平成26年度 情報工学コース卒業研究報告要旨

酒井 研究室	氏　　名	高 山　隆 之 介
卒業研究題目	関係データベースにおける問合せに関する ℓ-多様性のモデル計数ソルバを用いた判定法	

あらゆる企業や自治体などが関係データベースを用いて情報を保持，公開している昨今，機密情報の保護は重要な問題であり適切なセキュリティ下での管理が必要である．個人に関する情報を含む関係データに対して，全体データは公開せずに許可されたいくつかの問合せを通してデータが公開される状況において，それらの問合せとその結果から元の関係データを推論することにより重要な機密情報を盗む試みのことを推論攻撃と呼ぶ．直接的に機密情報を得るような問合せは禁止している場合にも，そのような推論攻撃に対して必ずしも安全であるとは限らない．そのため，関係データベースの管理者は推論攻撃によって機密情報が漏洩する可能性を把握し，必要に応じて一部の問合せを禁止するなどの対処を行う必要がある．

公開される関係データが推論攻撃に対してどの程度安全かを示す尺度の 1 つに ℓ-多様性がある．属性の集合は，それらの属性の値からレコードが一意に定まるとき準識別子と呼ばれ，データを公開する際には，準識別子の値でレコードが特定できないように，それらの値は加工または一部削除される．ℓ-多様性は，そのような関係データにおいて準識別子の値が同一であるレコードの各グループに対して機密属性の値の候補が少なくとも ℓ 個以上存在することを意味する．しかし，いくつかの問合せを通してデータが公開される場合，各々の問合せ結果が ℓ-多様性を満たしていても，複数の問合せの内容や結果を考慮した推論攻撃に対して安全であるとは限らない．

本研究では，問合せとその結果から推論して得られるデータ候補集合における ℓ-多様性を考え，それを問合せに関する ℓ-多様性と定義する．そして，関係データが問合せに関する ℓ-多様性を満たすかどうかを判定する方法を提案する．

提案手法では，問合せとして自己結合を含まない連言問合せを対象とし，モデル計数ソルバを用いて判定を行う．具体的な判定手順は，まず問合せとその問合せ結果から元の関係データに含まれる可能性のあるレコードの条件を表す制約式を生成する．次に，元の関係データの各レコードに対して，準識別子を表す変数の値をそのレコードの準識別子の値で固定する制約を制約式に加えたときに，制約式を満たすような機密属性を表す変数の値の候補の計数を行う．その候補数が ℓ 個未満であるレコードが存在しないとき，元の関係データは問合せに関する ℓ-多様性を満たすと判定する．機密属性の値の候補の計数には，制約ソルバである Sugar で制約式を CNF 式に変換し，モデル計数ソルバである sharpCDCL を用いた．

評価実験として，用意したレコード数 5000 の関係データと問合せに対して問合せに関する ℓ-多様性の判定を行い，制約式から変換された CNF 式の節数や機密属性を表す変数 (注目変数) の個数，判定にかかった時間の計測を行った．実験の結果，制約式の生成と制約式から CNF 式への変換は 1 秒未満で済み，実行時間のほとんどはモデル計数ソルバの実行時間であった．また，レコード数が増加するにつれて節数や注目変数の数が増加し，1 つのレコードに対する実行時間が長くなった．それに加え，レコード数が増加すると機密属性値の候補の計数を行う回数が増えるため，モデル計数ソルバの実行時間は大幅に長くなることもわかった．具体的には，属性数 11，レコード数 5000 の関係データに対してはモデル計数ソルバの実行時間は 118 秒であったが，その関係データのレコード数を 30000 に増やした関係データに対しては 2 時間半程度であった．