

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

Federated Learning によるデータシェアリング

研究分担者 シャルヴァアドリアン 順天堂大学 国際教養学部 講師
研究代表者 松田智大 国立がん研究センターがん対策研究所国際政策研究部 部長

研究要旨 国内、国際共同研究での個別データを外部に一切持ち出さずに、データ集約手法を導入する。オランダ国立総合がんセンターと共同で、Vantage6を開発し、日本のがん登録情報の、集計データの計算と共有のためのシステム利用の技術的実現性を実証し、実用して研究を進めることを目的とした。ミラノのRARECAREnetデータベースと、台湾と日本のアジア 2 国をステーションとした。解析を行う前に、STATA で同計算が実施できるコードを開発し、答え合わせができるようにした。各ステーションがモデルの数学的パラメータを繰り返し計算し、連合ポアソンモデルを実行する。各反復で、ステーションからの集約された統計は、中央で相対過剰死亡リスク（RER）の推定をするために結合され、RER の推定が収束したときに、アルゴリズムは終了する。Federated Learning による連携アルゴリズムは、3 カ国別々のデータベースを繰り返し分析し、STATA による計算と同じ結果を返した。国際共同研究としての頭頸部がんの相対生存率の比較にも成功し、各国間の差異についての検討を行った。欧州、日本だけでなく、今後アジア諸国においても、個別情報の国外提供が困難になると考えられる。しかしながら、過度なデータ保護体制を変えることをあきらめるべきではなく、古典的なデータ集約型の国際共同研究も、引き続き想定すべきである。

A. 研究目的

GDPR を初めとする情報保護に関する法律が多く、多くの国で適用され、匿名であろうと、個別情報の国外提供が困難となってきた。そこで、Personal Health Train (PHT) コンセプトを活用して、データ共有に関連する問題に対処することを企画した。PHT コンセプトは、複数の組織からのデータを、識別可能なデータが組織を離れることなく分析することを可能にするものであり、Vantage6 (<https://www.vantage6.ai>) は、PHT のオープンソース実装で、「連合学習 (Federated Learning)」の数学的原理を使用している。オランダ国立総合がんセンター (IKNL) は、Vantage6を開発し、この技

術は、がん登録情報の分散分析化を可能にしている。このプロジェクトの目的は、日本のがん登録情報の、集計データの計算と共有のためのシステム利用の技術的実現性を実証し、実用して研究を進めることである。

B. 研究方法

業務委託先である IKNL と共同で、実際に国際共同研究を実施する。欧州とアジアで罹患のパターンがことなる頭頸部がん (HNC) を対象とし、アジア諸国と比較してヨーロッパで観察された HNC の生存率の低さが、性別、年齢、HNC の部位や組織型などの既知の予後因子で調整し

た後も持続するかどうかを検証する

この研究では、水平に分割されたデータ（すなわち、異なる研究機関の患者コホートからデータを提供するが、類似した特性や項目を持つ）に適用した。連合学習は、計算をステーションでの部分と中央の部分に分けるといふ数学的原理に基づいている。ステーションは中央のサーバーとだけサブ計算を共有する。ミラノの RARECAREnet データベースと、台湾と日本のアジア 2 国をステーションとした。解析を行う前に、STATA で同計算が実施できるコードを開発し、答え合わせができるようにした。

各ステーションがモデルの数学的パラメータを繰り返し計算し、連合ポアソンモデルを実行する。各反復で、ステーションからの集約された統計は、中央で相対過剰死亡リスク（RER）の推定をするために結合され、RER の推定が収束したときに、アルゴリズムは終了する。

C. 研究結果

Federated Learning による連携アルゴリズムは、3 カ国別々のデータベースを繰り返し分析し、STATA による計算と同じ結果を返した。国際共同研究としての頭頸部がんの相対生存率の比較にも成功し、各国間の差異についての検討を行った。

D. 考察

Vantage6 は将来的にその解決策の一つとなりうるが、現時点では、パッケージとして完成していないため、導入に一定以上の IT の知識が必要であること、各サイトの PC 環境、ネット環境、セキュリティ環境

が大きく異なる中で、データ交換においてエラーが生じた場合の、問題解決が非常に難しい。また、基本的な解析のアルゴリズムは開発済みであるが、複雑な解析のアルゴリズム解析には非常に長い時間がかかり、検証も大きな負担がかかる。

欧州、日本だけでなく、今後アジア諸国においても、個別情報の国外提供が困難になると考えられる。しかしながら、過度なデータ保護体制を変えることをあきらめるべきではなく、古典的なデータ集約型の国際共同研究も、引き続き想定すべきである。

E. 結論

欧州、日本だけでなく、今後アジア諸国においても、個別情報の国外提供が困難になると考えられる。Federated Learning の技術が全世界的に一般化されるには、技術的なハードルを下げ、利活用のためのマニュアル等がある必要がある。

F. 健康危険情報

なし

G. 研究発表

なし

H. 知的財産権の出願・登録状況

1. 特許取得

該当なし

2. 実用新案登録

該当なし

3. その他

該当なし