

基礎理論とシステム基盤技術の融合による Society 5.0 のための
基盤ソフトウェアの創出

2021 年度年度採択研究代表者

2021 年度 年次報告書

菊池 浩明

明治大学 総合数理学部

専任教授

安全性と有用性の保証のあるヘルスケア匿名コホート基盤

§ 1. 研究成果の概要

初年度である 2021 年度は、本研究プロジェクトの 4 つの研究課題を設定し、達成目標を定めた。

本研究は、個人の権利利益の保護に配慮して、安全にヘルスケア情報を匿名化して提供し、生活改善や健康施策に有益な知見を得るためのヘルスケアコホート基盤の構築を目標としている。本年度は、課題 1(同意取得)、課題 2(匿名化ベンチマーク)、課題 3(差分プライベート匿名化)に着手した。

課題 1 では、国内のヘルスケアサービスのプライバシーポリシーを調査し、取得される個人情報の種別が適切に記載されているかを確認し、機械学習に適した固有表現情報を抽出した。また、各々の表現により記述されている現行のプライバシーポリシーの可読性を向上するために、国内 200 サイト (ALEXA Top 500)のプライバシーポリシーの 1,859 文から個人情報の種別、利用、第三者提供に関する抽出した 14 個のタームを選定し、それらを自動的なタグ付けするシステムを開発した。開発したシステムを 2,106 名の被験者により評価を行い、読みやすさ、理解度、簡素さの全てにおいて、ツールを用いたほうが向上するという統計的に有意な差を明らかにした。本システムの開発と評価について、成果 1)の国際会議で発表した。

課題 2 は、ヘルスケアデータの類型化を進めて、有用性や安全性を評価するための匿名化ベンチマークを目指している。本年度は、あるヘルスケア企業から提供された健康診断データと診療報酬データからなる匿名加工データについて、高血圧などの多くの疾患に罹患する相対リスクが、従来のコホート研究の結果と整合していることを確認した。

課題 3 では、理論的に安全性を保証する匿名化技術の構築を目指し、その準備として、米国 CDC から公開されているヘルスケアデータ NHANES を用いて糖尿病罹患リスクを予測するための匿名化コンテストの結果を分析した(成果 2))。

§ 2. 研究実施体制

(1) 菊池グループ

- ① 研究代表者: 菊池 浩明(明治大学 総合数理学部 専任教授)
- ② 研究項目
 - ・ 合成医療データを用いた匿名化ベンチマーク開発
 - ・ 差分プライバシーによる再識別に対して理論的な安全性を保證する加工方法の提案
 - ・ 匿名加工方式の安全性評価
 - ・ 積率母関数を用いた誤差範囲の推定

(2) 野島グループ

- ① 主たる共同研究者: 野島 良(情報通信研究機構 サイバーセキュリティ研究所 室長)
- ② 研究項目
 - ・ パーソナルデータ利用の説明に関する調査研究
 - ・ 匿名化ベンチマークにおける有用性と安全性指標の策定
 - ・ 差分プライベートな匿名化方式の提案
 - ・ 連携学習への応用
 - ・ 積率母関数を用いた誤差下限の推定

(3) 荒井グループ

- ① 主たる共同研究者: 荒井 ひろみ(理化学研究所 革新知能統合研究センター ユニットリーダー)
- ② 研究項目
 - ・ パーソナルデータ利用の説明に関する調査研究
 - ・ パーソナルデータ利用を説明する情報提示方法の開発

(4) 森グループ

- ① 主たる共同研究者: 森 由希子(京都大学 医学部附属病院 講師)
- ② 研究項目
 - ・ ヘルスケアデータの類型化と合成
 - ・ 実医療情報への適用と評価

【代表的な原著論文情報】

- 1) Sachiko Kanamori, Hirotsune Sato, Naoya Tabata and Ryo Nojima, “Construction of a Support Tool for User Reading of Privacy Policies and Assessment of its User Impact”, In Proceedings of the 8th International Conference on Information Systems Security and Privacy – ICISPP, pp. 412-419, 2022.
- 2) 菊池浩明, 馬 瑞強, 糖尿病罹患リスクを予測するヘルスケアデータの匿名化コンテスト PWS

Cup 2021, データの解析2022 年暗号と情報セキュリティシンポジウム SCIS 2022, 3B2-2, pp. 1-8, 2022.