

厚生労働科学研究費補助金（がん対策推進総合研究事業）
分担研究報告書
匿名化手法の検討・評価に関する研究

研究分担者 南 和宏 統計数理研究所教授
研究代表者 東 尚弘 国立がん研究センターがん対策研究所がん登録センターセンター長

研究要旨： がん登録情報に含まれる地域情報を位置座標に基づき、再帰的に分割する新規匿名化手法を開発した。提案された方法は、地理情報の粒度を柔軟に調整し、k-匿名化のためのレコードのグループが、従来の一般化手法よりもはるかに均一化できることを実証的に示した。

A. 研究目的

全国がん登録の情報には、医療機関の受診者に関する機密情報が含まれており、がん登録情報を用いた調査研究を行う際に、匿名化データからの機密情報が外部に漏洩しないような安全性の担保が必要である。現在、匿名化データの代表的な安全性指標として、k-匿名性および、その派生指標が多く提案されているが、がん登録情報に対して具体的にどの手法を選択すべきかその要件は明らかでない。本研究では、匿名データにおける安全性と有用性の適切なバランスを実現する匿名化手法の確立を目指す。

B. 研究方法

医療系マイクロデータであるがん登録情報を対象に地域情報の匿名化を柔軟に行う匿名化アルゴリズムを開発し、その有用性に関する実証的評価を行った。具体的には、がん登録情報の地域情報に国土交通省の位置参照情報を結合し、地域の位置座標に基づき地域領域を柔軟に分割する匿名化アルゴリズムの開発を行った。この提案手法の有効性を示すため、従来の地域レベルの調整による匿名化アルゴリズムも合わせて実装し、匿名処理で生成されるグループ間の均一性を有用性の指標として両者の比較を実証的に行った。

C. 研究結果

今回の提案手法をがん登録情報の住所情報に適用したところ、既存の地域レベルを調整する匿名化アルゴリズムと比較して、グループ間の均一性を定量化する Discenability 指標において、5%から16%の改善が確認された。また同一グループに含まれる地域情報の隣接性についても従来手法の結果に比べて際立った改善が実現できることが示された。

E. 結論

位置座標を利用する匿名データのグループ化処理において、生成されたグループの粒度が、地域レベルの一般化階層に基づく一般化処理よりもより均一性が実現できることが示された。

F. 健康危険情報

なし。

G. 研究発表

1. 論文発表

特になし。

2. 学会発表

1) 南 和宏. 公的統計における匿名化の取り組み. 日本計算機統計学会 JSCS フォーラム 2023「公的統計・レジストリデータの利活用～安全性と利便性の狭間で～」. 2023

2) Yutaka Abe, Kazuhiro Minami. Matching Attacks on Non-deterministic Algorithms for Cell Suppression Problem for Tabular Data, 2022 IEEE International Conference on BigData. 2022

3) Yuxin Liu, Kazuhiro Minami. Flexible Partitioning of Geographical Information based on GPS Coordinates for k-Anonymity. 2022 IEEE International Conference on BigData. 2022

4) 南 和宏. 差分プライバシーな合成データ生成技術の動向. コンピュータセキュリティシンポジウム. 2022

5) 南 和宏. 公的統計における擬似データのプライバシー保護と安全性評価. 2022 年度統計関連学会連合大会. 2022