分担研究報告書

全国がん登録情報の利用及び提供における情報の特性と安全管理措置に関する研究　（23EA1032）

研究分担者　南　和宏　統計数理研究所　学際統計数理研究系・教授

研究要旨
匿名化手法の包括的な調査、整理を行い、大域的一般化処理、局所的一般化処理、クラスタリングの各手法の代表的なアルゴリズムを実装した。さらに共通の実証基盤、データセットを整備し、匿名データの安全性、有用性に関する包括的な評価実験を行い、各手法の優位性を比較した。

A．研究目的

匿名化はがん登録情報に含まれる調査客体の識別を防止するために有効なプライバシー保護技術であるが、これまで様々な匿名化手法が提案されており、個別具体的な要件に応じてどの手法を選択すべきかは明らかでない。本研究では、k-匿名化の代表的な手法を開発し、共通の評価基盤上で実証的な評価を行うことで、安全性と有用性のトレードオフの関係性、手法間の優位性を明らかにすることを目的とした。

B．研究方法

k-匿名化を実現する代表的なアルゴリズムであるIncognite, Mondrian, K-member, Datafly等をPythonで実装し、各手法をUSセンサスのAdultデータセットに適用し、匿名データを作成した。匿名データの安全性はメンバーシップ攻撃手法の成功率で評価し、有用性については複数の汎用的な指標と具体的な予測モデルの予測精度で網羅的な評価を実施した。

C．研究結果

匿名データの有用性と安全性についてはパラメータkの値を変えることでトレードオフのバランスを調整できることを確認した。有用性については、特にクラスタリング手法を用いる匿名化アルゴリズムが特に高い性能を示し、データレコードの柔軟なグループ化が有用性保持に重要であることが示唆された。計算性能については、多くのアルゴリズムがデータの次元に対し、指数的な計算時間になるのに対し、トップダウン方式でデータ空間を分割するMondrianアルゴリズムの計算性能が高いことが確認された。

D．考察

全般的に柔軟に元データを分割するクラスタリング方式が様々な有用性指標で優れた成績を示した。また様々な汎用的な有用性指標があるなかで、Normalized Certainty Penalty (NCP) という情報量に基づく指標が実際の分析指標の性能と強い関係性を示しており、汎用的な有用性に基づき匿名化手法を選択できる可能性を示すことができた。

E．結論

代表的な匿名化アルゴリズムを実装し、統合的な評価基盤を整備することで、様々な匿名化アルゴリズムの安全性、有用性に関する定量的な比較が可能であることを示した。今後は更にl-多様性、差分プライバシーといった他の重要な安全性指標の実現手法を評価する基盤として拡充する予定である。

F．健康危険情報

G．研究発表

1.　論文発表

なし

2.　学会発表

1.杉山 拓海, 南 和宏. Empirical Evaluation of Differentially Private Microdata Generated via Frequency Tables with Additive Laplace Noise . 第110回コンピュータセキュリティ・第60回セキュリティ心理学とトラスト合同研究発表会, 2025年7月.

2.Tran Chanh Minh, Kazuhiro Minami. Evaluation of Data Utility, Classification Consistency, and Vulnerability of k-Anonymization Algorithms. 2026年暗号と情報セキュリティシンポジウム. 2026年１月.

3.Bisri Musthafa, Takuo Yamaguchi, Kazuhiro Minami. Enforcing l-Diversity via an Extension of the Incognito Algorithm for k-Anonymization. 第112回コンピュータセキュリティ合同研究発表会. 2026年3月.

H．知的財産権の出願・登録状況

1. 特許取得

なし

2. 実用新案登録

なし

3.その他

なし