

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)
総括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

研究代表者 木村 映善 (愛媛大学医学部 教授)

研究要旨

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、認定事業者の潜在的な利用者に関する意識調査、海外の医療情報の収集・活用に関する調査、匿名加工医療情報を用いたAI研究、次世代医療基盤法第25条下にかかるデータ融通の検討を実施した。

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きている。

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

本研究班は、2年目の活動として、以下に述べる4分野の取組を行った。

(1) 認定事業者を活用するAI研究・制度環境に関するアンケートの実施

認定事業者の潜在的な利用者である地方自治体、学術団体を選定し、Webアンケートシステムにて実施した。

(2) 海外の医療情報の収集・分析にかかる制度・事業の調査

9月に渡航し、英国で医療情報の収集、リンケージ、研究者のデータ利活用に供している組織として、NH S Digital (England); Scottish Safe Havens; Public Health Scotland; University of the West England; Imperial College London; Small Area Health Statistics Unit; Clinical Practice Research Datalinkを訪問し、聞き取り調査を実施した。ま

たその聞き取り調査及び提供された資料をベースに文献調査を実施した。

(3) 前年度から継続しているAI研究

前年度に構築した、LDI内に設置した機械学習環境を用いてLDIの匿名加工医療情報を分析した。J-MIMOから提供された匿名加工医療情報については東京大学医科学研究所のデータセンター内において解析を実行した。

(4) 次世代医療基盤法第25条下にかかるデータ融通の検討

定例会議、エンジニアWG、契約WG、審査WGを延べ30回開催し、次世代医療基盤法第25条下におけるデータ交換にかかる技術的課題、複数の認定事業者間にまたがる契約、審査の課題について検討し、最終提案にむけて資料を取りまとめた。

(倫理面への配慮)

本研究は、認定事業者より提供された匿名加工医療情報を利活用した研究や、個人情報を収集しないアンケートを実施しており、倫理委員会の付議対象ではない。LDIについては安全なデータセンター内で分析し、匿名加工医療情報を機械学習しその結果のみをLDIの確認のもとセキュリティ便で研究者に送付、確認する体制を構築し、情報セキュリティを確保している。J-MIMO提供の匿名加工医療情報の解析については東京大学において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、施設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

(1) 認定事業者を活用するAI研究・制度環境に関するアンケートの結果

アンケート依頼を送付した地方公共団体の部署は539部署となった。それらの部署からの総回答数461件中有効回答233件(有効回答率51.5%)であり、アンケート回答率は43.2%であった。31学術団体の会員からは総回答数713件があり、そのうち有効回答として474件(有効回答率66.5%)が得られた。全体的に

次世代医療基盤法・認定事業者に対する認知度は低

て、認定事業者の認知度の向上と認定事業者の利活用を推進するという2つの施策を推進する必要性が見いだされた。

(2) 海外の医療情報の収集・分析にかかる制度・事業
英国において、匿名加工情報を研究者に提供する代わりに、データ加工、分析環境も内包した**Trusted Research Environment(TRE)**というコンセプトを推進しており、TREによってはリモートアクセスも許可していることが確認された。このTREの運用にあたって英国のOffice for National Statisticsが考案した**Five Safe Models**、すなわち”安全な利用者”、”安全なプロジェクト”、”安全な施設/環境”、”安全なデータ”、”安全な分析結果”という5つの安全管理対策の軸にそった事業展開は、わが国でも認定事業者にTREのコンセプトを導入し事業拡大する際の参考になると思われる。次に、データのリンケージにあたって、直接識別子による名寄せとデータそのもののリンクの役務を**Indexer**と**Linker**という独立した組織に担わせることで個人情報露出を最低限におさえた秘匿性の高い運用を実現していることが確認された。本邦が学ぶべき点として、**TRE**、**Five Safe Models**、**Indexer**と**Linker**の分離、利用者への教育体制の拡充が確認された。

(3) AI研究

a) シーケンス解析

オーダーリング情報を対象にシーケンシャルパターンマイニングを実施、頻出医療指示パターン(SPM)の抽出を行う。そして、これらの最長共通部分を導出すると典型的な医療指示パターンを導出できている。複数の医療機関の頻出医療指示パターンを導出し比較することで、医療機関を越えて共通する指示パターンとバリエーションを抽出できることが確認された。

b) 敗血症の予測モデル

入院後の敗血症モデルを予測するための機械学習モデルを開発し、これまでに開発したモデルよりも高い指標(AUC : 0.769、F 値 : 0.528、適合率 : 0.536、再現率 : 0.523) を実現し、多施設から提供された大規模なデータを用いることのメリットが確認された。

c) (ϵ ,k)-乱択匿名化の開発

k-匿名化と差分プライバシー双方の要求を満たす新しい匿名化手法を開発し、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、統計量を高精度に維持できる可能性を提示した。

(4)次世代医療基盤法第25条下にかかるデータ融通の検討

審査WGでのとりまとめとして、複数の認定事業者にわたる申請は最初に申請を受理した認定事業者でワンストップ対応とし、その事業者が審査とリスク評価の審査結果を共有していく方向性を確認した。事業者間の申請書類、共有する項目について突合処理、検討を継続中である。契約WGでは、認定事業者間の医療情報の提供に関する契約書ひな型案を完成させた。エンジニアWGでは、データ授受の形式、手段、媒体、データの統合時のクレンジングの責任分界点、文字コードの扱いについて合意を形成した。

かった。主成分分析、順序カテゴリカル解析を通し

(5)結論

諸般の事情により、LDIの匿名加工医療情報を用いた研究が中断し、LDIとの新規の契約も停止したため、LDIの匿名加工医療情報を用いた研究と次世代医療基盤法第25条下のデータの名寄せの融通にかける検討に関して研究期間を延長することとなった。結果的には4種類のAIにかかる研究を実施し、1つは完了して国際学会の発表を実現し、当初の想定より研究範囲が広がっている。来年度に持ち越された次世代医療基盤法第25条下のデータ融通の検討、認定事業者の潜在的利用者へのアンケート、海外事例の調査を総合的に評価し、最終的に認定事業者にかかる厚生労働省行政の有意義な提言につなげていきたい。

D. 健康危険情報

本研究はリアルワールドデータ(日々の診療行為から発生したデータや、診療報酬、健診データ)の二次利用にもとづいた研究であり、侵襲性のある活動はない。また、次世代医療基盤法の認定事業者の匿名加工医療情報を用いる研究であるため、人を対象とする生命科学・医学系研究に関する倫理指針の対象外である。

認定事業者の潜在的利用者に対する認定事業者の認識にかかるアンケート調査についても、代表研究者が所属する愛媛大学において指针对象外でありIRBの付議対象とはならないという判断を受けている。

E. 研究発表

1) 論文発表

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 Bio medical Engineering Systems and Technologies - HEALTHINF, pp. 287-297.

1.木村映善、高田春樹、荒木賢二、黒田知宏、星佳芳、佐々木香織、伊藤伸介、次世代医療基盤法の認定事業者の認知状況、活用への意識に関する社会調査(投稿準備中)

2) 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.

1.木村映善、佐々木香織、伊藤伸介「英国における行政記録情報と医療情報をリンケージした利活用状況について～英国における機械学習での利用を前提とした医療情報の利活用状況の調査から～」『大規模データの公開におけるプライバシー保護の理論と応用』公的統計マイクロデータ研究コンソーシアム/統計数理研究所 共済 2022年12月9日

F. 知的財産権の出願・登録状況 該当なし