 算術演算をふくむ場合での包含関係の判定は重要であるが, 判定方法はより複雑になる.

例

- $Q_1 :$ p(X,Y) :- q(X,Z) & r(Z,Y)
  $Q_2 :$ p(X,Y) :- q(Y,W) & r(W,X)
- Q_1 の変数を凍結. p(x,y) :- q(x,z) & r(z,y)
- 正規データベース D = {q(x,z), r(z,y)} を作成.
- $\mathrm{p(x,y)} \notin T_{Q_2}(\mathrm{D})$ = {p(y,x)}
- $Q_1 \not\subseteq Q_2$

包含関係を使った問合せ最適化

- Q_1 :　p(X,Y) :- e(Y,X) & e(X,Z)
- Q_2 :　p(X,Y) :- e(Y,X) & e(X,Z) & e(U,X)
- Q_1 から Q_2 へは自明な包含写像が存在.

$$\begin{array}{ccc} X & Y & Z \\ \downarrow & \downarrow & \downarrow \\ X & Y & Z \end{array}$$

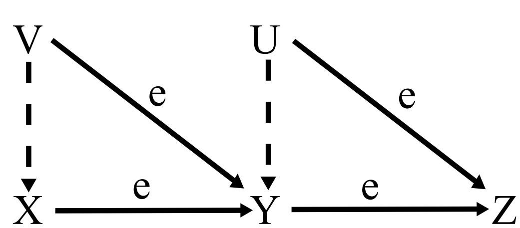
- Q_2 から Q_1 への包含写像.

$$\begin{array}{cccc} X & Y & Z & U \\ \downarrow & \downarrow & \downarrow & \downarrow \\ X & Y & Z & Y \end{array}$$

- Q_2 から e(U,X) を削除しても 同値な問合せ Q_1 が得られるので, e(U,X) は無駄な条件.

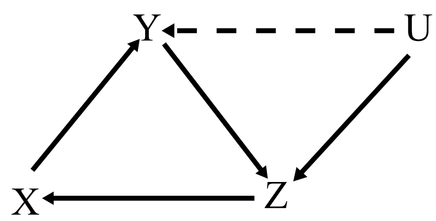
例　等価でかつ原子式の数が最小の問合せの生成.

- Q_2
  p(X,W) :- e(X,Y) & e(U,Z) & e(Y,Z) & e(V,Y) & e(Z,W)
- Q_1　p(X,W) :- e(X,Y) & e(Y,Z) & e(Z,W)



例　等価でかつ原子式の数が最小の問合せの生成.

- p(X,Z) :- e(X,Y) & e(U,Z) & e(Y,Z) & e(Z,X)
- p(X,Z) :- e(X,Y) & e(Y,Z) & e(Z,X)



問題

- Q_1:　h(X) :- red(X,Y) & blue(Y,Y) & red(Y,Z)
  Q_2:　h(X) :- red(X,Y) & blue(Y,Z) & red(Z,W)
  Q_3:　h(X) :- red(X,Y) & red(Y,Z)
- Q_1, Q_2, Q_3 の包含関係を調べよ.