F-027

ブロックチェーン技術と分散ファイルシステムを活用したクラスタ型連合学習の基礎検討

A Preliminary Study on Blockchain-based Clustered Federated Learning Utilizing Distributed File System

内山 光彩[1)]　向田 眞志保[1)]　小野 智司[1)]
Hiiro Uchiyama　Mashiho Mukaida　Satoshi Ono

1　はじめに

近年，機械学習におけるデータプライバシとセキュリティへの関心の高まりを背景に，分散型機械学習手法の一つである連合学習（Federated Learning: FL）[1] が注目されている．FL は，参加者が所持するデータを直接的に共有することなく，単一のグローバルモデルを複数の参加者が共同で学習する手法である．FL を用いることで，データに含まれるプライバシを保護しつつ，グローバルモデルの構築を行うことができる．

しかし，FL において，各参加者が所持するデータの分布が異なる場合は，グローバルモデルの性能が低下するという課題がある．この課題に対処するため，クラスタ型連合学習（Clustered FL: CFL）[2] が提案されている．CFL では各参加者が所持するデータを直接参照することなく，各参加者が所持するデータに適したモデルの構築を行う．一方で，FL や CFL では，中央サーバが処理やデータの管理を行っている．そのため，中央サーバが攻撃の標的となると，データのプライバシやセキュリティが侵害される場合がある．

このため本研究では，図 1 に示すようにブロックチェーン（BlockChain: BC）技術と分散ファイルシステム（InterPlanetary File System: IPFS）を用いたクラスタ型連合学習を提案する．提案手法では，BC ネットワーク上でクラスタ型連合学習を実行し，集約したモデルを IPFS に保存してそのハッシュ値を BC に記録する．これにより，単一の中央サーバを持たずに CFL を実現すること，ならびにモデルの整合性を保証しつつ BC のデータサイズ肥大化を抑制することが可能となる．

2　関連研究

FL においては，これまでに，各参加者が所持するデータの不均一性を対処するクラスタ型連合学習，およびプライバシやセキュリティの問題を対処する分散型連合学習が研究されている．クラスタ型連合学習においては，参加者の損失関数の幾何学的特性に基づいてクラスタを形成する Clustered-FL [3] が提案されている．しかし，各参加者のデータ量が限られている環境では，十分なエポック数でモデルの更新を行うことが難しく，適切なクラスタ形成が困難な場合がある．一方，IFCA [4] は，各参加者が属するクラスタを交互に推定し，各クラスタでモデルを最適化する手法である．しかし，クラスタ数を事前に設定する必要があり，各参加者が所持するデータの不均一性がわからない環境では柔軟性に欠ける可能性がある．この課題に対し，各参加者のモデルの推論結果を類似度としてクラスタリングした FLIS [5] が提案されている．この手法は，ソフトなクラスタリングによって柔軟なクラスタ型連合学習を実現している．さらに，各参加者が所持するデータに不均一性のない環境でも，既存手法に対する性能向上を実現している．

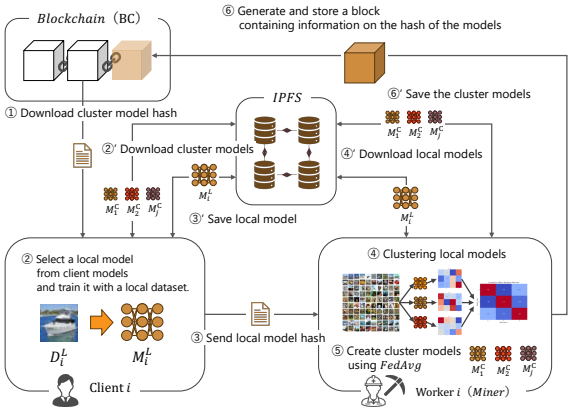

図 1　提案手法の概念図

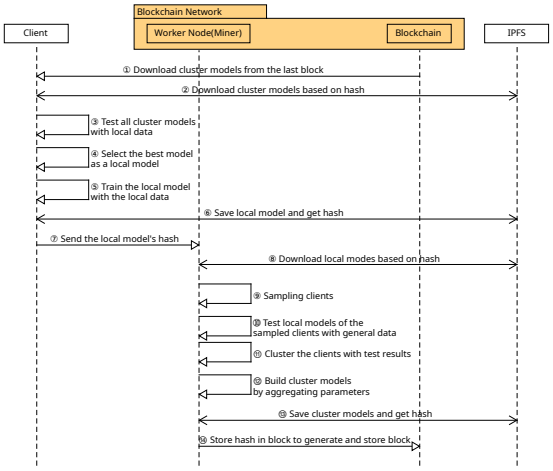

図 2　提案手法の実行フロー

分散型連合学習においては，ブロックチェーン（BC）技術が活用されている．ブロックチェーン（BC）技術は，分散型台帳技術とも呼ばれ，P2P ネットワーク上で分散型の台帳を管理する技術である．BC ネットワークは，デジタル台帳技術を基にした分散型のデータベースシステムであり，取引やデータの記録をブロックと呼ばれる単位でチェーン状に連結し，ネットワーク上の複数のノード（コンピュータ）に分散して保存する．例えば，Kim ら [6] は，BC を用いて参加者の学習の履歴を BC ネットワークにより管理し，連合学習を行うためのフレームワークを提案している．このような分散型連合学習は，連合学習におけるセキュリティとプライバシの問題に対処し，連合学習の信頼性を向上させるための有望なアプローチとして注目されている．

1)　鹿児島大学　Kagoshima University

3　提案手法

本研究では，CFL 手法である FLIS を基本とし，FLIS のデータと学習過程を BC 技術と分散ファイルシステムである InterPlanetary File System（IPFS）を用いて管理する手法を提案する．提案手法の概念図を図 1 に示す．本手法では，各参加者が，自身が有するローカルデータに適したローカルモデルを構築することを目的とし，類似するデータを所有する参加者と協力してローカルモデルの品質を向上させるためにクラスタリングを行う．

提案手法の処理手順を図 2 に示す．まず，各参加者は，ブロックチェーンの最後のブロックに格納されたクラスタモデルのハッシュ値を参照し，IPFS からハッシュ値に紐づくクラスタモデルをダウンロードする[1)]．次に，各参加者のローカルデータを用いてローカルモデルの訓練を行う．このとき，ローカルモデルの初期値パラメータは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使用する．各参加者は訓練後のモデルパラメータを IPFS に保存し，そのハッシュ値をワーカに送信する．

ハッシュ値を取得したワーカ（マイナ）は，IPFS からモデルパラメータをダウンロードする．次に，一般テストデータ（補助的データまたは合成データ）に対し，各参加者のローカルモデルによって推論を行う．ワーカはこの推論結果をもとに参加者間の類似度を計算し，クラスタリングを行う．クラスタリング後，各クラスタのモデルパラメータを FedAvg アルゴリズムによって平均化し，それを各クラスタのモデルとして構築する．ワーカは構築した各クラスタのモデルを IPFS に保存し，そのハッシュ値を BC ネットワークのブロックに格納する．最後に，格納したブロックは BC ネットワークの承認が得られた後に，BC の最後のブロックとして追加される．これらの処理を繰り返すことで，FLIS における中央サーバを用いることなく，動的なクラスタリングに基づいた CFL を実行することができる．

4　評価実験

本研究では，CFL を BC ネットワーク上で行うことによる識別性能への影響を検証するため，シミュレーションによる実験を行った．本実験では，提案手法と BC を用いずに中央サーバを用いる FLIS の性能比較，および CFL を BC ネットワーク上で行う際の 1 ブロックあたりのデータサイズについての評価を行った．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 [7]．クラスタモデル，ローカルモデルは，畳込み層を 2 層含む小規模な CNN とした．参加者の総数を 100 人，参加者のサンプリング率を 10%，通信ラウンドの上限を 50 回とした．

本実験ではまず，提案手法と従来手法である FLIS の正解率を比較した．表 1 に，FLIS と提案手法の正解率を示す．表より，提案手法は BC 技術と IPFS を用いても，FLIS と同等の正解率を維持していることがわかる．これは，提案手法が FLIS とほぼ同様の処理手順を持つためであり，BC 技術と IPFS を組み込んだことによる正解率への影響は小さいと考えられる．

1）　参加者は，自身が属するクラスタのクラスタモデルだけでなく，すべてのクラスタモデルをダウンロードする．

表 1　FLIS と提案手法の正解率の比較

Algorithm	Accuracy
FLIS	$86.37 \pm 0.68\%$
提案手法	$85.75 \pm 0.49\%$

表 2　1 ブロックあたりのデータサイズ

Algorithm	Data Size
提案手法（IPFS なし）	2.56 [MB]
提案手法（IPFS あり）	64 [byte]

続いて，提案手法における BC のデータサイズと，モデルを BC に直接保存する場合 [8] のデータサイズの比較を行った．表 2 に，各手法における 1 ブロックあたりのデータサイズを示す[2)]．表 2 から，提案手法では，モデルのハッシュ値のみをブロックチェーンに格納することで，モデルを直接保存する場合と比べて，データサイズを大幅に削減できることがわかる．

5　結論

本研究では，クラスタ型連合学習をブロックチェーン技術と分散ファイルシステムを用いて実行する手法を提案した．シミュレーション実験により，中央サーバを用いない提案手法であっても，従来のクラスタ型連合学習手法である FLIS [5] と同等の正解率を維持できることを確認した．また，提案手法では IPFS を活用することで BC のデータ容量の増加を抑制できることを示した．今後の課題として，セキュリティ面での課題に対する手法を検討，および非同期型の構成へと提案手法を拡張することが挙げられる．

参考文献

[1] Brendan McMahan and Daniel Ramage. Federated learning: Collaborative machine learning without centralized training data. Google Res. Blog, vol. 3, 2017.
[2] Christopher Briggs, Zhong Fan, and Peter Andras. Federated learning with hierarchical clustering of local updates to improve training on non-iid data. pp. 1–9, 2020.
[3] Felix Sattler, Klaus-Robert Müller, and Wojciech Samek. Clustered federated learning: Model-agnostic distributed multitask optimization under privacy constraints. IEEE transactions on neural networks and learning systems, Vol. 32, No. 8, pp. 3710–3722, 2020.
[4] Avishek Ghosh, Jichan Chung, Dong Yin, and Kannan Ramchandran. An efficient framework for clustered federated learning. Advances in Neural Information Processing Systems, Vol. 33, pp. 19586–19597, 2020.
[5] Mahdi Morafah, Saeed Vahidian, Weijia Wang, and Bill Lin. Flis: Clustered federated learning via inference similarity for non-iid data distribution. IEEE Open Journal of the Computer Society, Vol. 4, pp. 109–120, 2023.
[6] Hyesung Kim, Jihong Park, Mehdi Bennis, and Seong-Lyun Kim. Blockchained on-device federated learning. IEEE Communications Letters, Vol. 24, No. 6, pp. 1279–1283, 2019.
[7] Qinbin Li, Yiqun Diao, Quan Chen, and Bingsheng He. Federated learning on non-iid data silos: An experimental study. pp. 965–978, 2022.
[8] 内山光彩, 鈴木昇太, 小野智司. 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. 情報処理学会第 86 回全国大会, 2024.

2）　本表に示すデータサイズは，ブロックチェーンの標準構成要素を除外し，比較対象としている手法固有のデータのみを含めている．