2E-05

サーバレス Federated Learning のための分散最適化

田谷昭仁† 戸辺 義人† 西尾理志‡ 守倉正博‡ 山本高至‡
青山学院大学理工学部情報テクノロジー学科†
京都大学大学院情報学研究科通信情報システム専攻‡

１．はじめに

画像処理や自然言語処理を中心として，機械学習の研究や実用化が盛んに進められている．一方で，学習に利用できるデータ量が多ければ多いほど性能を向上できる機械学習では，サービス提供者間にデータ収集の能力格差が生じるため，大量のユーザ数を抱え，データ収集が容易な大企業が市場を独占することが予想される．また，このような大企業がサービス停止した場合に，収集データや学習済み機械学習モデルの喪失が社会的損失となることも懸念される．

本稿では，単独のサービス提供者に依存するのではなく，小規模データを所有する個人や中小企業などが独立した学習器を学習させ，それらの連携により，大規模データを使用した学習と同程度の性能を達成する学習手法を提案する．提案手法では，SNS（Social Networking Service）などのネットワークを介して学習器同士が学習途中のモデルを共有し，最終的には全学習器がネットワーク上の全データを使って学習した場合と同等の学習モデルを獲得することを目標とする．提案手法は中央局が不要で参加する学習器同士が対等であるため，サービスの寡占化を防ぐことが可能である．

２．関連研究

機械学習の分散学習アルゴリズムとして，プライバシー保護を目的とした FL（Federated Learning）[1]が提案されている．これはスマートフォンなどのユーザ端末が端末内データにより深層学習を行い，学習モデルのパラメータをサーバにアップロードし，中央サーバでパラメータを集約することで分散学習を実現している．[2]では co-distillation と呼ばれる分散学習を提案している．co-distillation では複数の学習器が学習途中に，同一データに対する出力を共有することで，性能向上を図っている．

[1]ではサービス提供者にモデルが集約されるため，サービスの寡占化が進む問題が解決されない．また，[2]では学習器が他のすべての学習器と情報共有するため，大規模化が困難である．本稿で提案する手法は co-distillation を隣接する学習器に対して適用し，FL のアルゴリズムから中央サーバを廃することでネットワーク上のオープンな協調学習システムを実現する．

３．分散学習アルゴリズム

図 1 に提案する分散学習環境を示す．スマートフォンなどのユーザ端末iに独立した学習器が搭載され，それぞれが端末内のデータセット$\mathcal{D}_i$を利用して画像分類等の学習を行い，学習モデルf_iを更新する．各端末で利用可能なデータセット$\mathcal{D}_i$として少数の教師ラベル付きデータを想定する．すなわち，$\mathcal{D}_i$の要素はモデルの入出力のペア(x, y)である．ネットワーク上の教師ラベル付きデータ全体の集合$\mathcal{D}$を$\mathcal{D} := \bigcup_{i\in\mathcal{U}} \mathcal{D}_i$と定義する．ただし，$\mathcal{U}$はユーザ端末全体の集合を表す．また，各端末はインターネット上の公開データ$\mathcal{D}_\mathrm{p}$にアクセス可能とする．ここで，$\mathcal{D}_\mathrm{p}$に教師ラベルは付与されていないものとする．ユーザ端末は公開データ$\mathcal{D}_\mathrm{p}$に対して，学習途中のモデル

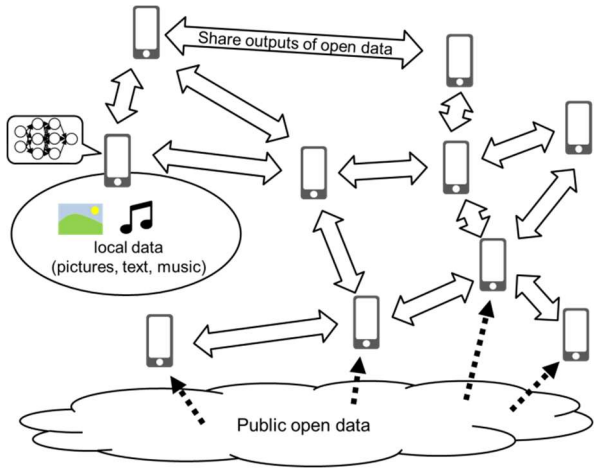

図 1. サーバレス Federated Learning

Distributed Optimization for Serverless Federated Learning
†Akihito TAYA, Yoshito TOBE / Aoyama Gakuin University
‡Takayuki NISHIO, Masahiro MORIKURA, Koji YAMAMOTO/ Kyoto University

の出力を計算し，その結果をネットワーク上の隣接端末と共有する．

この時，機械学習は以下の形式の最適化問題として定義される．

$$\underset{f}{\text{minimize}}\ L(f) := \sum_{(x,y)\in\mathcal{D}} l\big(y, f(x)\big) \qquad (1)$$

ただし，$L(f)$は損失関数であり，一般的に平均二乗誤差や交差エントロピーなどが利用される．協調学習では，ユーザ端末は$\mathcal{D}$の一部である$\mathcal{D}_i$のみを利用して学習モデルf_iを独立に更新するが，最終的には全ユーザで同一かつ$L(f)$を最小化する$f^\star$に収束することを目指す．

深層学習では学習モデルfをパラメータwにより表現し，(1)をwについての最適化問題として定義するのが一般的である．もし，$L\big(f(w)\big)$がwについて凸関数であれば，(1)は分散合意最適化[3]によって最適化できるが，深層学習では学習モデル$f(x;w)$が非凸関数であるため，wを最適化することはできない．しかし，損失関数に平均二乗誤差や交差エントロピーを利用する場合，$L(f)$がfに対して凸汎関数になるため，分散合意最適化を関数空間上で実行できれば，すべてのf_iが最適解$f^\star$に収束することが期待される．提案手法では[3]における変数の平均化の操作を学習モデル同士の距離を減少させることで，関数空間上での分散合意最適化を実現する．

提案手法の擬似コードをアルゴリズム 1 に示す．3 行目で各ユーザ端末は確率的勾配法を使って自身の学習モデルf_iを更新する．ここでは，関数空間でのアルゴリズムであることを強調するためにf_iを Fréchet 微分D_{f_i}によって更新する表記としているが，実装上は近似として，パラメータw_iについての勾配を計算しw_iを更新する．[1]ではパラメータw_iをサーバにアップロードするが，提案アルゴリズムでは co-distillation を

アルゴリズム 1．サーバレス Federated Learning

1: **while** not converged **do**
2: **for** $i \in \mathcal{U}$ **do**
3: $f_i \leftarrow f_i - \eta D_{f_i} \sum_{(x,y)\in\mathcal{D}_i} l\big(y, f_i(x)\big)$
4: $y_i(x) \leftarrow f_i(x),\ \forall x \in \mathcal{D}_\mathrm{p}$
5: **end for**
6: share y_i with neighbors
7: **for** $i \in \mathcal{U}$ **do**
8: $\overline{y(x)} \leftarrow \frac{1}{|\mathcal{N}_i|}\sum_{j\in\mathcal{N}_i} y_j(x),\ \forall x \in \mathcal{D}_\mathrm{p}$
9: $f_i \leftarrow f_i - \eta_\mathrm{c} D_{f_i} \sum_{x\in\mathcal{D}_\mathrm{p}} \big\|\overline{y(x)} - f_i(x)\big\|^2$
10: **end for**
11: **end while**

η, η_c：学習係数，$\mathcal{N}_i$：ユーザiの隣接ユーザ

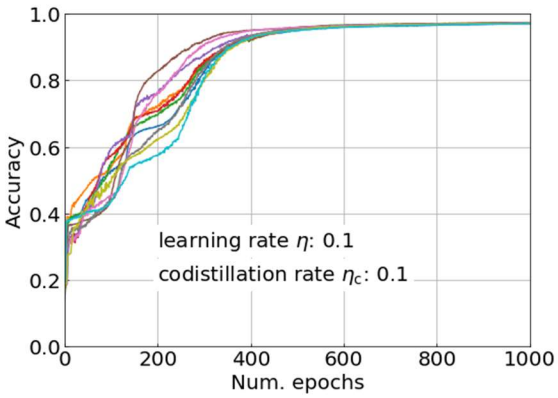

図 2．MNIST の学習結果

隣接ユーザ間で行い（4, 6, 8, 9 行目），ユーザ間で同一の学習モデルに収束するようにしている．この更新式は関数空間上での学習モデル同士の距離を減少させている．なお，このアルゴリズムの収束とは関数としての収束，$f_i(x) = f_j(x), \forall x \in \mathcal{D}, \forall i, j \in \mathcal{U}$を目指し，パラメータ$w$を収束させるものではない．

4．シミュレーション評価

提案する分散学習アルゴリズムを MNIST で評価した．10 台のユーザ端末がリング状のネットワークを形成し，隣接する 2 台のみと公開データに対する出力を共有させた．ユーザごとに所持するデータの偏りがあることを想定し，各端末は 2 種類のラベルでそれぞれ 500 組のデータを学習に利用し，公開データとしては 10000 枚の画像を学習に利用した．図 2 に評価結果を示す．10 種類の手書き文字認識の正解率の学習経過を端末ごとに示している．学習が進むにつれて，互いの学習モデルが同一のモデルに収束し，分類の正解率が収束している．収束後の正解率のユーザ間平均は約 97%であった．また，それぞれの端末は 2 種類のラベルのデータしか利用できないにもかかわらず，隣接ユーザと公開データに対する出力を共有するだけで 10 種類のラベルすべての分類に成功していることがわかる．

謝辞

本研究は立石財団及び AOYAMA VISION「AI 研究拠点形成プロジェクト」の助成を受けたものです．

参考文献

[1] McMahan, H. B., et al.: Communication-Efficient Learning of Deep Networks from Decentralized Data, Proc. of AISTATS, 2017.

[2] Anil, R., et al.: Large Scale Distributed Neural Network Training through Online Distillation, Proc. Of ICLR, Apr, 2018.

[3] Nedic, A., et al.: Constrained Consensus and Optimization in Multi-Agent Networks, IEEE Trans. Autom. Control, Vol.55, No.4, pp.922-938 (2010).