ISS-SP-006

モデル蒸留に基づいた 分散環境下における連合学習

阿比留　祥太†　　田村　慶一††
†広島市立大学情報学部　††広島市立大学大学院情報科学研究科

1. はじめに

近年，プライバシー保護やデータ送信の帯域圧迫を回避する観点からデータをサーバに送ることなく，デバイスに分散させた状態で学習を行う連合学習が注目されている．連合学習の中でも FedMD[1]は，モデル蒸留に基づいており，サーバにすべてのデータを集める必要がなく，各参加デバイス(以下，クライアントと呼ぶ)が学習したモデルも秘匿にできるという特徴を持つ．FedMD ではモデル蒸留のためにクライアントが持つモデルに対する公開データの出力をサーバに集める必要がある．そこで，本研究では，サーバを設置できない場合を考えて，クライアント同士が直接接続された分散環境下でのモデル蒸留に基づく新しい連合学習を提案する．

2. FedMD

FedMD で用いる公開データは，通常，各クライアントから一部を収集したり，類似したドメインのデータを用いる．FedMD のアルゴリズムは以下の通りである．

(1) 各クライアントは，公開データとクライアントが持つプライベートデータを用いてモデルの転移学習を行う．
(2) クライアントは公開データに対するモデル出力をサーバに送信する．
(3) サーバは，受け取った出力を平均化しクライアントに返送する．
(4) クライアントは受け取った平均化された出力にモデル出力が近づくように知識蒸留を行う．
(5) クライアントはプライベートデータを用いてモデルを再学習する．
(6) (2)～(5)を収束するまで繰り返す．

3. 分散環境下でのモデル蒸留に基づく連合学習

分散環境下でのモデル蒸留に基づく連合学習では，モデル出力の交換を接続されたクライアントのみで行うことで学習を行う．また，分散環境下でクライアントから公開データを抽出することのリスクや情報共有ができないことを踏まえ，クライアントから集める公開データとは異なるデータ(mismatch データと呼ぶ．)を用いて蒸留を行う．

(1) クライアントは mismatch データに対するモデル出力を隣接クライアントに送信する．
(2) クライアントは隣接クライアントから受け取った出力を平均化し，平均化した出力にモデルの出力が近づくように知識蒸留を行う．
(3) プライベートデータでモデルを再学習する．
(4) (1)～(3)を収束するまで繰り返す．

4. 評価実験

評価実験には手書き文字データセットの EMNIST を連合学習用に拡張した FEMNIST を用いる．FEMNIST は，62 クラスの多クラス分類でクライアント間のデータ分布は非 IID に従う．全 40 クライアントのモデルはすべて 2 層の CNN で，各手法の評価はクライアントそれぞれのプライベートテストデータに対する精度の平均とする．

分散環境の設定は，リンク数の平均 5.5，分散 11.4 のスケールネットワークである．比較する手法は，サーバの有無と公開データが match データ/mismatch データであるかのふたつのオプションのそれぞれを組み合わせた手法とクライアントのみで学習する local-only と従来通りサーバに全データを集める centralized の計 6 つの手法である．各手法 5 回ずつ行った際の平均精度を以下の表 1 に示す．

表 1　各手法の精度

サーバ	データ	平均精度
あり	Mismatch	74.03%
あり	Match	78.25%
なし	Mismatch	73.75%
なし	Match	77.67%
Local-only		68.36%
Centralized		85.99%

実験結果からサーバを用いない場合の学習でも，用いる場合と比べて同程度に精度が向上していることが分かる．

5. まとめ

本研究では，分散環境下でのモデル蒸留に基づく連合学習手法を提案した．サーバのない場合でもクライアントのみで学習した場合より精度が向上することが分かった．今後の課題として，分散環境下での実用的なアルゴリズムとそれに伴った蒸留方法の改善が考えられる．

謝辞

本研究は広島市立大学特色研究費，科研費獲得支援研究費の助成を受けたものである．

参考文献

[1] Daliang Li and Junpu Wang . Fedmd: Heterogenous federated learning via model distillation. arXiv preprint arXiv:1910. 03581, 2019.