1X-04

ブロックチェーン技術を活用したパーソナライズド連合学習におけるクラスタリングの基礎検討

内山　光彩　　鈴木　昇太　　小野　智司 †
鹿児島大学 †

概要

パーソナライズド連合学習（Personalized Federated Learning: PFL）は，個々の参加者が持つ分布の異なるデータに適したモデルを構築することを目的とする．本研究は，FLIS の学習をブロックチェーンネットワーク上で実行し，中央サーバを用いることなく PFL を実行する手法を提案する．

1　はじめに

近年，参加者が所持するデータを共有せずに訓練データとしての利用を許す連合学習 (Federated Learning: FL) [1] が注目を集めている．連合学習は，データに含まれるプライバシを保護しつつ，複数の参加者が共同で単一のグローバルモデルを構築することを目的とする．

連合学習では，各参加者が所持するデータの分布が異なる場合に，グローバルモデルの性能が低下してしまう課題がある．このような不均一性の課題に対し，複数のモデルを構築することで，各参加者が所持するデータに適したモデルを構築するパーソナライズド連合学習が提案されており，例えば，参加者をクラスタリングすることで類似するデータを所持する参加者との共同訓練を行う Federated Learning by Inference Similarity（FLIS）[2] が提案されている．しかし，FLIS を含む中央集権型の連合学習では，中央サーバがデータと学習過程を管理することにより，中央サーバが攻撃の標的となり，データのプライバシやセキュリティが侵害される恐れがある．

このため本研究は，図 1 に示すように，FLIS のデータと学習過程をブロックチェーン（Blockchain: BC）技術を用いて管理する手法を提案する．本手法は，FLIS の性能を維持しつつ，FLIS のデータ管理と学習過程を BC ネットワーク上で実行することで，中央サーバを用いることなく PFL を実行することが可能となる．実験により，モデルの性能を維持しつつ BC ネットワーク上で PFL を実行可能であることを確認した．

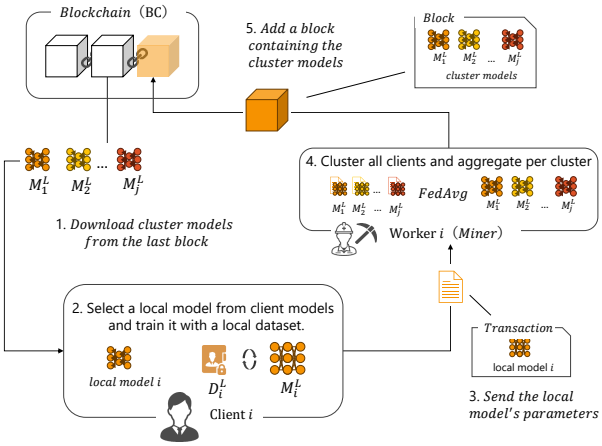

図 1: 提案手法の概念図

2　関連研究

参加者が所持するデータの不均一性に着目して複数のモデルを訓練するパーソナライズド連合学習が提案されている．FL+HC [3] は，サーバ側でモデルの重みやモデルの更新情報の比較に基づいてクラスタを形成する．しかし，参加者のデータ量は限られており，参加者がモデルを十分に訓練することが難しいため，モデルの重みやモデルの更新情報の比較に基づくクラスタリングは推奨されていない．IFCA [4] は，クラスタ数を事前にサーバ上で設定する必要がある．FLIS は，各参加者のモデルの推論結果に基づいて類似度を計算することで，FL+HC や IFCA の問題を解決して，クラスタを作成し，クラスタ単位でモデルを訓練することを可能とする．一方で，FLIS を含む中央集権型の FL および PFL は，中央サーバがデータと学習過程を管理することによる脆弱性が存在する．

A Preliminary Study on Blockchain-based Personalized Federated Learning Using Clustering
† Hiiro Uchiyama, Shota Suzuki, Satoshi Ono, Kagoshima University

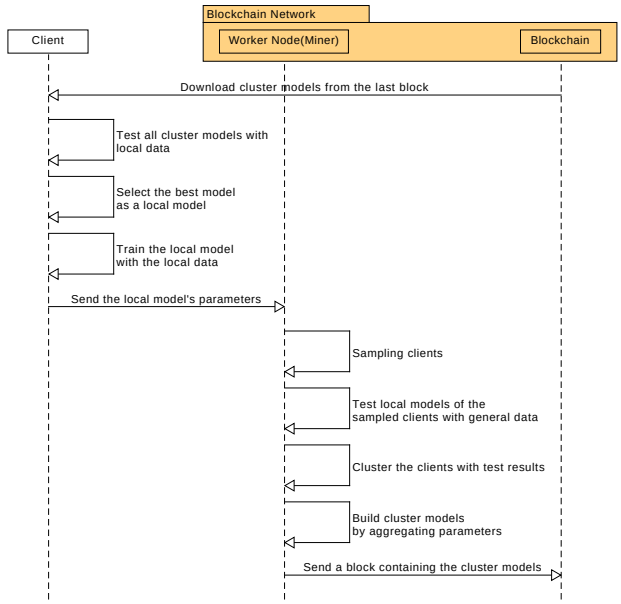

図 2: 提案手法のメインループ

3 提案手法

提案手法の構成および処理手順を図 2 に示す．各参加者は分布の異なるローカルデータを所持し，クラスタ毎のモデル（クラスタモデル）をブロックチェーンからダウンロードし，ローカルデータ上で最小の損失をもたらすクラスタモデルのパラメータをローカルモデルとする．次に，各参加者は，ローカルデータを用いてローカルモデルの訓練を行い，訓練後のモデルパラメータをトランザクションに格納し，BC ネットワーク上のワーカに送信する．

BC ネットワーク上のワーカ（マイナ）は，参加者から送付されるトランザクションを集約し，クラスタリングを行ってクラスタモデルを構築する．まず，参加者をランダムにサンプリングする．ワーカは続いて，各参加者のローカルモデルに対して一般テストデータ（補助的データまたは合成データ）を入力して推論を行う．この結果をもとに参加者間の類似度を計算し，これを要素として持つ隣接行列を構築し，クラスタリングを行う．その後，ワーカはクラスタ毎にモデルパラメータを FedAvg アルゴリズムを使用して平均化し，各クラスタモデルを構築する．最後に，全クラスタモデルを含むブロックを構築し，BC ネットワークの承認を得た後に，ブロックチェーンに最後のブロックとして追加する．これらの処理を繰り返すことで，データと学習過程を安全に管理しつつ，動的なクラスタリングに基づいた PFL を実行することができる．

表 1: FLIS と提案手法の正解率の比較

Algorithm	Accuracy
FLIS	86.37 ± 0.68%
提案手法	85.75 ± 0.68%

4 評価実験

本手法の評価を行うために，FLIS と提案手法の比較を行った．データセットは Cifar-10 を対象とし，各参加者が所持するデータのラベル分布に 20% の不均一性のある環境を作成した [5]．BlockSim [6] を用いて，シミュレーション環境を実装した．

表 1 に，各参加者が所持するテストデータにおける，各クラスタモデルの正解率を示す．表 1 から，本手法は FLIS と同程度の正解率を維持していることがわかる．

今後，実行コストや通信データ量などのスケーラビリティの検証を行う．

5 結論

本研究では，動的なクラスタリングに基づいた PFL である FLIS を，データ管理と学習過程を BC ネットワーク上で実行する手法を提案した．実験結果から，本手法が FLIS と同程度の正解率で，実行可能であることがわかった．今後は，非同期型のステムへの拡張を検討する．

参考文献

[1] B. McMahan and D. Ramage, “Federated learning: Collaborative machine learning without centralized training data,” Google Res. Blog, vol. 3, 2017.

[2] M. Morafah et al., “Flis: Clustered federated learning via inference similarity for non-iid data distribution,” IEEE Open Journal of the Computer Society, 4, 109-120., 2023.

[3] F. Z. . A. P. Briggs, C., “Federated learning with hierarchical clustering of local updates to improve training on non-iid data,” In 2020 International Joint Conference on Neural Networks (IJCNN) (pp. 1-9). IEEE., 2020.

[4] D. Y. A. Ghosh, J. Chung and K. Ramchandran, “An efficient framework for clustered federated learning,” Adv. Neural Inf. Process. Syst., pp. 19586-19597, 2020.

[5] “Federated learning on non-iid data silos: An experimental study,”

[6] M. Alharby and A. van Moorsel, “Blocksim: An extensible simulation tool for blockchain systems,” Front. Blockchain, 09 June 2020 Sec. Financial Blockchain.