組織間水平連合学習による社会実装

小澤, 誠一

(Citation)
オペレーションズ・リサーチ,69(5):241-246

(Issue Date)
2024-05-01

(Resource Type)
journal article

(Version)
Version of Record

(URL)
https://hdl.handle.net/20.500.14094/0100496779

組織間水平連合学習による社会実装

小澤　誠一

組織の機微なデータを他組織と共有しなくても，高度な AI を協調して構築できる連合学習は，プライバシー保護を重視する社会実装に不可欠な技術である．連合学習の有望な応用として，特殊詐欺の被害口座やマネーロンダリングを行う不正口座を検知する社会課題がある．銀行の顧客口座情報や取引履歴は，いわゆる要配慮個人情報に該当し，金融機関が組織間を超えてデータ提供することは極めて難しい．しかし，メガバンク以外の銀行が十分な不正口座のデータを確保することは簡単ではない．本解説では，このような背景から行っている連合学習の銀行不正送金検知への応用の取り組みについて紹介し，連合学習の今後の課題について述べる．

キーワード：プライバシー保護，連合学習，AI の社会実装，銀行不正送金検知

1. はじめに

内閣府の第 5 期科学技術基本計画で提言された Society 5.0 の実現に向けて，その前提となるデジタル化については一定の進展がみられたが，諸外国のようなデータ連携・活用による新たなビジネスモデル創出や社会課題の解決は十分に進んでいるとはいえない．国内において，IT プラットフォーマー企業など一部の企業を除いて，データホルダー間のデータ連携・活用が進んでない理由には，信頼性とセキュリティが担保されたデータ流通環境の整備やプライバシーに配慮した AI・データ分析基盤が整っていない技術的要因がある．

一方，海外では，2010 年代の GAFA に代表される IT プラットフォーマーによる国際的な情報独占によりビッグデータ解析と AI（特に深層学習）の技術が飛躍的に進展し，データから巨大な富を生む仕組みが確立され，優秀な人材が AI・データサイエンス分野に集まるようになった．しかし，昨今の個人情報に対する意識の高まりやサイバー攻撃の高度化に伴って，パーソナルデータを一極集中で収集し，解析するこれまでのデータ解析スタイルが大きなリスクをもつことになり，IT プラットフォーマーでも差分プライバシーをはじめとしたプライバシー保護技術を導入し，さらにデータを個人（エッジ）や組織の外に出さないで AI を学習・利用する連合学習 (Federated Learning) が注目されるようになり，医療や金融などの分野への社会実装が急速に進んでいる．このような状況で，実現すべき新たな社会の具現化目標として，サイバー空間とフィジカル空間とのダイナミックな好循環を生み出し，安心してデータや AI を活用した新たな価値創出を行う Society 5.0 が掲げられ，第 6 期科学技術・イノベーション基本計画が令和 3 (2021) 年 3 月 26 日に閣議決定された．これを受け，わが国自身の社会課題の克服や産業競争力の向上に向けた AI に関する総合的な政策パッケージとして「AI 戦略」が取りまとめられ，令和 4 (2022) 年 4 月に，その最新版である「AI 戦略 2022」が策定された．その戦略目標 2「産業構想力」を達成する重要技術の一つとして，「プライバシーや機密情報を保護しながら学習可能な連合学習などの技術」が挙げられている．

本解説記事では，プライバシーを保護したうえで安全にデータを解析するための「プライバシー保護データ解析技術」の一つとして，複数組織がもつデータを互いに中身を知らずに目的の計算が可能な複数組織間連合学習を取り上げ，その応用事例を紹介する．

2. 組織間プライバシー保護連合学習

2.1 プライバシー保護データ解析

プライバシー保護データ解析とは，収集されたパーソナルデータが何らかの理由で漏えいして，攻撃者（興味本位でのぞき見する者も含む）が入手したとしても，そこから個人を特定できないようにする技術である．これには，匿名化，差分プライバシー [1, 2]，秘密計算，準同型暗号 [3] を用いる方法などがあり，それぞれを単独または組み合わせて使うことにより，データの秘匿化を実現する．ただし，一般に匿名性と有用性はトレードオフの関係にあり，匿名性を高めると，データ漏えいによる損失は小さくできるが，そのデータから得られる情報量は少なくなる [4]．これに対し，連合学習は匿名性と有用性を両立させる手法として期待されている．

連合学習 [5] は，2016 年に Google によって発表さ

おざわ　せいいち
神戸大学数理・データサイエンスセンター/大学院工学研究科
〒 657–8501　兵庫県神戸市灘区六甲台町 1–1
ozawasei@kobe-u.ac.jp

れた学習フレームワークである．複数のデータ所有者が互いのデータを秘匿したまま協調し，一つの機械学習モデルを構築する枠組みである．Yang ら [6] は，学習データの分担方法によって「水平連合学習」と「垂直連合学習」に分類した．Google が目指す連合学習では，まず機械学習モデルがクラウドからユーザのスマートフォン端末にダウンロードされる．そして，端末上でユーザのデータを使って訓練し，モデルの更新情報のみを暗号化してクラウドに送り返し，他ユーザから送られてくる更新情報と平均化して，端末内の機械学習が更新される．

類似のアイデアにより，複数組織間がもつデータで一つの深層学習モデルを協調して学習する方式を Phong ら [7] が提案している．この機械学習モデルは DeepProtect と呼ばれ，複数の組織内で学習した深層学習モデルの勾配情報を加法準同型暗号で暗号化して中央サーバーに集め，中央サーバーで暗号化したまま勾配情報を加算する．加算された勾配情報は各組織に戻されて，それぞれがもつ秘密鍵を使って復号し，深層学習モデルの更新を行う．Google 連合学習との違いは，更新情報が中央サーバで復号されないで加算され，機械学習モデルのアップデートはクライアント側で行う点である．Google 連合学習や DeepProtect は，互いに不足する情報を補って学習することに重点を置いており，個々の組織だけではカバーできない希少事例を直接共有せず，複数の組織が連携して学習する仕組みを提供している．この点で，これらは社会実装向きの技術として期待されている．

2.2 組織間水平連合学習の定式化

連合学習のプライバシー保護は，組織間で直接データを共有する代わりに，各組織に課された学習誤差の最小化問題を他組織と共有することで達成していると考えられる．つまり，同じ特徴量をもつ N 組織がデータセット $\mathcal{D}_i$ $(i = 1, \ldots, N)$ を有しており，共通の AI モデル $\mathcal{M}(\theta)$（グローバルモデルという）を協力して学習するとき，全組織の総誤差関数 $J(\mathcal{M}(\theta), \mathcal{D})$ が次式で与えられるよう設定するのが組織間水平連合学習の問題である．

$$\min_{\theta} J(\mathcal{M}(\theta), \mathcal{D}) = \min_{\theta} \sum_{i=1}^{N} J(\mathcal{M}(\theta), \mathcal{D}_i) \quad (1)$$

ここで，$\mathcal{D} = \{\mathcal{D}_i\}_{i=1}^{N}$ であり，式 (1) の左辺は全組織のデータを結合して誤差関数を求めることを意味するが，プライバシー保護の観点からデータを結合せず，式 (1) の右辺に示すように組織ごとに求めた誤差関数の和で表されるように設定する．

このとき，グローバルモデル $\mathcal{M}(\theta)$ を最急勾配法に基づいて学習するものとすれば，全組織で求めるべき勾配ベクトルは次式で与えられることになる．

$$\nabla_\theta J(\mathcal{M}(\theta), \mathcal{D}) = \sum_{i=1}^{N} \nabla_\theta J(\mathcal{M}(\theta), \mathcal{D}_i) \quad (2)$$

この式が意味するのは，全組織で協調してグローバルモデル $\mathcal{M}(\theta)$ を学習する際，互いの勾配ベクトル $\nabla_\theta J(\mathcal{M}(\theta), \mathcal{D}_i)$ を共有すればよく，一般に勾配ベクトルは組織の多数のデータから求められる統計情報となるため，個人情報には該当せず，プライバシーが保護されるというわけである．

組織間連合学習が満たすべき式 (1) の関係は，多くの最適化問題で強い制約にはならず適用範囲は広い．また，グローバルモデル $\mathcal{M}(\theta)$ は任意の深層学習モデルでよいし，勾配ブースティング決定木アンサンブルや敵対的生成ネットワークモデルなどでもよい．

次節では，勾配ブースティング決定木アンサンブルの連合学習モデルである eFL-Boost (Efficient Federated Learning for Gradient Boosting Decision Trees) を紹介する．

3. 決定木アンサンブル連合学習モデル eFL-Boost

3.1 関連研究

Zhao ら [8] は，グローバルモデルを各組織に順に与えて，各組織のデータで更新する Tree-based Federated Learning (TFL) を提案した．TFL では，グローバルモデルを組織数だけ送信すれば学習できるため，通信コストを抑えることができるメリットがある．しかし，決定木アンサンブルにおいて，組織ごとに増やす決定木が他組織に推測できる可能性があり，決定木ノードの閾値や葉ノードの重みから個人情報が漏洩するリスクがある．TFL では，この漏洩リスクを制御するため差分プライバシーを導入しているが，データなどに加える摂動が性能劣化を導くことは避けられない．

Tian ら [9] は，組織間でデータを共有しなくても，全組織のデータを結合して学習する勾配ブースティング決定木アンサンブルの計算手続に等価となるアルゴリズムを提案しており，これを FederBoost と呼んでいる．アルゴリズムの性質上，XGBoost などの勾配ブースティングアルゴリズムと同等の性能が得られるが，組織間で頻繁に通信を行う必要があり，他の組織に情報が漏洩するリスクもある．特に，ノードごとに求

表 1　変数の定義

Notations	Descriptions
n	全組織のデータ数
n_d	d 番目の組織のデータ数
n_f	特徴数
U	モデル更新回数（= 決定木数）
l	決定木の深さ
λ	正則パラメータ
$\mathbf{X}$	データ行列
$\boldsymbol{y}$	目標値
$(\mathbf{X}_d, \boldsymbol{y}_d)$	組織 d のデータ・目標値ペア
$\boldsymbol{g}_d, \boldsymbol{h}_d$	組織 d で求められた勾配ベクトル
$\mathbb{G}, \mathbb{H}$	葉ノードでの蓄積勾配ベクトル
$\boldsymbol{w}$	葉ノードの重みベクトル
$\mathcal{D}$	組織の集合
$\mathcal{A}gg$	アグリゲータ
$\mathcal{B}$	ビルダー ($\mathcal{B} \in \mathcal{D}$)
$\mathbb{T}_i$	i 番目の木構造
$\mathcal{T}_i$	i 番目の決定木
$\{\mathcal{T}_1, \ldots, \mathcal{T}_U\}$	グローバルモデル

める勾配ヒストグラムは，木の深さに依存した回数の通信が必要であり，そのヒストグラムを構成するデータが少ないとき，入力データの目的変数に関する情報を推測することができる可能性がある．

3.2　eFL-Boost の学習アルゴリズム

TFL や FederBoost の欠点を改良するため，山本ら [10] は，1 回の更新で求める決定木の構造を 1 組織に限定して求める eFL-Boost を提案した．表 1 に eFL-Boost で使う変数の定義を示す．

以下では，全組織で協調して行う計算を大局計算，個別の組織で行う計算を局所計算と呼び，安全性と予測精度を犠牲にせず，組織間の通信コストを削減することを eFL-Boost の目的とする．決定木 $\mathcal{T}$ は，木構造 $\mathbb{T}$ と葉の重み $\boldsymbol{w}$ に分けられ，前者はグラフノードの閾値を含み，後者はモデルの出力を決定する．原理的に，閾値を決定するには，決定木の深さだけ全組織と通信が必要であり，葉ノードの重みは 1 回の組織間通信で済む．葉ノードの重みはモデル出力を決定し，予測精度に大きな影響を与えることから，全組織のデータを使って求めるべきである．そこで，eFL-Boost では，高い通信コストが必要な割に予測精度への影響が小さいと考えられる木構造の決定は局所計算に基づき，葉ノードの重みは全組織が協調して行う大局計算に基づくべきとの方針が取られている．

eFL-Boost は水平型連合学習を仮定しており，学習アルゴリズムは以下の三つのパートに分けられる．

組織：全組織で同じ特徴量をもつが，固有のデータセットをもつ．組織 $d \in \mathcal{D}$ は，ビルダーによって共有された木構造に基づき，自らのデータを使って，葉ノードに関する勾配ベクトル $\mathbb{G}_d$ と $\mathbb{H}_d$ を計算する．

ビルダー：更新ごとに $\mathcal{D}$ から一つの組織が $\mathcal{B}$ として選ばれ，自組織のデータを使って，木の構造を決定する．

アグリゲータ：データを所有している組織とは独立した存在であり，各組織で計算された勾配ベクトル $\mathbb{G}_d$，$\mathbb{H}_d$ が送られてきて，それを統合する．そして，統合した勾配ベクトルからはノードの重みを計算し，各組織に送信する．

なお，ビルダーは特定の組織に決めておいてもよいし，順に組織を回していくことで決めてもよい．プライバシー保護の観点から，ビルダー $\mathcal{B}$ は葉ノードに属するデータが非常に少なくなるような木構造を選ぶべきではない．よって，葉ノードに割り当てられるデータ数に対して下限を設けて，個人情報の漏洩が問題とならないようにすべきである．

Algorithm 1 に eFL-Boost の学習アルゴリズムを示す．まず，アグリゲータは $\mathcal{D}$ からビルダー $\mathcal{B}$ を決定する．次に，各組織 $d \in \mathcal{D}$ は，前回に求めた決定木 $\mathcal{T}_{i-1}$ と各組織のデータを使って，勾配ベクトル $\boldsymbol{g}_d$ と $\boldsymbol{h}_d$ を更新する．そして，ビルダー $\mathcal{B}$ は，自らの組織データと勾配ベクトル $\boldsymbol{g}_d$ と $\boldsymbol{h}_d$ を使って，i 番目の木構造 $\mathbb{T}_i$ を決定し，それを全組織に送信する．各組織は，その木構造 $\mathbb{T}_i$ に基づき，自らのデータを使って蓄積勾配ベクトル $\mathbb{G}_d$，$\mathbb{H}_d$ を計算してアグリゲータに送信する．アグリゲータは，各組織から送られてきた蓄積勾配ベクトルから，葉ノードの重みを計算し，それを各組織に送信する．以上の計算を指定された繰り返し回数 U だけ実施して，学習を終了する．

eFL-Boost では，FederBoost で組織間通信の回数が多くなることを改善するため，木構造が N 組織から選ばれたビルダーのデータのみで決定される．それゆえ，各回で求められる木構造が最適なものでないという意味で，勾配ブースティング決定木アンサンブルを正確には実現していない．しかし，毎回の更新でビルダーを入れ替えていくことで各組織のデータ分布が木構造に反映されると期待され，弱識別器をアンサンブルして強識別器にするブースティング手法の特徴を考えれば，木構造の最適性が保証されない点の影響は軽減されると期待される．eFL-Boost は，連合学習で必要となる大局計算と局所計算のバランスを制御することで，安全性と性能のトレードオフを解決した初めて

Algorithm 1: eFL-Boost の学習アルゴリズム

Input: U: Number of updates,
$(\mathbf{X}_d, \boldsymbol{y}_d)$: Dataset for $d \in \mathcal{D}$,
λ: Regularization parameter
Output: $\{\mathcal{T}_1, \ldots, \mathcal{T}_U\}$: Global model

begin
  for $i \leftarrow 1$ **to** U **do**
    $\mathcal{A}gg$ selects $\mathcal{B}$ from $\mathcal{D}$.
    for each data owner $d \in \mathcal{D}$ **do**
      d updates gradients $\boldsymbol{g}_d$ and $\boldsymbol{h}_d$ based on $\mathcal{T}_{i-1}$ and $(\mathbf{X}_d, \boldsymbol{y}_d)$
    end
    $\mathcal{B}$ computes $\mathbb{T}_i$ from $\mathbf{X}_B, \boldsymbol{g}_B$ and $\boldsymbol{h}_B$.
    $\mathcal{B}$ sends $\mathbb{T}_i$ to $\mathcal{D}$.
    for each data owner $d \in \mathcal{D}$ **do**
      d computes $\mathbb{G}_d$ and $\mathbb{H}_d$ from $\mathbf{X}_d, \boldsymbol{g}_d$ and $\boldsymbol{h}_d$.
      d sends $\mathbb{G}_d$ and $\mathbb{H}_d$ to $\mathcal{A}gg$.
    end
    $\mathcal{A}gg$ computes $\mathbb{G} = \sum_{d\in\mathcal{D}} \mathbb{G}_d$ and $\mathbb{H} = \sum_{d\in\mathcal{D}} \mathbb{H}_d$.
    $\mathcal{A}gg$ computes $\boldsymbol{w}_i = -\frac{\mathbb{G}}{\mathbb{H}+\lambda}$.
    $\mathcal{A}gg$ sends $\boldsymbol{w}_i$ to $\mathcal{D}$.
    for each data owner $d \in \mathcal{D}$ **do**
      d combines $\mathbb{T}_i$ and $\boldsymbol{w}_i$ to obtain $\mathcal{T}_i$.
    end
  end
end

のモデルといえる．

4. 組織間連合学習の社会実装

4.1 金融業界における非競争領域の課題

金融機関の取引は 24 時間 365 日休むことなく行われ，大量の通常取引に隠れて，特殊詐欺やマネーロンダリングなどに紐づいた不正金融取引が少なからず報告されている．具体的には，令和 3 (2021) 年度における「特殊詐欺」の認知件数は 17,570 件 [11]，特定事業者（士業者を除く）から所管行政庁に届け出られた「疑わしい取引」の件数は 58.3 万件程度 [12] であり，1 日当たりで，それぞれ約 48 件，約 1,598 件となる．しかし，これはすべての金融機関で合計した件数であり，2017 年 3 月 31 日現在で銀行数が 534 であることを考えると，1 行当たりで起こる特殊詐欺や疑わしい取引は相当少ないといえる．この事実が意味するところは，どんなに高度な AI を導入したとしても，不正案件（機械学習では正例データと呼ぶ）が少なすぎて，個々の金融機関の努力だけで高精度な不正送金検知を実現することは困難ということである．このような背景もあり，令和 3 (2021) 年 8 月に財務省より発表された「マネロン・テロ資金供与・拡散金融対策に関する行動計画」において，2024 年春を期限とした「取引スクリーニング，取引モニタリングの共同システムの実用化」が行動計画に盛り込まれた．

これを受け，全国銀行協会は 2022 年 10 月 13 日に AML/CFT 業務の高度化・共同化を図ることを目的とした株式会社を新たに設立すると決定し，2023 年 1 月 6 日に株式会社マネー・ロンダリング対策共同機構を設立し，2024 年度以降の段階的なサービス提供を図るとしている [13]．本共同機構が提供するのは，AI スコアリングサービスと業務高度化支援サービスであり，前者に AI が導入されている．しかしながら，AI は既存のルールベースシステムでアラートが出たものに対してリスクをスコアリングし，さらに AI の学習は事前に収集した複数銀行の口座取引データで行われる．よって，AI の更新は頻繁には行えず，犯罪手口の変化には追従できる仕組みをもっていない．

銀行がもつ顧客口座データは個人の金融資産に関する情報であり，当該銀行の外にデータ提供することは難しいが，水平型連合学習を用いて互いに協力し合うことで，実質的に不正取引や不正口座の正例データを増やすことができ，検知精度の向上が期待される．このような背景から，情報通信研究機構 (NICT) と神戸大学，株式会社エルテスは，JST CREST「人工知能」研究領域にて，データの利活用とプライバシー保護を両立できるプライバシー保護データ解析技術の研究開発および実用性検証に取り組み，データを外部に開示することなく機密性を保ったまま機械学習を行う組織間連合学習を活用し，複数の金融機関と連携して不正送金などを自動検知するシステムの実現を目指し，実証実験に取り組んだ．

その結果，目標としていた不正送金の検知精度 80% 以上を達成するとともに，一銀行では検知できなかった不正送金の被害に遭った取引（被害取引）の検知や，不正送金に悪用された口座（不正口座）の早期検知を確認した．具体的には，各銀行の検知目的に応じて，被害取引の検知を目的とする「被害検知グループ」と，不正口座の検知を目的とする「加害検知グループ」とで，それぞれ実証実験を行った [14]．被害検知グループの実証実験には 2 行の銀行が参加し，「単独組織のデータのみを用いた通常の機械学習モデル（個別学習モデル）による検知精度」と「2 行のデータを用いた DeepProtect モデル（連合学習モデル）による検知精度」の比較を行った．その結果，連合学習モデルにより検知精度が向上し，1 日当たりの被害取引のアラート件数を 600 件としたときの検知精度は，当初目標としていた 80%以上を達成した．また，個別学習モデル

では検知できなかった不正取引が検知された事例も確認した．ここで検知率は，検知漏れの少なさを表す指標である再現率で示している．

加害検知グループの実証実験には 4 行の銀行が参加し，「通常取引に使われていた口座が，あるタイミングから犯罪に悪用され，取引停止・凍結されたもの」を不正口座と定義し，不正口座を検知することとした．個別学習モデルとハイブリッドモデル（個別学習モデルと連合学習モデルの組合せ）で比較し，モデルの性能を検知率に加えて，不正口座に対して凍結時点より何週前に検知できるかという検知タイミングで評価した．その結果，個別学習モデルと比較してハイブリッドモデルが高い性能を示し，検知率 80%以上を達成した．さらに，実データでの不正口座凍結よりも 20〜50 週程度の早期検知が可能であることが確認された．

5. さいごに

本解説記事では，組織が有する機微なデータを利活用する仕組みとしてプライバシー保護データ解析を紹介し，その中でもプライバシー保護と高性能性を両立できる組織間連合学習について解説した．まず，組織間連合学習の目的関数を定式化し，個々の組織で最適化すべき損失関数の和で表すことにより，勾配法ベースの機械学習モデルであれば，どのようなモデルでも適用可能であることを示した．そのうえで，深層学習モデルをベースにした組織間連合学習モデルとして，NICT が開発した DeepProtect を紹介し，勾配ブースティング決定木アンサンブルを連合学習に拡張した eFL-Boost について，そのアルゴリズムの詳細を述べた．

また，非競争領域での社会課題のソリューションとして，組織間連合学習が有望であることを銀行不正送金検知実証実験を例に取って述べた．しかし，この実証実験を通して，組織間連合学習における以下の現実的な問題も顕在化している．

問題 1. 顧客口座・取引情報に組織間不整合が解消されず，完全な「水平型連合学習」にならない．

問題 2. 提供される特徴量の数に違いがあり，銀行間で情報量格差が生じる．

問題 3. 犯罪ラベルの定義（凍結口座 and/or 疑わしい取引）が銀行間で異なる．

表 2 に，問題 1 で述べた組織間不整合の一例を示す．これからわかるように，銀行 3 は多くの特徴量を提供できない．よって，銀行 1〜4 で完全な水平型連合学習にするためには，全組織で特徴量が揃っている必要があるが，三つの特徴量しか採用できないことになる．これは銀行 1，2，4 で多くの特徴量を考慮できないことになり，水平型連合学習が性能の劣化をもたらす可能性が高い．また，問題 2 で述べたように，銀行 3 が連合学習で提供できる特徴量が少ないため，銀行 3 以外の銀行にとって協調するメリットが少ないと感じるかもしれない．これを解消する一つの方法は，銀行 3 の四つの特徴量と銀行 4 の一つの特徴量を欠損値として扱い，欠損値補完を導入することが考えられる．また，銀行 1〜4 のデータを使って連合学習するだけでなく，比較的特徴量が揃っている銀行 1，2，3 のデータを使って別の連合学習モデルを学習し，個別学習のモデルを合わせてハイブリッド化することによって，性能改善が図れる可能性がある．今後の課題として検討していく．

表 2　顧客口座・取引情報の組織間不整合の例

	特徴量	銀行1	銀行2	銀行3	銀行4
1	年齢	○	○	○	○
2	法人格（法人・個人）	○	○	○	×
3	口座開設期間	○	○	×	○
4	国籍	○	○	×	○
5	口座・取引場所距離	○	○	×	○
6	入出金時刻	○	○	×	○
7	入金額	○	○	○	○
8	出金額	○	○	○	○

謝辞

本解説記事は，JST CREST「プライバシー保護データ解析技術の社会実装」（課題番号：JPMJCR19F6）で NICT とエルテスとの共同して得られた研究成果を一部引用している．

参考文献

[1] C. Dwork, “An ad omnia approach to defining and achieving private data analysis,” *Privacy, Security, and Trust in KDD*, F. Bonchi, E. Ferrari, B. Malin and Y, Saygin (eds.), Springer, pp. 1–13, 2008.

[2] 寺田雅之，“差分プライバシーとは何か，”システム/制御/情報，**63**, pp. 58–63, 2019.

[3] C. Gentry, “Fully homomorphic encryption using ideal lattices,” In *Proceedings of the STOC*, pp. 169–178, 2009.

[4] 佐久間淳，『データ解析におけるプライバシー保護』，講談社，2016.

[5] J. Konečnỳ, H. B. McMahan, F. X. Yu, P. Richtárik, A. T. Suresh and D. Bacon, “Federated learning: Strategies for improving communication efficiency,”

arXiv preprint, arXiv:1610.05492, 2016.
[6] Q. Yang, Y. Liu, T. Chen and Y. Tong, "Federated machine learning: Concept and applications," *ACM Trans. on Intelligent Systems and Technology*, **10**, pp. 1–19, 2019.
[7] L. T. Phong, Y. Aono, T. Hayashi, L. Wang and S. Moriai, "Privacy-preserving deep learning via additively homomorphic encryption," *IEEE Trans. on Information Forensics and Security*, **13**, pp. 1333–1345, 2018.
[8] L. Zhao, L. Ni, S. Hu, Y. Chen, P. Zhou, F. Xiao and L. Wu, "Inprivate digging: Enabling tree-based distributed data mining with differential privacy," In *Proceedings of the IEEE INFOCOM 2018-IEEE Conference on Computer Communications*, pp. 2087–2095, 2018.
[9] Z. Tian, R. Zhang, X. Hou, J. Liu and K. Ren, "Federboost: Private federated learning for GBDT," *arXiv preprint*, arXiv:2011.02796, 2020.
[10] F. Yamamoto, S. Ozawa and L. Wang, "eFL-Boost: Efficient federated learning for gradient boosting decision trees," *IEEE Access*, **10**, pp. 43954–43963, 2022.
[11] 警視庁,「特殊詐欺対策ページ」, https://www.npa.go.jp/bureau/safetylife/sos47/circumstances/ (2024 年 2 月 12 日閲覧)
[12] 警察庁,「犯罪収益移転防止に関する年次報告書(令和 4 年)」, https://www.npa.go.jp/sosikihanzai/jafic/nenzihokoku/data/jafic_2022.pdf (2024 年 2 月 12 日閲覧)
[13] マネー・ローンダリング対策共同機構, https://www.caml.co.jp/index.html (2024 年 2 月 12 日閲覧)
[14] 神戸大学,「プライバシー保護連合学習技術を活用した不正送金検知の実証実験を実施」, https://www.kobe-u.ac.jp/ja/news/article/2022_03_10_01/ (2024 年 2 月 12 日閲覧)

