

令和4年厚生労働科学研究費 補助金 政策科学総合研究事業
 (臨床研究等ICT基盤構築・人工知能実装研究事業)
 分担研究年度終了報告書

大規模データの利活用研究の加速のための研究

研究分担者 渋谷 哲朗 (東京大学医科学研究所 教授)

研究要旨

本研究では AI 技術を用いた研究に関する技術的課題を明らかにし、認定事業者を利用した研究を加速する施策の提言につなげることをめざし、実際に事業者より提供された匿名加工医療情報を用いた、差分プライバシーの観点から匿名加工医療情報の AI 研究の可能性の評価を行った。

A. 研究目的

健康医療ビッグデータをこれまでの統計解析手法に加えてAIを適用させることにより、これまでに省みられることのなかった事象に光をあて、医療の質の向上・均てん化・診療支援と医療分野のイノベーションに貢献することが期待されている。このため、要配慮個人情報収集し、匿名加工された医療情報を円滑に利活用する社会的仕組みとして、医療分野の研究開発に資するための次世代医療基盤法が施行された。この次世代医療基盤法の認定事業者から匿名加工医療情報を提供頂いてICT・AI技術を利用した研究や開発が進展することが期待されている。

しかし、認定事業者は稼働したばかりであり、医学研究に用いられている事例はまだない。また、機械学習手法は多数の変数を要求する傾向がありながら、匿名加工による変数削減とリスクのトレードオフの関係下に変数の数に制約を課せられる可能性があることから、匿名加工医療情報を用いた機械学習の研究について懐疑的な見方もでてい

本研究ではAI技術を用いた研究に関する技術的課題を明らかにし、認定事業者を利用した研究を加速する施策の提言につなげることを目的とする。

B. 研究方法

第二号事業者より提供された匿名加工医療情報を用いて、差分プライバシーの観点から匿名加工医療情報のAI研究の可能性の評価を行うことで、提供された情報の評価を行った。

(倫理面への配慮)

本研究は、研究に用いられた情報は匿名加工情報であり、個人情報を明らかにすることのない研究である。なお、提供されたデータの解析には、高度に管理されたスパコンおよび、研究者のみがアクセスできる、施錠された部屋に設置されたPCのみを用いて行った。

C. 研究結果

健康医療データの利用におけるデータ保護の主要な概念としてk-匿名化と差分プライバシーがある。今回提供されたデータはすでにk-匿名化されたデータではあるが、差分プライバシーの観点では厳密には保護されたデータであるわけではなく、差分プライ

バシー保護を要求する国などで用いることはできない。このため、我々は、この2つの概念の双方を満たすデータ加工の方法として、(ε,k)-乱択匿名化とよぶ新しい手法を開発した。この手法はk-匿名化と差分プライバシーの双方の要求を満たす新しい手法であり、EMアルゴリズムと組み合わせることにより、高いプライバシーレベルを保ったまま、高精度な人工知能解析も行えることを実験を通して示した[1]。

この研究によって、差分プライバシー加工をk-匿名化に加えて行う新しい匿名化手法を開発することに成功し、将来のデータの利活用に向けた新しい視点を見出すことができた。また、この研究は提供データを用いて検証を行ったものであり、今回の提供データが実際の差分プライバシー研究および人工知能研究に活用可能な非常に有用性の高いものであることも明らかになった。

D. 研究発表

1.Akito Yamamoto, Eizen Kimura, and Tetsuo Shibuya, (ε, k)-Randomized Anonymization: ε-Differentially Private Data Sharing with k-Anonymity, In Proceedings of the 16th International Joint Conference on Biomedical Engineering Systems and Technologies - HEALTHINF, pp. 287-297. (本論文に関連して、筆頭著者の山本章人は令和4年度東京大学総長賞を受賞)

E. 学会発表

1.Tetsuo Shibuya, "Privacy preserving technologies toward safe medical science", The 47th Annual Meeting of the Japanese Society for Investigative Dermatology (JSID), December 2-4, 2022.

F. 知的財産権の出願・登録状況
 該当なし