公開ポリシを考慮したデータ交換フレームワークにおける問合せクラスの拡張

2016SC098 山口流星
指導教員：石原靖哲

1　はじめに

近年，ビックデータを利用したビジネスなどにより，異なる企業間でのデータの共有が行われている．データを共有する際データベースの構造が異なる場合がある．そのため，異なるデータベース構造間でデータをやり取りするデータ交換 [1] の技術に注目が集まっている．このようなデータ交換フレームワークにおいて個人情報などの重要な情報はデータ公開に制限を設ける必要がある．

本研究では，データ交換フレームワークにおいてソース側データベースに対する公開ポリシが与えられている状況を想定し，ターゲット側においてその公開ポリシをどのように反映させるのかについて考える．福嶋の研究 [2] では，ターゲット側公開ポリシとマッピングによる問合せがソース側公開ポリシの問合せよりも解像度が低い [3] ことを必要条件と考え，ターゲット側公開ポリシが満たすべき安全性要件の定式化を行っている．さらに，安全性要件を満たすようなターゲット側公開ポリシを導出するアルゴリズムを提案している．しかし，現状では自己結合のない連言問合せのクラスに制限されている．自己結合とは同じ関係に結合演算を適用することであり，自己結合を許すことにより表内で一部の値が重複する組の列挙などを行うことができる．そのため，本研究では問合せクラスを自己結合のある連言問合せに拡張した際のターゲット側公開ポリシの導出アルゴリズムを提案することを目標とする．

2　先行研究

文献 [2] では，問合せクラスを自己結合のない連言問合せとして，問合せ解像度という概念を用いて安全性を定義し，安全性を満たすようなターゲット側公開ポリシの導出アルゴリズムが提案されている．

2.1　問合せ解像度

問合せの解像度とは直観的にデータベースのインスタンスの違いを識別する能力である．以下，図 1 のデータベースインスタンス D_1, D_2, D_3 を用いて説明する．

問合せを $\pi_{\text{病名, 性別}}$ にすると D_1 と D_3，D_2 と D_3 は区別できるが D_1 と D_2 は区別できない．また問合せを $\pi_{\text{性別}}$ とすると D_1, D_2, D_3 に対してすべて同じ問合せ結果となり，区別できない．このように問合せがデータベースインスタンスを区別する能力のことを問合せ解像度という．そしてデータベースインスタンスの集合を細かく区別できるほど問合せ解像度は高いという．したがって，D_1, D_2, D_3に対しては$\pi_{\text{病名, 性別}}$の方が$\pi_{\text{性別}}$よりも解像度は高い．このことを式では$\pi_{\text{病名, 性別}} \succeq \pi_{\text{性別}}$と表す．

D_1

病名	年齢	性別
がん	40代	男
がん	60代	男
肺炎	50代	女
肺炎	80代	男

D_2

病名	年齢	性別
がん	50代	男
がん	60代	男
肺炎	60代	女
肺炎	70代	男

D_3

病名	年齢	性別
がん	60代	男
肺炎	50代	男
がん	70代	女
肺炎	60代	男

図 1　データベースインスタンスの例

2.2　データ交換フレームワーク

本研究で扱うデータ交換フレームワークを図 2 に示す．

- S：ソース側データベースインスタンス
- ソース側公開ポリシ V：S からデータを制限して公開する問合せ
- ビュー A：S から公開ポリシ V によって得られる問合せの結果
- マッピング M：データベースの構造が異なるソース側データベースからターゲット側データベースへのデータ交換手続き
- T：S から M を経由して得られたターゲット側データベースインスタンス
- ターゲット側公開ポリシ W：T からデータを制限して公開する問合せ
- ビュー B：T から公開ポリシ W によって得られる問合せ結果

2.3　安全性要件の定式化

ソース側公開ポリシを V，マッピングを M，ターゲット側公開ポリシを W とする．

- (秘匿性に関する条件)：ある連言問合せ X が存在して 2 つの合成問合せ $W \circ M$ と $X \circ V$ が等価となる．

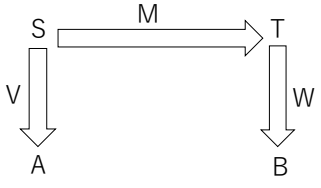

図 2　本研究で扱うデータ交換フレームワーク

- (可用性に関する条件)：上記の秘匿性に関する条件を満たす任意の W'に対し，$W' \circ M \succeq W \circ M$ ならば $W \circ M \succeq W' \circ M$ となる．

2.4 ターゲット側公開ポリシの導出アルゴリズム

自己結合のない連言問合せのクラスにおいて，A と T を Datalog の記法で書き直す．

- $A(Y) : -S_1(X_1)..., S_n(X_n), C_V(X_1....X_n)$
- $T(Z) : -S_1(X_1)..., S_n(X_n), C_M(X_1....X_n)$

ただし，ソース側データベースインスタンスの関係を S で表し，問合せの条件を C_V, C_M で表す．この状況で安全性要件を満たす W を求めるアルゴリズムの方針を以下に示す．

1. $A(Y)$ と $T(Z)$ の規則を比較し，一方の規則に存在するが他方の規則に存在しない項集合を特定する．
2. 1. で特定した項集合を存在しない方の規則に追加することで，同型の規則が導出できるか判定する．
3. 2. で同型の規則が導出できたとき，$T(Z)$ の右辺に追加した項集合が W である．

3 自己結合のあるターゲット側公開ポリシの導出アルゴリズムの提案

3.1 問合せクラスを拡張した際の課題

問合せクラスを自己結合のある連言問合せに拡張すると，項の対応を考慮する必要がある．例えば，以下の Datalog を考える．

- $A(Y) : -S(A, B, C), S(B, A, C)$
- $T(Z) : -S(A, B, C), S(C, A, B)$

この場合 $A(Y)$ の $S(A, B, C)$ が $T(Z)$ の $S(A, B, C)$ に対応するのか，$S(C, A, B)$ に対応するのかの 2 通りが考えられる．一般的には項の対応の数だけ W の候補が存在するといえる．

3.2 提案アルゴリズム

問合せクラスを自己結合のある連言問合せに拡張した際のターゲット側公開ポリシの導出アルゴリズムを，2.4 節のアルゴリズムをサブルーチンとして用いる形で提案する．

- $A(Y) : -S(A, B, C), ..., S(B, A, C), C_V(X_1....X_n)$
- $T(Z) : -S(A, B, C), ..., S(C, A, B), C_M(X_1....X_n)$

の場合について考える．

1. Datalog の述語 $A(Y)$ の $S(A, B, C)$ と $T(Z)$ の S の 1 つと対応付けを決めてから 2.4 節のアルゴリズムを開始する．
2. 2.4 節の 1.の動作で特定される項を対応付けに応じて変形する．
3. 2.4 節のすべての動作が完了したら対応付けを変更して同じ動作を行う．すべての対応付けにおいて動作が完了したとき終了する．

この条件を 2.4 節アルゴリズムに追加することによって W が導出される．

4 提案アルゴリズムにおける制約ソルバの利用

Sugar[4] とは，制約充足問題 (CSP) を解く制約ソルバである．CSP とは与えられた制約を満たす解を探索する問題である．Sugar では入力として与えられた整数の有限領域上の CSP を SAT に符号化した後，外部の SAT ソルバを用いて SAT の解を求め，それを元の CSP の解に変換することにより解を求めることができる．本研究で提案したアルゴリズムにおいて導出される 2 つの問合せ式 $W \circ M$ と $X \circ V$ を制約ソルバである Sugar を用いて等しいことを判定できるか検討した．2 つの問合せを Datalog の記法で書き直す．

- $B(Y) : -S^1(X_1, X_2, ..., X_n), S^2(X_{n+1}, ..., X_{2n}), ..., S^m(X_{nm+1}, X_{nm+2}, ..., X_{2nm})$
- $B'(Z) : -S^1(X'_1, X'_2, ..., X'_n), S^2(X'_{n+1}, ..., X'_{2n}), ..., S^m(X'_{nm+1}, X'_{nm+2}, ..., X'_{2nm})$

上記の式に対して以下に示す制約式の構成方針を提案した．

(Rule1) 同じ名前の変数を定義する．

(Rule2) S の対応を総当たりで割り当てる．

5 まとめ

本研究では，自己結合のある連言問合せのクラスにおいてターゲット側公開ポリシの導出アルゴリズムを提案した．提案アルゴリズムが安全性要件を満たすことの証明も行った．また制約ソルバ Sugar を用いてアルゴリズムによって導出される 2 つの問合せが等価となることを判定する制約式の構成方針を提案した．今後の課題としては，本研究で提案した Sugar を用いた制約式の構成方針の動作検証することである．また，マッピングを一般的なスキーママッピングのクラスに拡張することも今後の課題である．

参考文献

[1] Pablo Barceló. Logical foundations of relational data exchange. *SIGMOD Rec.*, Vol. 38, No. 1, pp. 49–58, June 2009.

[2] 福嶋啓二, 石原靖哲, 藤原融. データ交換フレームワークにおける問合せ解像度に基づいたデータ公開. 電子情報通信学会技術研究報告, SS2018-44, pp. 103–108, 2019.

[3] 廣田祐一, 橋本健二, 石原靖哲, 藤原融. データベースの推論攻撃に対する問合せ解像度に基づいた安全性定義の提案. コンピュータセキュリティシンポジウム 2008 論文集 B5-2, pp. 467–472, 2008.

[4] 田村直之, 丹生智也, 番原睦則. SAT 型制約ソルバー Sugar と Scala インターフェイス について. 日本ソフトウェア科学会第 28 回大会講演論文集, 2011.