様式Ｃ－１９

科学研究費助成事業（科学研究費補助金）研究成果報告書

平成２４年　５月１８日現在

機関番号：３２６８９
研究種目：挑戦的萌芽研究
研究期間：2009～2011
課題番号：21650022
研究課題名（和文）　情報損失を最小化可能なプライバシー保護のための匿名化アルゴリズム

研究課題名（英文）Privacy-preserving anonymization algorithm　with information loss minimization

研究代表者
岩井原　瑞穂（ IWAIHARA　MIZUHO ）
早稲田大学・理工学術院・教授
研究者番号：40253538

研究成果の概要（和文）：
個人情報を含むデータベーステーブルの匿名化問題において,個人の属性値からプライバシー属性値への推論の確率(精度)と逆方向の推論の確率をプライバシー制約として与え，さらに情報損失をコスト関数として定義し，これらの条件を満たす最適な匿名化を求める手法を開発した．さらに関連して大規模分散環境におけるアクセス制御手法，およびソーシャルネットワークサービスにおけるプライバシー設定の支援手法について研究した．

研究成果の概要（英文）：
For anonymizing database tables containing personal information, we proposed a new framework such that privacy constraints are given as inference probabilities (accuracies) between person-identifying attributes and sensitive attributes. Also, information loss due to anonymization is modeled as cost functions, and an algorithm that finds an optimal anonymized table was developed.　As related research, we also developed access control methods for globally distributed environments, as well as user assistance for privacy settings in social networking services.

交付決定額

（金額単位：円）

	直接経費	間接経費	合　計
2009(21)年度	1,400,000	0	1,400,000
2011(22)年度	1,000,000	0	1,000,000
2012(23)年度	800,000	240,000	1,040,000
年度			
年度			
総　計	3,200,000	240,000	3,440,000

研究分野：
科研費の分科・細目：　情報学，メディア情報学・データベース
キーワード：データベース，セキュリティ，プライバシー，アルゴリズム，匿名化

１．研究開始当初の背景

医療情報や顧客情報等にデータマイニングを行い，有用な知識を引き出すことは大変重要である．一方，個人を特定する情報やプライバシー情報が含まれるこれらのデータを分析する場合は，個人のプライバシーを侵害しないように，データの匿名化を行うことが必要である．氏名や電話番号など個人を特定する情報(ID 属性)は除去できる一方で，個人の特定には至らないが該当する個人の集合を狭めるため，組み合わせることにより個人が特定可能になる属性（準 ID 属性-例えば性別，年齢，住所の郵便番号）が知られている．ID 属性を除いたテーブルからさらに準

ID 属性による個人の特定を阻止するために，テーブルに含まれる値の匿名化を行うことが必要であり，その代表的な匿名化操作として一般化（例えば「25 歳」を「20 代」に置き換える）と除去(性別を非開示にする)がある．また，匿名化の指標として k 匿名性(１つのタプルの準 ID 属性に該当する個人は少なくとも k 人以上いる)がよく知られている．一方，k 匿名性の弱点も指摘されており，ひとつは攻撃者がつきとめたい個人の準 ID 属性値を知っている場合，あるいは住民票や名簿など外部知識を有する場合には，その個人の準 ID 属性値にマッチする匿名化されたタプルを求め，そこからプライバシー属性の値が判明する場合がある．l 多様性という 1 つの準 ID 属性のタプルごとに少なくとも l 個のプライバシー属性値が対応しなければならないというプライバシー制約でこのような推論を防げる．他にもプライバシー制約や匿名化アルゴリズムに関する多くの論文が著名な国際会議で発表され，近年のデータベース分野における最も活発な研究テーマの１つとなっている．

２．研究の目的

本研究は，これまでとは異なるテーブルの匿名化手法として，確率テーブルを用いた手法を開発する．確率テーブルとは，通常のタプルにそのタプルの情報が真であるための確率を付与した確率タプルの集合である．確率テーブルを匿名化テーブルとして用いることにより，１つの準 ID 属性のタプルについて，対応するプライバシー属性値の集合を確率分布付きで表現できる．これにより，テーブルの利用者は真の値を知る代わりに値の確率分布を分析に用いることができる．このような従来とは異なる匿名化手法の開発と共に，アクセス制御やソーシャルネットワークサービスにおけるプライバシーなど，新たな分野への応用を試みる．

３．研究の方法

(1)　個人からプライバシー属性値への推論確率(IS 確率)，およびプライバシー属性値から個人への推論確率(IS 確率)による漏えい確率モデル(IS-SI 制約)の定式化

任意の個人の属性値から，匿名化テーブルを探索したときに得られるプライバシー属性値の確率(IS 確率)を一定値 β 以下，逆に病名や高所得など知りたいプライバシー属性値から，その値を持つ個人が特定される可能性(SI 確率)を一定値 γ 以下とするモデルを定式化する．IS 確率では，既存のモデルで考慮されていなかった，個人が複数のプライバシー属性値を持つ場合（複数の病名を持つ等）に自然に適用できる．

(2)　情報損失が最小可能な対応増加による匿名化アルゴリズム

従来の一般化や除去による匿名化アルゴリズムは，情報損失を最小化しながら k 匿名性などのプライバシー制約を満たすのは NP ハードであることが知られ，良い近似解を求める研究にシフトしていた．これに対し本研究では，一般化や除去など知られている匿名化とは異なるアプローチとして，準 ID 属性タプルとプライバシー属性値の対応を表す二部グラフの枝を追加することにより，IS 制約および SI 制約を充足する解を求める匿名化アルゴリズムを開発する．この対応増加によれば，情報損失を枝のコストとして表現でき，情報損失最小の対応増加を最小コスト流問題として多項式時間で解くことができる．

(3)　対応増加による匿名化アルゴリズムの実装と評価

(2)の匿名化アルゴリズムを実装し，実データを用いた性能評価を行う．評価の指標としては，実行時間および多様な情報損失コストモデルについて比較する．さらに提案手法と既存手法について，情報損失や効用の面から匿名化の結果の比較評価を行う．

(4) IS-SI 確率モデルの拡張

発展課題として提案手法の概念を，位置データやソーシャルグラフなどプライバシー情報を含む多様なデータの匿名化に適用できるように，モデルを拡張する．

(5) プライバシーデータの扱いと密接な関わりのあるアクセス制御について，近年重要性を増しているクラウドコンピューティングを想定した大規模分散環境を対象とし，アクセス制御の対象であるデータ（モノ）とアクセス主体（人）双方について効率的処理方法を開発する．具体的には，データについては XML 文書を対象とし，アクセス主体については大規模分散環境におけるロールベースアクセス制御を対象とする．

(6) ソーシャルネットワークサービスでは，利用者が自分のプライバシー情報を管理するために，開示範囲の設定を行うが，適切な設定の推薦を行うために，多くの利用者のプライバシー設定を収集する．プライバシー設定は非開示であるため，参加する利用者にプライバシー設定を提供してもらう手法と，公衆からアクセス可能な属性から，その利用者は公衆に開示していると判断する手法を組み合わせる．

４．研究成果

匿名化アルゴリズムに関して，準 ID 属性の組み合わせからプライバシー属性値への推論確率（IS 確率）およびその逆方向の推論確率（SI 確率）というあらたな匿名化制約を定式化した．さらに拡張として，presence attack という攻撃手法への対応を検討した．

Presence attack とは，匿名化テーブルにある個人が含まれているかを，個人の準 ID 属性値の一部を用いて攻撃者が特定するものである．このために，新たに PI 確率を導入し，個人の持つ準 ID タプルが匿名化テーブルに含まれる確率が PI 確率制約の範囲内に含まれるという PI 制約を導入した．

次に，匿名化により生じる情報損失を定量的に評価するモデルならびに，情報損失最小化の目標となるコスト関数の定式化を行った．コスト関数への重み付けを可能とし，コスト関数間の相対的重要度を任意に与えられるようにした．

以上で定式化した問題を解くアルゴリズムとして，問題が劣モジュラであることを利用し，最小コストフロー問題に帰着できることを示した．これにより与えられた任意の IS,SI 制約とコスト関数について，コスト最小の解を多項式時間で求められることが分かった．

PI 制約を充足させるためには，元のテーブルには含まれない個人の準 ID 属性タプルを追加して，保護する個人の準 ID 属性タプルを隠ぺいする方法を取る．PI 制約を充足させるには，やはり最小コストフロー問題への帰着を用いる．全体としては，QI－SI 制約とコスト関数最小化の問題を解き，次に PI 制約を充足させる順序を取る．

プロトタイムシステムの開発として，コスト関数の記述処理系の開発ならびに情報損失を最小化する匿名化アルゴリズムの実装を行った．

開発した匿名化アルゴリズムについて評価を行ない，5 万人規模のベンチマークデータに対する匿名化を，現実的な時間で行なえることを確認した．さらに評価を通してアルゴリズムの改良ならびに実装の効率化を行った．また情報損失の最小化目標として与えるコスト関数によって，実際に情報損失が制御できることを確認した．本匿名化手法の形式的定義を整理し，既存手法との定性的比較を行った．

クラウドコンピューティングでは，アクセス制御ルールの適用は複数のサーバにまたがって行われると考えられる．XML 文書に対するアクセス制御ルールの適用を効率化するために，XML 文書を格納する関係データベースにおいて，アクセス制御ルールの評価結果をキャッシュする方式を考察した．本方式では XML 文書が一部更新されても，更新されていない部分の評価結果は再利用できるという効率化を行った．

またアクセス制御には利用者をロールとして抽象化するロールベースアクセス制御がある．企業内の大規模分散環境において，ロールアクセス制御ポリシーの実装するには，大域的なポリシーを分散したノードにおいて実装する必要がある．異なる要求事項を持つ分散ノードそれぞれに適したロールアクセス制御ルールをモデル駆動で設計する手法を開発した．

ソーシャルネットワークサービスでは，利用者の個人属性や写真等を一般に公開するか，Friend リンクでつながれた利用者にのみ公開するかといったプライバシー設定機能がある．本機能は詳細な設定が可能になる一方で煩雑化しており，プライバシー設定の支援が必要になる．プライバシー属性オントロジーを用いて，類似する概念からリスク値を求める手法を開発した．プライバシー設定を収集し，その統計的傾向および開示属性の共起関係から，利用者に適した情報開示設定を推薦する手法の開発を行った．

さらに，最大のソーシャルネットワークサービスである Facebook について，利用者が公開あるいは非公開としている項目と公開している個人属性の収集を行なった．プライバシースコアという情報の公開度を表わす指標と，個人属性の相関を求め，相関の強さの情報を用いて，個々の利用者にプライバシー設定の推薦を行なう手法を開発した．

５．主な発表論文等

（研究代表者、研究分担者及び連携研究者には下線）

〔雑誌論文〕（計 4 件）

1. Toshikazu Munemasa and Mizuho Iwaihara, “Trend Analysis and Recommendation of Users' Privacy Settings on Social Networking Services,” 3rd Int. Conf. Social Informatics (SocInfo2011), Lecture Note in Computer Science 6984, pp. 184-197, Oct. 2011 (査読あり).
2. 近藤誠一, 岩井原瑞穂, 吉川正俊, 小宮崇, 虎渡昌史, ``異種分散環境におけるロールベースアクセス制御のモデル駆動設計手法," 情報処理学会論文誌 Vol. 52, No. 5, pp. 1882-1898, 2011 年 5 月 (査読あり).
3. Erwin Leonardi, Sourav S. Bhowmick, and Mizuho Iwaihara, “Efficient Database-Driven Evaluation of Security Clearance for Federated Access Control of Dynamic XML Documents,” Proc. Database Systems for Advanced Applications (DASFAA2010), Lecture Note in Computer Science 5981, pp. 299-306, April 2010 (査読あり).

4. 近藤 誠一, 岩井原 瑞穂, 吉川 正俊, 虎渡昌史, “異種分散環境におけるロールベースアクセス制御の定量的リスク評価,” 情報処理学会論文誌, Vol. 50, No. 11, pp.2727-2739, Nov. 2009 (査読あり).

〔学会発表〕(計 5 件)

1. Liming Shen, and Mizuho Iwaihara, "Trend Analysis of Privacy Settings and User Classifications in Social Network Services," DEIM2012 Forum, F10-2, Mar. 2012.
2. Ding Xiaochen, Mizuho Iwaihara, “Evaluating Risks for Card-style Identity Management,” 第 9 回情報科学技術フォーラム, D-018, 2010 年 9 月.
3. 宗政俊一, 岩井原瑞穂. “ソーシャルネットワーキングサービスにおけるプライバシ設定の収集と集約,” 第 9 回情報科学技術フォーラム, D-017, pp. 129-130, 2010 年 9 月.
4. Fan Zhongyi, Mizuho Iwaihara, “Utilizing Social Annotations for User's Privacy Policy Management,” 電子情報通信学会 データ工学研究会報告 DE2010-16, pp. 13-18, August 2010.
5. 岡野光太郎, 岩井原瑞穂, Gail-Joon Ahn, 吉川正俊, “ソーシャル・ネットワーキング・サービスの相互接続におけるリスク評価を用いたアクセス制御ポリシー設定支援,” 電子情報通信学会第 16 回 Web インテリジェンスとインタラクション研究会, 2009 年 10 月.

６．研究組織

(1)研究代表者

岩井原 瑞穂, 早稲田大学, 理工学術院, 教授, (40253538)