大規模医療データの活用効率化と患者の個人情報保護の両立に向けて

顧　　玉　杰*

Towards Enhancing Both the Efficient Utilization of Large-scale Medical Data and the Protection of Patients' Privacy

Yujie GU*

In this study, we introduce a framework called Abstention-Aware Federated Voting (AAFV), designed to collaboratively and securely train heterogeneous local models while preserving data privacy. AAFV achieves this by incorporating two key components into the local model predictions: an abstention-aware voting mechanism and a differential privacy mechanism. To evaluate the effectiveness of AAFV, we apply it to real-world task of in-hospital mortality prediction. Experimental results demonstrate that AAFV delivers strong performance in both predictive accuracy and privacy protection.

1．研究背景と目的

近年，医療分野において機械学習は，医療画像診断や疾病リスク予測などの応用で顕著な成果を上げている．高精度なモデルの構築には多様かつ大規模なデータが必要だが，地域や医療機関ごとにデータに偏りがあり，モデル性能に悪影響を与える可能性がある．こうした偏りを軽減するには複数機関のデータを活用することが有効だが，個人情報保護法（HIPAAやGDPRなど）によりデータ共有には制約がある．この課題に対し，連合学習は，機密データを共有せずに複数機関が協調してモデルを訓練できる有望な手法である．しかし，従来のFedAvgアルゴリズム[1)]は，モデルの異質性や知的財産保護の要請に十分対応できない．また，推論攻撃によってプライバシーが侵害されるリスクも存在する．本研究の目的は，現実の医療応用においてモデル異質性とプライバシー保護の両立したフレームワークの構築である．

2．研究方法

本研究では，新たなフレームワークであるAbstention-Aware Federated Voting（AAFV）を提案した[4)]．AAFVは，ローカルクライアントのプライバシーと機密性を同時に保証するために，piecewiseメカニズムと新規なアブステンション対応型の投票メカニズムを統合するものである．AAFVについて，以下の各フェーズに分けて説明する．

1）事前学習：連合学習は協調的なシステムであり，フリーライダーは許容されない．そのため，協調学習の通信を開始する前に，各クライアントは自身のローカルデバイス上で，プライベートなラベル付きデータセットを用いてローカルモデルを事前学習する．このプロセスは，各サンプルに対応する正解ラベルと予測を一致させる典型的な教師あり学習で実行する．

2）ローカル投票：ローカルモデル間の異質性を橋渡しするため，補助的な公開ラベルなしデータセットD（以下，データセットD）を使用する．まず，各クライアントはデータセットDに対して予測を生成する．次に，プライバシー保護を強化するために，ローカル予測に対してpiecewiseメカニズム[2)]を適用し，摂動された予測を生成する．その後，各クライアントは摂動済み予測に基づいて，ローカル投票を行う．ここで，摂動の影響を軽減し，高信頼な予測を選択するために，アブステンション対応型投票メカニズムを提案する．一方，クライアントが当該サンプルに対する予測に十分な信頼がない場合，学習過程に混乱を与えないよう棄権する．このようにして得られた各クライアントの投票は，中央サーバにアップロードされ，集約処理が行われる．

3）集約処理：中央サーバは，全クライアントからアップロードされた未ラベルデータに対するローカル投票を収集し，それに基づいて集約処理を行う．まず，各サンプルに対して，陽性投票の総数をカウントする．なお，棄権された．投票は無効とみなされ，カウントには含めない．その後，中央サーバはローカル投票の多数決に基づいて，未ラベルデータセットに対するグローバル投票を生成する．この処理によって，最終的な集約予測が決定される．これら

2025年4月7日 受理
* 豊田理研スカラー
九州大学大学院システム情報科学研究院

の集約された投票は，未ラベルデータに対する擬似ラベルとして機能し，さらなる学習のために各クライアントに返送される．

4）再学習：各クライアントは，中央サーバから受け取った擬似ラベル付きの未ラベルデータを使用して学習を継続する．まず，無効なサンプルを除外することで，高信頼な擬似ラベル付きデータセットを構築する．その後，各クライアントは自身のプライベートデータセットとデータセットDを組み合わせて，ローカルモデルの精度向上を図る．この学習プロセスは，正解ラベル（または擬似ラベル）と予測結果を一致させる典型的な教師あり学習として行う．この手法により，各クライアントはプライベートなデータやローカルモデルの詳細を共有することなく，他のクライアントから学習することが可能となり，モデルの機密性および知的財産を保護することができる．

3．研究結果と考察

提案するAAFV手法の有効性を検証するために，入院患者の死亡予測という実用的なタスクを対象に実験を実施する．電子健康記録（EHR）は，研究者や医療機関に対して，入院患者の情報を分析する機会を提供する．大規模なEHRデータセットを活用することで，病院は現在の治療法の有効性を評価し，患者ケアの改善につなげることができる．本実験では，文献3）に従ってMIMIC-IIIデータセットの特徴量を抽出し，7,488次元の特徴量を持つ23,944サンプルからなる入院患者の死亡予測用データセットを構築する．

MIMIC-IIIデータセットの高次元な特徴空間に対応するため，多層パーセプトロン（MLP）を使用する．各層の後には活性化関数としてReLUを適用し，効率的な計算とモデルの適合性向上を図る．

図1に示すように，AAFVはプライバシーバジェット1において，安定かつ有効な性能を示す．全体として，3つのモデルすべてにおいて平均で約3%の精度向上が確認される．特に，MLPとSVMモデルは，FedAvgフレームワークで訓練された同等のモデルに比べて顕著な性能向上を示し，p値の平均は0.0025と，有意水準0.05を大きく下回る統計的有意性が確認されている．ですので，AAFVはテスト精度およびプライバシー保護の両面において，その有効性と機密性を示すことが確認された．

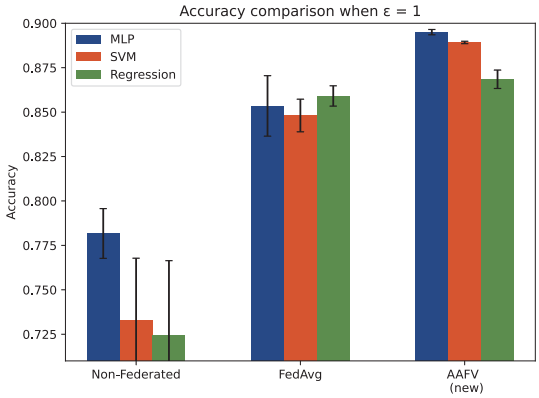

図1

4．まとめ

本研究では，新たなフレームワークであるAAFVを提案した．AAFVは異質なローカルモデルを，プライバシーを保ちつつ協調的に学習させることを可能にするフレームワークである．実医療データセットに基づく実験結果から，AAFVはFedAvgや非連合な手法と比較して，同等のプライバシーレベルを保ちつつ，テスト精度において一貫して優れた性能を示すことが確認された．今後は，より多様な実医療応用におけるAAFVの有効性を検証することが課題である．

謝辞

本研究の遂行にあたり，公益財団法人豊田理化学研究による豊田理研スカラー研究助成のご支援を賜り，心より感謝申し上げます．本研究はオークランド大学のDr. Jingfeng Zhangと九州大学の大学院生Xu氏との共同研究の成果です．

REFERENCES

1) H. McMahan, E. Moore, D. Ramage, S. Hampson and B. A. Arcas, *International Conference on Artificial Intelligence and Statistics (AISTATS)*, 2017.
2) N. Wang, X. Xiao, Y. Yang, J. Zhao, S. C. Hui, H. Shin, J. Shin and G. Yu, *IEEE Annual International Conference on Data Engineering,* 2019, pp. 638-649.
3) S. Wang, M. McDermott, G. Chauhan, M. Hughes, T. Naumann and M. Ghassemi, *arXiv:* 1907.08322.
4) Y. Xu, J. Zhang and Y. Gu, *IEEE Conference on Artificial Intelligence,* 2024, pp. 1142-1147.