連合学習の通信量を削減するための Energy スコアを利用した知識蒸留手法の検討

東　桔也†　　宮崎　智††　　大町　真一郎††
† 東北大学工学部電気情報物理工学科　　†† 東北大学大学院工学研究科

1. はじめに

近年，プライバシーを保護しながらデータを集約せずにモデルを学習させることができる連合学習が注目を集めている．一般的な連合学習では多数のクライアントで学習させたモデルのパラメータを集約することで，データ自体をサーバーに集約することなく大規模なデータで学習したのと同等のモデルを構築することが可能になる．一方で，モデルパラメータではなくモデルの出力のみをサーバーに集約して知識蒸留する手法も提案されている．

知識蒸留[1]は，大きなデータセットで学習済みの教師モデルから生徒モデルに知識を移行する方法の一つである．知識蒸留を連合学習に拡張した DS-FL[2]では，全クライアントで共有可能なラベルなしデータに対するモデル出力を集約して蒸留する．この手法は通信量を削減しつつ，モデルパラメータを集約する手法と同等の精度を獲得した．

本稿では，各クライアントが送信するモデル出力の Energy スコア分布に着目することで学習を効率化させると同時に，通信量を削減できる手法を提案する．

2. 提案手法

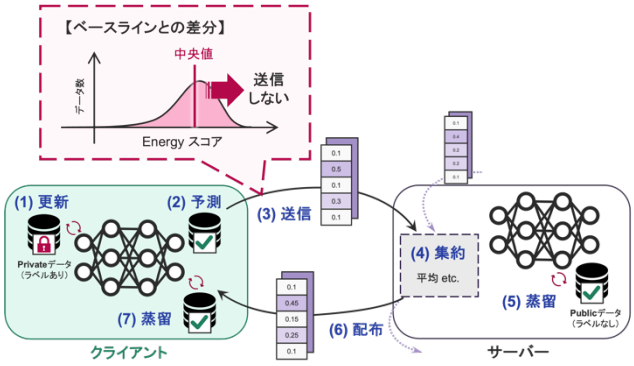

図1. 提案手法(DS-FL+)の概要図

図 1 に示した提案手法(DS-FL+)では，ラベルなしデータに対するモデル出力から(1)式の Energy スコアを算出し，中央値以上のスコアを持つデータの予測はサーバーに送信しない．

$$E(\mathbf{x};f) = -\log \sum_{i}^{K} e^{f_i(\mathbf{x})} \qquad (1)$$

ここで，$\mathbf{x}$は入力，fは分類器，Kは出力の次元数である．Energy スコアを用いることで，学習データセットにないデータを識別(分布外検出)できる[3]．

各クライアントにおいて十分に学習されたデータに対しての予測のみを送信するため，予測時に Energy スコアの閾値を設定する．ベースラインの DS-FL では，クライアントが蒸留後に自身の所有するデータでモデルを再び更新する．しかし，このデータ分布が全体のデータ分布と異なる場合(Non-IID)に，ローカルモデルでは以前蒸留したグローバルなパラメータが失われてしまう可能性がある．

Energy スコアの中央値を閾値とすることで，より正確な予測のみをサーバーに送信でき，かつ各ラウンドでサーバーに送信するデータサイズも 1/2 になるため，蒸留の効率化と通信量の削減を両立できると考えた．

3. 実験結果

クライアントとサーバーのモデルに ResNet-18 を，データセットに CIFAR-10 を使用し，それ以外の設定は[2]に従った．各クライアントが持つデータのクラス数を最大 2 とした Non-IID データの場合の結果を図 2 に示す．

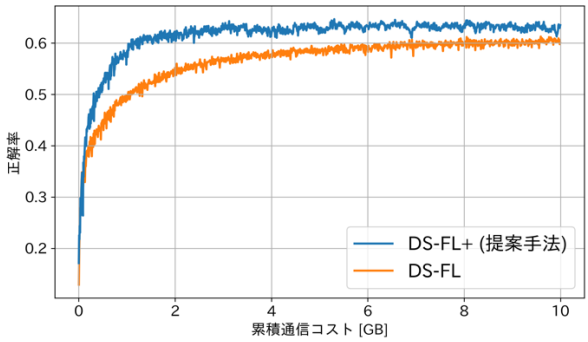

図 2. 通信コストと正解率の関係

提案手法は，ベースラインである DS-FL の約 1/5 倍の通信コストで正解率 60%に到達した．

4. まとめ

偏りが強い Non-IID データの場合に，Energy スコアの閾値を利用することで連合学習を効率化し，通信量を大きく削減することができた．

参考文献

[1] Hinton, G. et al. Distilling the Knowledge in a Neural Network. *arXiv preprint arXiv:1503.02531*, 2015.

[2] Itahara, S. et al. Distillation-Based Semi-Supervised Federated Learning for Communication-Efficient Collaborative Training With Non-IID Private Data. In *IEEE TMC*, 2021.

[3] Liu, W. et al. Energy-based Out-of-distribution Detection. *arXiv preprint arXiv: 2010.03759*, 2021.