Computer Security Symposium 2019
21 - 24 October 2019

個人差のあるデータによる協調機械学習における精度向上手法の提案とプライバシー保護に対する一考察

香川 椋平[1,a)]　西村 拓真[1]　松永 昌浩[1]

概要：宅内設置センサ等から得られる，生活に関するデータを想定し，プライバシーに配慮しつつ機械学習を行うことを目指す．そこで，クライアントデバイスから生データを収集することなく，機械学習モデルの更新に必要な情報のみを収集する協調機械学習に着目した．ただし，生活に関するデータは個人ごとの生活パターンによって大きく偏ったデータが蓄積されると考えられる．このような状況を想定して協調機械学習を行ったところ，精度の大幅な低下がみられた．本稿では，生活に関するデータを学習データと想定した際の精度低下に対し，それを解決する手法を提案する．さらに，偏りのあるデータを用いた際の協調機械学習のプライバシー保護について考察を行う．

キーワード：協調機械学習, 非独立同分布, プライバシー

Accuracy Improvement Method for Collaborative Machine Learning with Individually Different Data and Consideration on Privacy

Ryohei Kagawa[1,a)]　Takuma Nishimura[1]　Masahiro Matsunaga[1]

Abstract: We aim to perform machine learning with considering privacy, when using life activity data from home sensors . Therefore, we focused on collaborative machine learning. However, the data is biased because it depends on lifestyle patterns,and in such a case the accuracy of the collaborative machine learning decreases. In this paper, we propose an accuracy inprovement method for collaborative machine larning using life activity data. Furthermore, we consider the privacy protection of collaborative machine learning with biased data.

Keywords: Collaborative Machine Learning,non-Independent and Identically Distributed,privacy

1. はじめに

IoT 機器の発達や，スマートホーム機器の普及によって，様々なセンサが利用可能となってきている．これらのデータを用いることで，人の行動を推定する試みが行われている [1]．健康状態や生活状態をセンシングされたデータから推定することで，見守りサービス等，様々なサービスの実現が可能となる．生活に関するデータを用いてサービスを行う場合， 複数のデータを組み合わせることにより，推定可能な生活行動の増加や推定精度の向上が期待される．また，将来の予測や，予測精度をより向上させるために，サービスを提供する個人のデータのみではなく，複数の人のデータを用いる必要がある．例えば，病気の予測をする際には，その人のデータしか利用できないと，学習データにはその人が過去にかかったことのある病気のデータしか存在しないため，初めてかかる病気の予測ができない恐れがある．このような場合には，複数の人のデータを参照し，その病気になる要因を分析することで，より高精度な予測が可能となることが期待される．

本稿では，生活に関するデータにおいて健康状態や生活行動を推定するモデルを機械学習を用いて構築することを想定する．

[1] セコム株式会社 IS 研究所
Intelligent Systems Laboratory, SECOM CO.,LTD.
[a)] ryo-kagawa@secom.co.jp

© 2019 Information Processing Society of Japan

しかし，生活に関するデータは有用であると同時に，センシティブな情報であるため，サービス提供事業者が用途に応じて必要なデータを収集し，適切に扱う必要がある．

例えば，機械学習モデルの精度向上目的に利用する場合には，個人の詳細情報を利用する必要はなく，データ全体の傾向を知るための情報を収集することが望ましい．データ全体の傾向を知るための情報のみを収集する際に用いられる技術の一例として，差分プライバシー [2], [3] を満たすノイズ付加が用いられることがある．しかし，多次元データに対するノイズ付加には課題がある．宅内設置センサ等のデータに，ノイズを付加して収集する場合には，あらかじめ必要な情報を定めて，要約したデータに対してノイズを付加することが望ましい．しかし，機械学習によって，生活データを学習する場合には，必要な情報をあらかじめ定めることが困難である．そこで，プライバシーの尊重と機械学習モデルの精度向上の両立をめざし，協調機械学習に着目した．協調機械学習とは，クライアントデバイスにて学習を行い，モデルのパラメータをサーバで収集することによって，複数のクライアントのデータを学習した共通モデルを学習する手法である．この手法によって，機械学習を行う際にクライアントから生データを収集せずに，機械学習モデルの更新に必要な情報のみを収集し，プライバシーに配慮しつつ共通モデルを学習することが可能となる．

本稿では，この協調機械学習手法を生活に関するデータに対して適用することを目指す．その場合に想定される，クライアントデータセットを意図的に偏らせて協調機械学習の既存手法である Federated Averaging[6] を行ったところ，学習の効率が低下する問題が見られた．そのような状況においての協調機械学習の精度向上手法として SOFA:Similarity Oriented Federated Averaging を提案した．SOFA によって，学習に参加するデバイスを大幅に増やすことなく，多様性を確保したクライアント選択が可能となる．その結果，クライアントデータセットに偏りを生じさせた場合において分類モデルの精度向上が確認できた．さらに，3 節の検証において学習が困難であった VAE に関しても，同じラベルのデータが潜在空間で近くなるように学習することができた．

本稿の構成は以下となる．まず 2 節では，本研究において着目した協調機械学習について説明し，既存研究で利用されている手法について説明する．3 節では生活に関するデータを想定し，各クライアントデータセットの分布を偏らせた状況下での Federated Averaging の検証を行い，課題の洗い出しを行う．4 節では，課題に対する解決手法としてクライアントから送信される学習モデルの類似度に着目した SOFA:Similarity Oriented Federated Averaging を提案する．5 節において SOFA について実験と評価を行い，協調機械学習をプライバシー保護に適用する際の課題について考察する．最後にまとめと今後の展望について述べる．

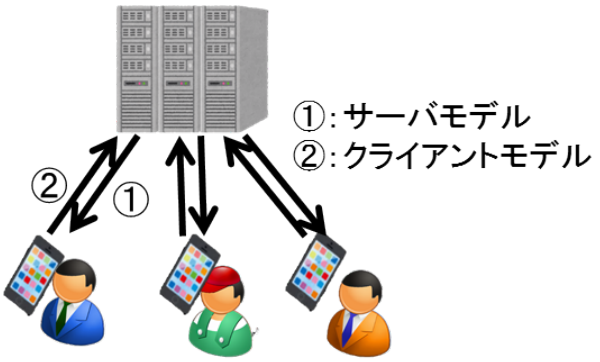

図 1　協調機械学習
Fig. 1　Collaborative machine learning.

2. 関連研究

本節では，協調機械学習手法について説明を行う．一般的な機械学習において，複数のクライアントのデータを用いて学習した共通モデルを構築するためには，サーバに，クライアントからデータ収集し，学習に参加するすべてのクライアントのデータを保持したグローバルデータセットを構築する．そのグローバルデータセットを用いて共通モデルを学習する．

対して，協調機械学習では，すべてのクライアントのデータを収集したグローバルデータセットは構築せず，クライアントがそれぞれ保持しているクライアントデータセットを用いて学習を行う．**図 1** のように，サーバが保持している学習中のモデルをクライアントに送信し，クライアント側のデバイスにて学習を行いモデルを更新し，サーバに送り返す．学習に使用されるクライアントは，サーバの更新のたびにサーバ選択割合 C によって指定した台数をランダムに選択する．これを繰り返すことによって，学習データを収集せずとも共通モデルの学習が可能となる．

複数マシンの学習結果を統合することで学習を行う手法としては，深層学習の分野においては隠れ層の増加に伴い増加していく学習時間を減らすために，複数の GPU マシンなどを用いて高速化を図るという文脈での分散学習が行われている [4], [5]．しかし，この場合の学習に用いるデータは，すべてのデータを保持したグローバルデータセットを共有又は，シャッフルして分割したデータセットを学習に用いる．その点において，協調機械学習とは異なっている．

この協調機械学習を適用した例として，Google が発表した Federated Averaging[6], [7], [8] と呼ばれる手法がある．

この手法の流れを Algorithm1 に示す．

Federated Averaging は現在スマートフォンのキーボードアプリ Gboard[9] でテストされており，個人のスマートフォンに蓄積された情報をもとに共通の予測変換モデルの

Algorithm 1 Federated Averaging

function SERVER: 共通モデルの学習
  $K \leftarrow$ 全クライアント数
  $C \leftarrow$ 各ラウンドにおけるクライアント選択割合
  $w \leftarrow$ 初期モデルパラメータ
  for round t=1,2$\cdots$ **do**
    $m \leftarrow max(C * K, 1)$
    $S_t \leftarrow m$ クライアントをランダムに選択
    for each client $k \in S_t$ in parallel **do**
      $w_{t+1}^k \leftarrow$ ClientUpdate (k, w_t)
    end for
    $w_{t+1} \leftarrow \sum_{k=1}^{K} \frac{n_k}{n} w_{t+1}^k$
  end for
end function

function CLIENTUPDATE(k, w) : クライアント k によるモデルの学習
  for epoch i=1,2$\cdots$ **do**
    $w \leftarrow$ モデルパラメータ更新
  end for
  return w
end function

学習が行われている．入力した文章の内容や入力時間等を事業者が知ることなく，予測の為の情報のみを事業者が収集することが可能となる．

3. 偏りのあるデータに対する先行研究の検証

本稿では，生活に関するデータを協調機械学習で分析することに着目する．生活パターンは頻繁に変わるものではないと考えられる．そのため，各クライアントデータセットは，様々な人のデータを集めたグローバルデータセットの分布とは異なり，各自の生活パターンに従った，偏ったデータのみで占められていると考える．例えば，朝型の生活や夜型の生活，自炊派や外食派，また，仕事・季節等の要因によってデータが偏る場合がある．

つまり，協調機械学習によって生活に関わるデータを学習することを想定した場合には，各クライアントデータセットの分布が全デバイスのデータを集めた分布とは異なるものになると考えられる．

本節では，偏りがあるクライアントデータセットによる影響について示す．生活に関わるデータセットは，データ数が少ない場合が多く，本実験に用いることができなかったため，意図的に偏りを生じさせた手書き数字データセット MNIST を用いて，Federated Averaging によって協調機械学習を行う．3.2 項では，先行研究 [6] でも行われていた分類モデルの学習の結果を示す．さらに，3.3 項では，先行研究では行われていない，教師なし学習である Variational AutoEncoder(VAE)[10] に関して，協調機械学習を行うことができるのかを検証する．

3.1 検証条件

クライアントデータセットを意図的に偏らせた場合と偏らせない場合の Federated Averaging の動作を検証する．本研究では説明が無い限り [6] を参考にクライアントの選択割合は $C = 0.1$ で固定し，ラウンドごとの共通モデルに対して精度評価を行う．さらに，クライアントごとにデータに偏りがないデータ分割を「IID(independent and identically distributed)」，データに偏りがあるデータ分割を「non-IID」と呼ぶ．

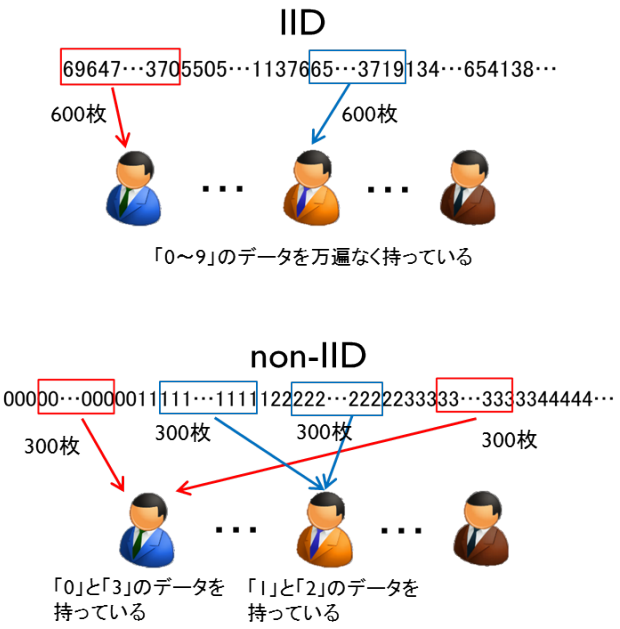

図 2　IID と non-IID のデータセット
Fig. 2　IID and non-IID dataset.

評価実験では，**図 2** に示すように，学習データ 60000 枚をクライアント 100 人に 600 枚ずつランダムに割り当てた分割を IID，0〜9 の数字のうち 300 枚ずつ 2 種類を持つように割り当てた分割を non-IID とする．

3.2 分類モデルの結果

入力が MNIST 画像 で，隠れ層 2 つを持つニューラルネットワークを用いて分類モデルを学習させる．0〜9 の 10 クラス分類であるため，出力層は 10 次元である．

IID と non-IID のデータセットを用いた場合の比較を**図 3** に示す．non-IID の場合には，学習の精度が下がることが読み取れる．

3.3 教師なし学習 VAE の結果

本稿では新たに，データにラベルを付与することが難しく教師無し学習での運用を行う場合も想定して，教師無し学習の一例として VAE を学習する実験を行った．

VAE は，高次元である入力データの特徴を，低次元の潜在空間で表現できるように学習を行う．教師データは利用せず、入力と出力が同じになるように学習することで，入力データの特徴を保持したまま潜在変数に圧縮する．本実

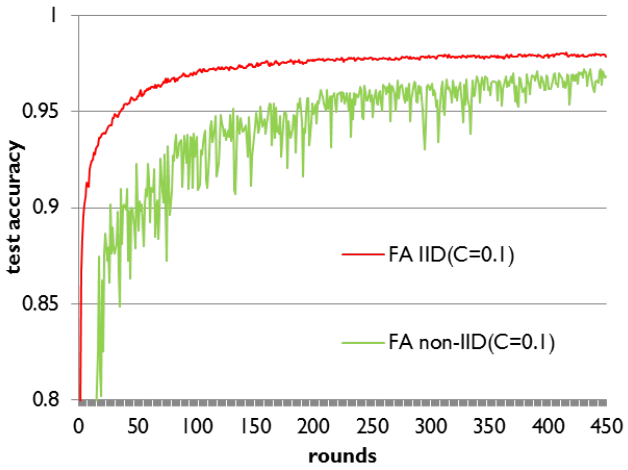

図 3　IID と non-IID のデータセットを用いた場合の精度比較

Fig. 3　Test set accuracy of Federated Averaging on IID and non-IID data.

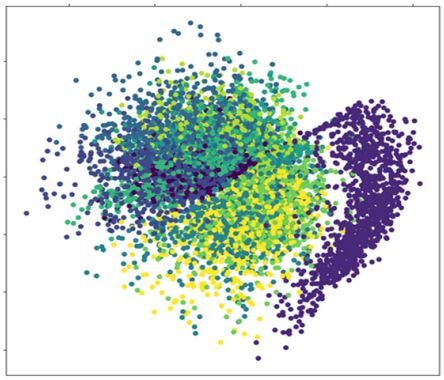

図 5　non-IID データにおける Federated Averaging による VAE の潜在空間

Fig. 5　Latent space by VAE with Federated Averaging on non-IID data.

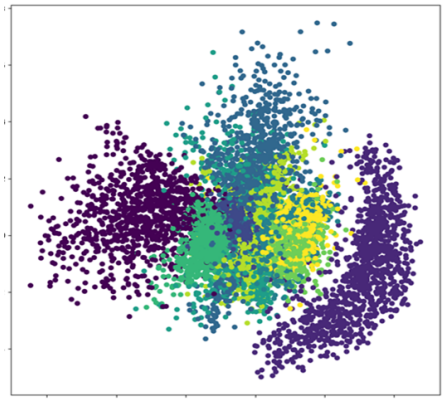

図 4　IID データにおける Federated Averaging による VAE の潜在空間

Fig. 4　Latent space by VAE with Federated Averaging on IID data.

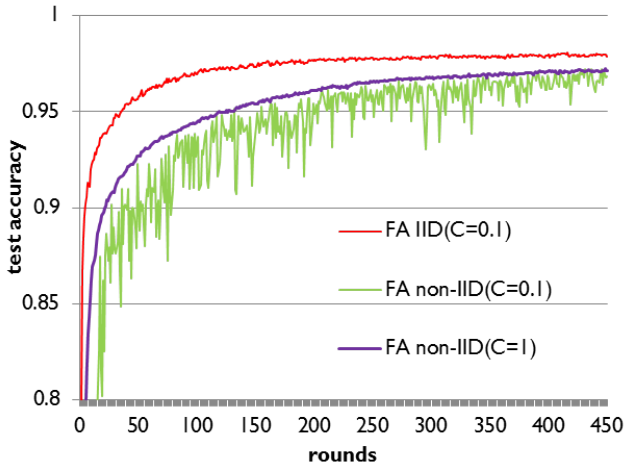

図 6　クライアントの選択割合を上げる影響

Fig. 6　The influence of increasing client selection rate.

験では入力と出力が 784 次元，潜在空間を 2 次元として，Federated Averaging を行い，入力データの特徴を保持したまま潜在空間へ圧縮されるかを確認する．

学習後の潜在空間を**図 4**，**図 5** に示す．この図にはラベルの数字ごとに色を付けている．

VAE のエンコーダ部分では元データを復元するのに十分な情報を残して低次元の潜在変数空間へ圧縮するように学習が行われるため，同じクラスのデータの潜在変数空間上への写像は距離的に近くなりやすい．しかし，non-IID のデータを用いて Federated Averaging を行った場合には，図 5 のように様々なクラスのデータが入り混じり，入力データの特徴を保持できていないため，学習がうまくいかないことが確認できた．

3.4　評価

3.2,3.3 項にて示したように，non-IID の場合に学習がうまくいかない理由について考察を行う．協調機械学習では，各クライアントにおいてそれぞれが保持するクライアントデータセットを用いて学習を行っている．本稿では，クライアントデータセットは偏りが生じていることを想定している．そのため，各ラウンドにて選択されたクライアントが保持するデータの分布が，全ユーザの分布と異なることが，学習の精度低下の原因であると考えた．

Federated Averaging におけるデバイスの選択方法がランダムである為，デバイスの選択数を増やすことで，全クライアントデータの分布に近づけることかできる．クライアントの選択割合を $C = 1$ にした場合の学習結果を**図 6** に示す．3.2 項にて行った $C = 0.1$ 時の IID と non-IID の検証結果に加えて，non-IID $C = 1$ で学習を行った結果を示している．

non-IID におけるクライアント選択割合を変化させた 2

パターンを比べると，クライアントの選択割合を上げた方が test accuracy の上がり方が早くなり IID での学習に近づくことが確認できる．

そのため，ラウンドに参加したクライアントの学習データが多様なものであることが精度向上につながるのではないかと考えた．しかし，各クライアントのデータは送信されないため，サーバ側では学習データを参照することができない．ラウンドにおけるクライアントはクライアント選択割合 C に応じてランダムに選択される．そのため，多くのクライアントからデータを受信すれば，参加したクライアントの学習データは，グローバルデータセットの分布に近づく．ただし個人デバイスを想定した場合には，デバイスが必ず通信可能な状況下にあるとは限らなかったりサーバが同時に通信できる台数に限りがあったりする．さらに，学習に参加するデバイスとサーバとの間には，学習済モデルのパラメータの通信が生じる為，通信コストも増大する．よって，学習参加デバイスを大幅に増やすことは現実的ではないと考えている．そのため，これ以降は再び $C = 0.1$ の条件下で，クライアント間でデータ分布に偏りがある場合の学習の改良を目指す．

4. 提案手法

本節ではクライアント間で偏ったデータ分布を持つ場合の学習における精度低下に対する改善手法を述べる．さらに，提案したアルゴリズムについて 3 節と同様の実験を行うことで性能評価を行う．

4.1 提案手法のアルゴリズム

前述のとおり，多様な学習データをラウンドの学習に加えることで精度向上につながると考えている．しかし，協調機械学習の設定上，クライアントに保存されているデータセットをサーバが確認することはできない．そこで，本稿では，クライアントから送信される学習モデルの類似度に着目した SOFA:Similarity Oriented Federated Averaging を提案する．

クライアントから送信されるパラメータは，学習データを参照し更新される．よって，パラメータが類似するクライアント同士のデータセットも，類似していると考えられる．そこで，類似しているクライアントを同ラウンド内では選択しないことでクライアントの多様性を確保する提案手法のアルゴリズムについて Algorithm2 に示す．

SOFA では，クライアントの選択時に高類似度なペアを許さないように選択する処理を加えている．高類似度なペアを把握するために，サーバ側でパラメータを受け取った際にクライアント間でパラメータのコサイン類似度を計算する．一度類似度ペアに選ばれると，それ以降のラウンドでは同時に学習に参加させないようにする．これにより生データの受信は行わずにクライアントの選択方法を変えることが可能になり，多様な学習データを持つクライアントを選択することができる．

Algorithm 2 Similarity Oriented Federated Averaging

function SERVER: 共通モデルの学習
  $K \leftarrow$ 全クライアント数
  $C \leftarrow$ 各ラウンドにおけるクライアント選択割合
  $w \leftarrow$ モデルパラメータ初期化
  $P_n \leftarrow$ 類似ペアリスト
  $T \leftarrow$ 類似ペア登録閾値
  for round t=1,2$\cdots$ **do**
    $m \leftarrow max(C * K, 1)$
    $S_t \leftarrow m$ クライアントをランダムに選択
    ただし S_t 内に P_n に登録されている類似ペアは存在しない
    for each client $k \in S_t$ in parallel **do**
      $w_{t+1}^k \leftarrow$ ClientUpdate (k, w_t)
    end for
    for client $k_1, k_2 \in S_t$ **do**
      if $\cos(\Delta w_{t+1}^{k_1}, \Delta w_{t+1}^{k_2}) > T$ **then**
        クライアントペア（k_1, k_2）を P_n に登録
      end if
    end for
    $w_{t+1} \leftarrow \sum_{k=1}^{K} \frac{n_k}{n} w_{t+1}^k$
  end for
end function

function CLIENTUPDATE(k, w) : クライアント k によるモデルの学習
  for epoch i=1,2$\cdots$ **do**
    $w \leftarrow$ モデルパラメータ更新
  end for
  return w
end function

5. 実験および評価

本節では，提案手法に対して 3 節と同様の実験を行うことで性能評価を行う．IID 及び non-IID の場合における分類モデルの精度評価及び，non-IID の場合における VAE に適用した結果を示す．

5.1 分類モデルの精度評価

IID データでの学習の結果を**図 7** に示す．使用したモデル及び条件は 3 節にて用いたものと同様である．

IID データでの学習の場合，すべてのユーザがある程度類似しているが，高類似度なペアが存在しないために，実質的な処理が変化しない．そのため，図 7 に見られるように Federated Averaging と SOFA の精度に差は現れず，グラフが重なっている．そのため提案手法が，偏りのないデータでの協調機械学習に対しても悪影響を与えないと考えられる．

続いて，non-IID データでの学習の結果を**図 8** に示す．non-IID データでの学習の場合，提案手法の方が test accuracy の上がり方が安定して早くなった．こちらの場合では実際に高類似度なペアも出現しており，提案手法が上手

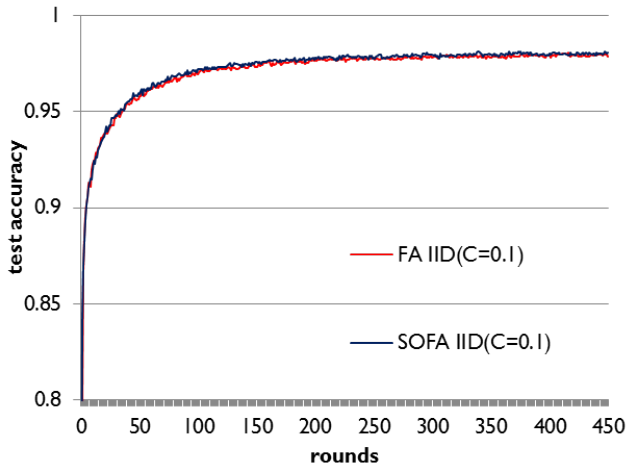

図 7　IID データにおける Federated Averaging と SOFA の精度比較

Fig. 7　Test set accuracy for Federated Averaging and SOFA on IID data

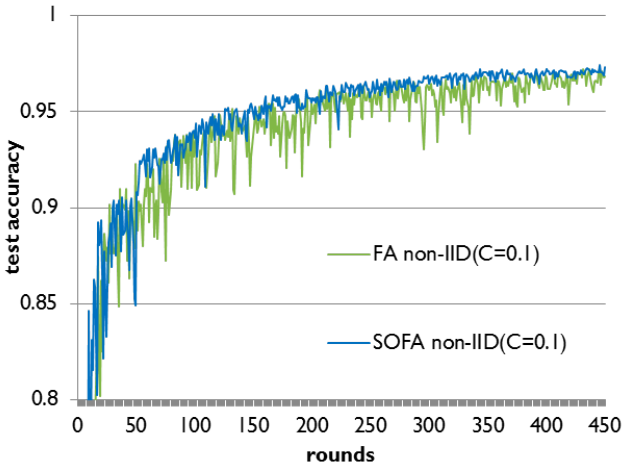

図 8　non-IID データにおける Federated Averaging と SOFA の精度比較

Fig. 8　Test set accuracy for Federated Averaging and SOFA on non-IID data

く働いていると考えられる．

5.2　VAE への適用

3 節の検証において学習が困難であった，non-IID データにおける VAE の学習に提案手法を適用する．使用したモデル及び条件は 3 節にて用いたものと同様である．結果を図 **9** に示す．

同じクラスのデータが潜在空間で近くなっていることから，提案手法によって VAE の学習性能が向上したと考えられる．

6.　議論

6.1　プライバシー保護に関する考察

non-IID データによる Federated Averaging における，

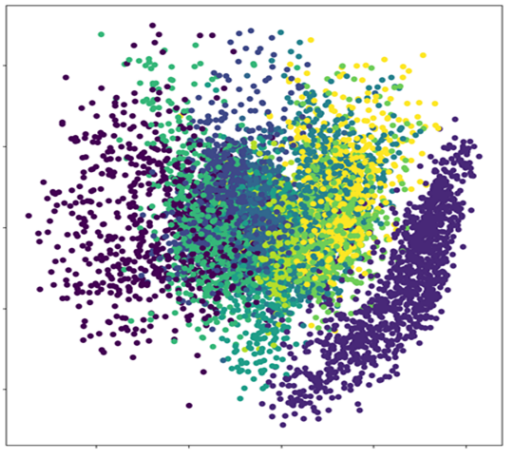

図 9　non-IID データにおける提案手法を用いて学習した VAE による潜在空間

Fig. 9　Latent space by VAE with SOFA on non-IID data

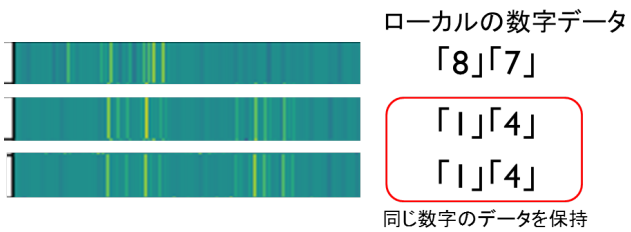

図 10　non-IID による Federated Averaging にてクライアントから送信されたモデルの更新差分

Fig. 10　Client model parameters of Federated Averaging on Non-IID data.

あるラウンドにクライアントによって更新されたパラメータの差分を可視化したものを図 **10** に示す．これは，モデルパラメータのうち出力層に繋がるパラメータの一部を，横方向に可視化したものである．

同じ数字種を持っているクライアントが送信するパラメータが類似しているため，協調機械学習で収集するものは，クライアント毎のモデルパラメータのみあるが，生データを送信しなくとも，それを学習したモデルパラメータに学習データの影響が表れていることがわかる．

そのため，あるクライアントの情報を事前知識として持っている場合には，類似したモデルを送信した，他のクライアントに関してもデータの傾向を把握することが可能となる．

生活に関わるデータは生活パターンによってデータが偏ると考えている．そのため，データの傾向を把握することで，データに偏りが生じた要因である生活パターンも推定することができる．例えば，電力を朝によく利用する傾向がある場合には，朝に起床して家事を行う生活をしているという推測をすることができる．

生活パターンがプライバシー情報であるかは，データが利用される目的や提供先，提供者の感情によって変わると考えている．仮に，推定される生活パターン自体をプライバシー情報とした場合には，協調機械学習による保護のみでは不足であると考えている．そのため，これらの情報までも隠したい場合には，送信するパラメータの内，一部のみをランダムサンプリングすることや差分プライバシー [2] を満たしたノイズ付与など，さらなるプライバシー保護手法を講じる必要がある．

6.2 今後の予定

本節において，意図的に偏りを生じさせたデータセットを用いて，提案手法の検証を行った．その結果，パラメータに含まれる情報を利用することで，non-IID データによる協調機械学習において分類モデルの精度が向上することが確認できた．しかし，パラメータに含まれる情報もセンシティブなものである場合には，パラメータに対してさらなるプライバシー保護手法が必要となり，モデルパラメータからクライアント間の類似度を算出する本手法の適用が難しくなる．この場合においても適用可能な手法を検討したいと考えている．さらに，この実験は手書き数字データセット MNIST を用いて行ったが，実際に生活データを用いて検証を行いたいと考えている．しかしながら，先に述べたように，生活に関わるデータセットは，データ数が少ない場合が多く，協調機械学習の評価に用いるためには，より多い件数のデータが利用できるデータセットの登場が望まれる．

7. おわりに

本稿では，生活データの様な偏ったデータセットを想定して，分類モデル及び教師なし学習である VAE に対して Federated Averaging の性能評価を行った．その結果，偏ったデータセットを利用した際には精度が低下するという問題が確認できた．さらに，そのような状況においての協調機械学習の精度向上手法として SOFA:Similarity Oriented Federated Averaging を提案した．SOFA は，パラメータのみからでも得られる情報を利用することで，学習に参加するデバイスを大幅に増やすことなく，多様性を確保したクライアント選択が可能となる．その結果，クライアントデータセットに偏りがある場合の協調機械学習において分類モデルの精度向上が確認できた．さらに，3 節の検証において学習が困難であった VAE に関しても，同じラベルのデータが潜在空間で近くなるように学習することができた．しかし，6.1 項にて述べたように，パラメータにはクライアントデータセットの偏りが表れるため，その情報もプライバシー情報とした場合には，SOFA が適用できなくなる．この場合においても，適用できる手法について検討を進めていきたいと考えている．

参考文献

[1] Wang, J., Chen, Y., Hao, S., Peng, X. and Hu, L.: Deep learning for sensor-based activity recognition: A survey, *Pattern Recognition Letters*, Vol. 119, pp. 3–11 (2019).

[2] Dwork, C.: Differential privacy, *Encyclopedia of Cryptography and Security*, pp. 338–340 (2011).

[3] Dwork, C., Roth, A. et al.: The algorithmic foundations of differential privacy, *Foundations and Trends® in Theoretical Computer Science*, Vol. 9, No. 3–4, pp. 211–407 (2014).

[4] Dean, J., Corrado, G., Monga, R., Chen, K., Devin, M., Mao, M., Senior, A., Tucker, P., Yang, K., Le, Q. V. et al.: Large scale distributed deep networks, *Advances in neural information processing systems*, pp. 1223–1231 (2012).

[5] Das, D., Avancha, S., Mudigere, D., Vaidynathan, K., Sridharan, S., Kalamkar, D., Kaul, B. and Dubey, P.: Distributed deep learning using synchronous stochastic gradient descent, *arXiv preprint arXiv:1602.06709* (2016).

[6] McMahan, H. B., Moore, E., Ramage, D., Hampson, S. and othersmcmahan2016communication: Communication-efficient learning of deep networks from decentralized data, *arXiv preprint arXiv:1602.05629* (2016).

[7] Konečnỳ, J., McMahan, H. B., Yu, F. X., Richtárik, P., Suresh, A. T. and Bacon, D.: Federated learning: Strategies for improving communication efficiency, *arXiv preprint arXiv:1610.05492* (2016).

[8] McMahan, B. and Daniel Ramage, R. S.: Google AI Blog : Federated Learning: Collaborative Machine Learning without Centralized Training Data (2016). `https://research.googleblog.com/2017/04/federated-learning-collaborative.html`.

[9] Google: Gboard, now available for Android (2016). `https://blog.google/products/search/gboard-now-on-android/`.

[10] Kingma, D. P. and Welling, M.: Auto-encoding variational bayes, *arXiv preprint arXiv:1312.6114* (2013).