連合学習技術の最前線
～研究開発の取り組みと応用事例の紹介～

MSIISM Conference 2024 @ 野村コンファレンスプラザ日本橋

2024年10月11日

日本電信電話株式会社　社会情報研究所　深見匠
株式会社NTTデータ数理システム　シミュレーション＆マイニング部　村田智也

目次

前半パート (NTTデータ数理システム村田)

1. 連合学習の基礎
2. 差分プライバシーに基づくプライバシー保証
3. プライバシー保護つき連合学習の共同研究の紹介
4. 前半パートのまとめ

後半パート (NTT社会情報研究所深見)

1. 連合学習の取り組み紹介
    1.1. 医療支援
    1.2. 悪性通信検知
    1.3. 言語処理
2. 後半パートのまとめ

目次

前半パート (NTTデータ数理システム村田)

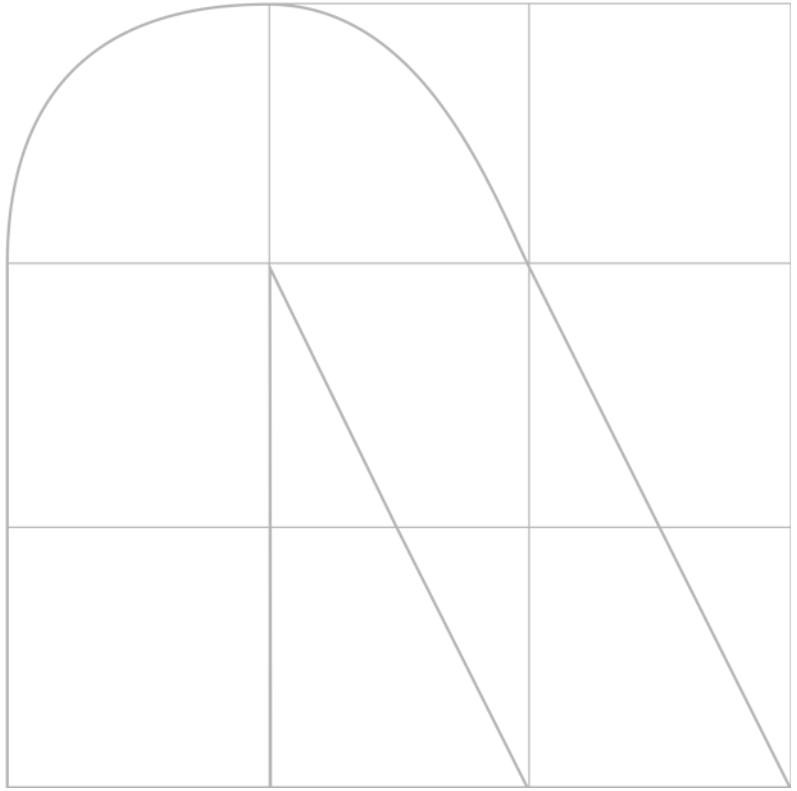
1. 連合学習の基礎
2. 差分プライバシーに基づくプライバシー保証
3. プライバシー保護つき連合学習の共同研究の紹介
4. 前半パートのまとめ

後半パート (NTT社会情報研究所深見)

1. 連合学習の取り組み紹介
    1.1. 医療支援
    1.2. 悪性通信検知
    1.3. 言語処理
2. 後半パートのまとめ

01

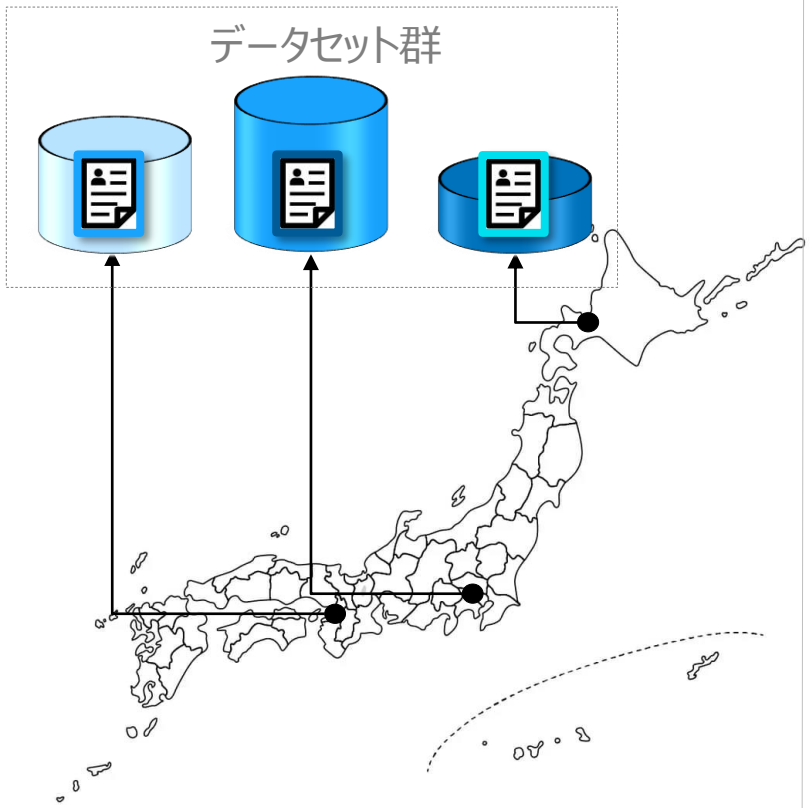
連合学習の基礎



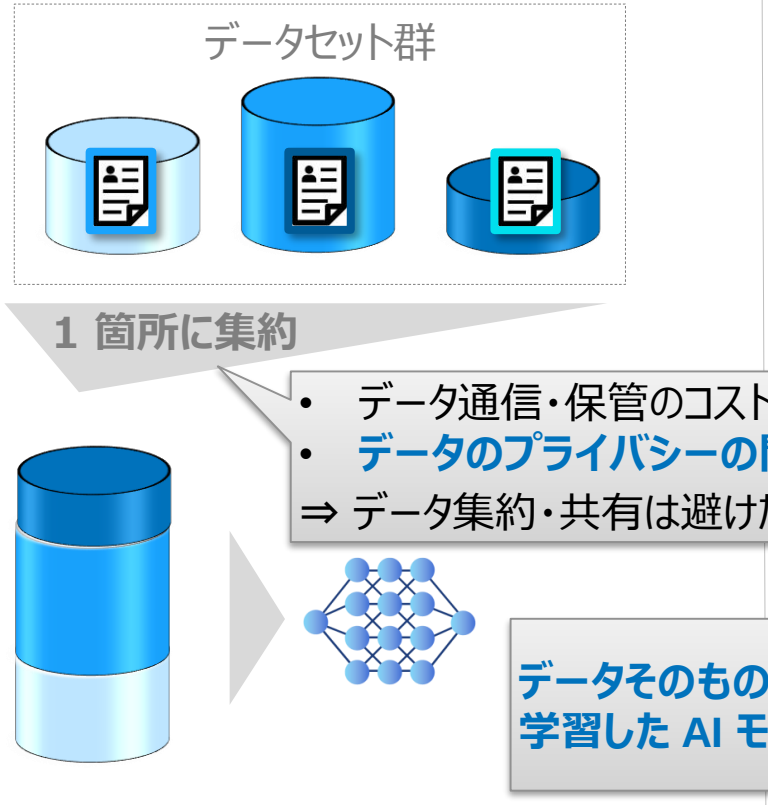
1.1 連合学習とは

連合学習 (Federated Learning) とは、**分散したデータセット群を分散させたまま AI モデルを学習**する仕組み。

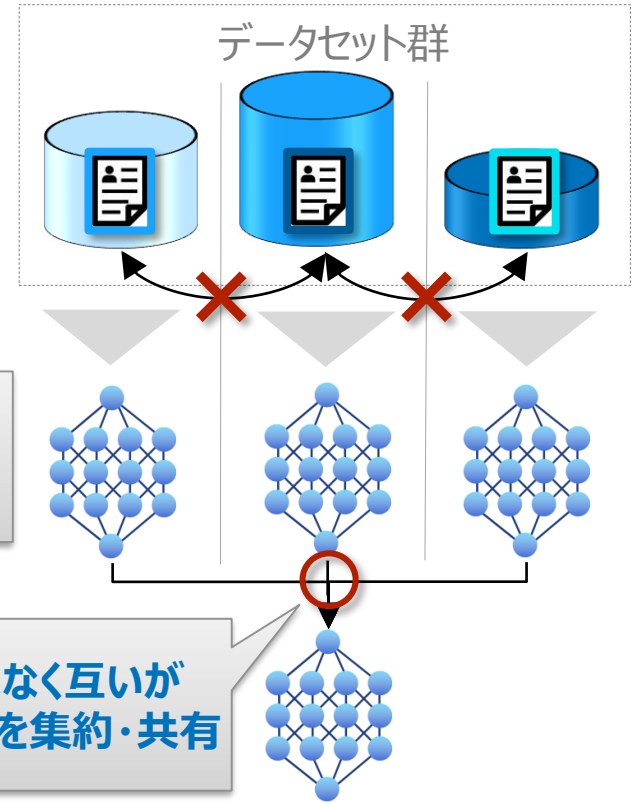
分散したデータセット群の仮想例



従来の AI 学習とその問題点



連合学習のイメージ



1.2 連合学習における二つのカテゴリー

	クロスデバイス学習	クロスサイロ学習
概要	クライアントはある**個人の IoT デバイス**に対応 デバイス間で協力して AI モデルを学習	クライアントは地理的に離れた**組織**に対応 組織間で協力して AI モデルを学習
クライアント例		
クライアント数	非常に膨大 ($\sim 10^{10}$)	$2 \sim 100$
学習参加形式	学習中一部クライアントのみランダムに参加 学習参加中の脱落も起きうる	全クライアントが常に学習に参加 学習参加中脱落は起きない

1.3 連合学習の活用事例

クロスデバイス学習

- **(IT・通信) スマートフォン予測変換**

  Google は、スマートフォンユーザーのプライベート情報を含む予測変換履歴を学習データとして連合学習で予測変換モデルを構築。

- **(IT・通信) 音声認識**

  Apple は、iPhone の音声認識システム Siri の改良にユーザーの音声データを利用した連合学習を活用。

- **(医療) ヘルスモニタリング用活動推定**

  中山大学・アリゾナ州立大学は、活動センサーから個人の活動 (歩く、寝る、座るなど) 推定モデルを連合学習で構築

クロスサイロ学習

- **(医療) 酸素投与判断**

  NVIDIA は、20 医療機関の胸部 X 線やバイタル情報等を用いて連合学習で COVID-19 酸素投与判断モデルを構築。

- **(医療) 創薬**

  AMGENやAstraZeneca などの 10 の 製薬企業は、各企業の創薬データを共有せずに連合学習で創薬モデルを共同構築。

- **(金融) 不正送金検知**

  日本国内金融機関 5 行は、顧客の金融取引データを用いて不正送金検知のための共同モデル開発の実証実験を実施。

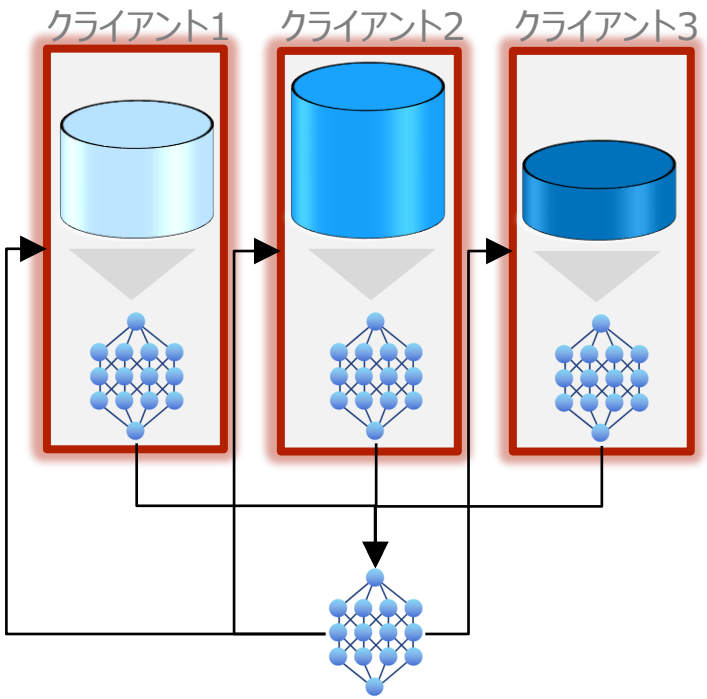
世界中で連合学習を活用したイノベーションが広がっている

1.4 連合学習の標準アルゴ: FedAvg

FedAvg は以下①②③のステップを繰り返す:

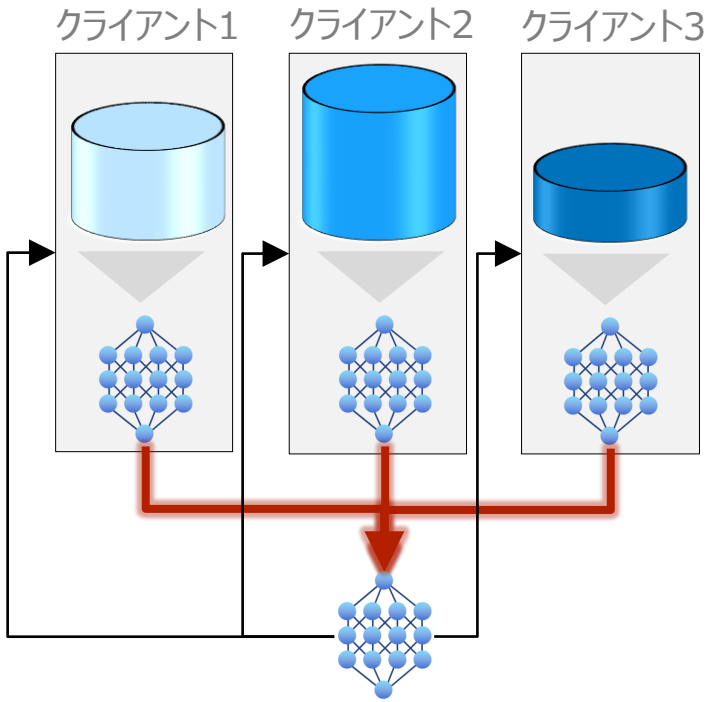
①ローカル更新:

各クライアントは独立に自データセットでローカルモデルを学習



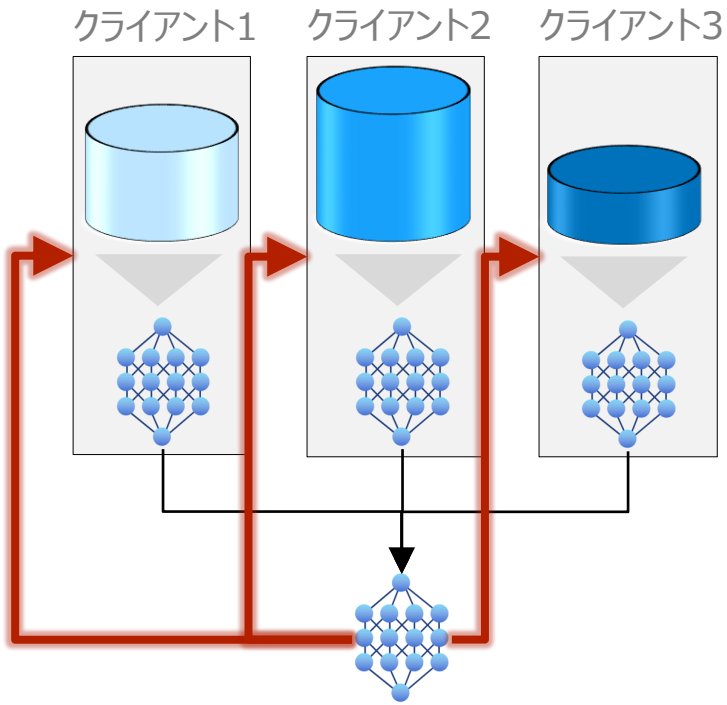
②モデル集約:

①の各クライアントのモデルを平均しグローバルモデルを作成



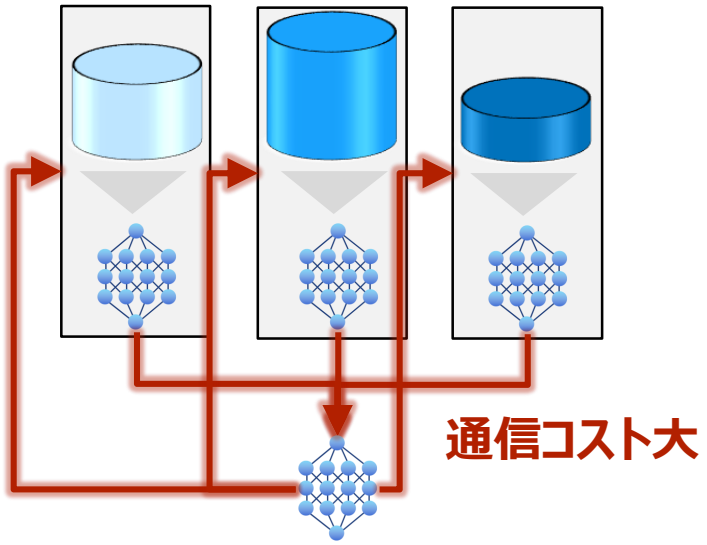
③モデル再配布:

②で作成したグローバルモデルを各クライアントに配布し、①に戻る



1.5 連合学習における三つの課題

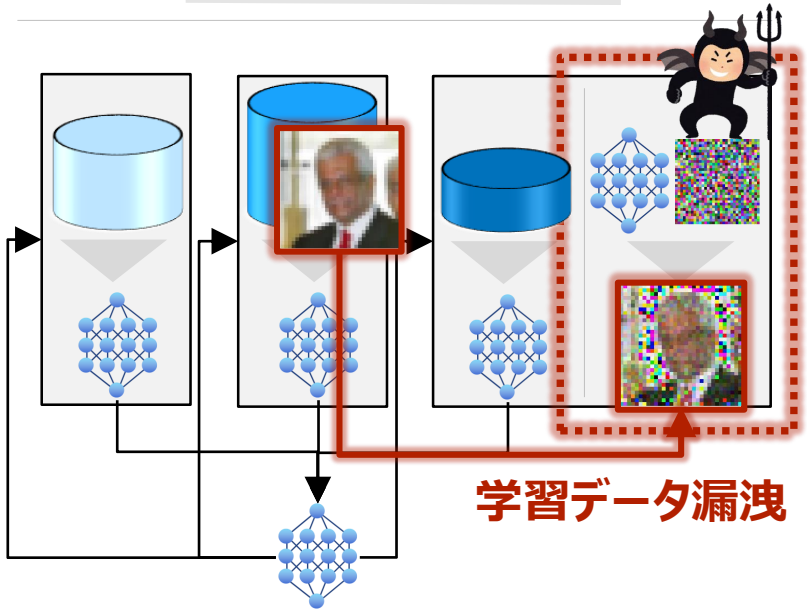
①通信コスト



AIモデルを通信によって共有する必要
⇒ 通信コストがボトルネックに

FedAvg を改良した**高通信効率**な
連合学習アルゴが必要

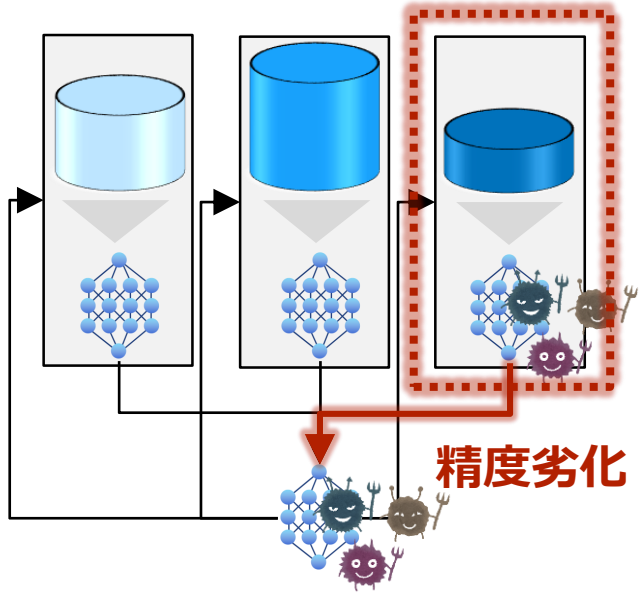
②プライバシー保護



共有モデルは学習データ情報を反映
⇒ 共有モデルからデータを復元可能

差分プライバシー、秘密計算に基づく
プライバシー保護アルゴが必要

③頑健性



クライアントが意図しない挙動
⇒ 学習モデルの精度劣化が発生

一部クライアントの異常挙動に耐性の
ある連合学習アルゴが必要

1.6 連合学習の活用に向けてまず検討すべきこと

本番環境の分散システム構築だけでなく、**事前に検討・検証しておくべき重要な観点**が複数ある：

必要に応じて

①前提条件確認 → **②要件整理** → **③事前検証**

- データが分散して存在？
- データは 1 箇所に集約 NG？
- 自データで学習したモデルの共有は OK？

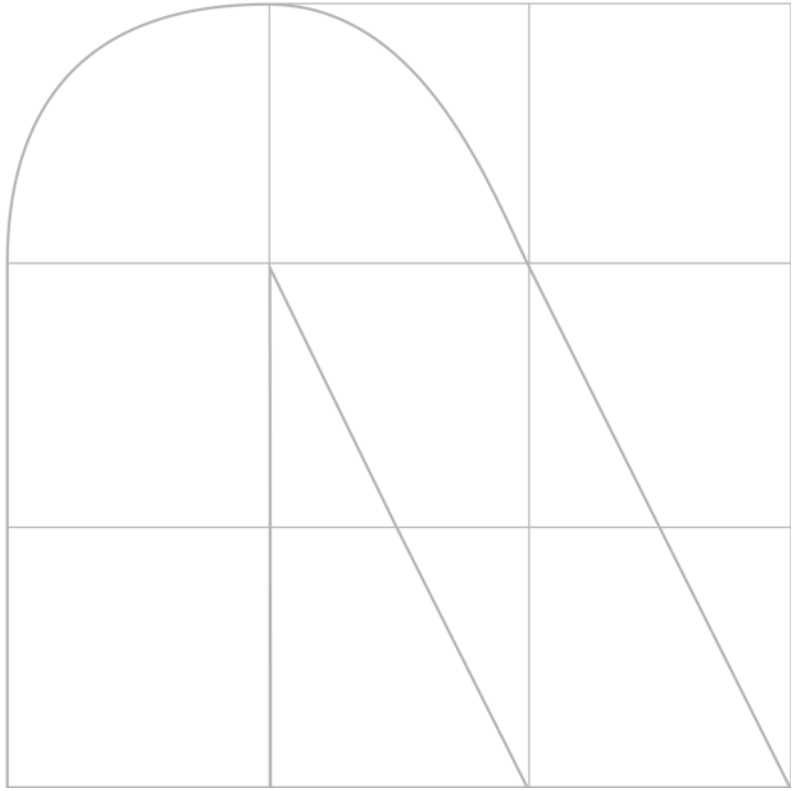
- 学習対象モデルは？サイズは？
- データは量・質はクライアント間でどの程度均一？
- 各クライアントの計算・通信スピードは？
- 中央集権 or 非中央集権？
- 追加のプライバシー保護・頑健性保証は必要？
- モデルの個別化は必要？

- 単独学習より連合学習の方が精度向上？
- データ集約の場合と比べ、連合学習はどの程度精度劣化？
- 全体の学習時間のボトルネックはどこか？改善の必要は？

＊ 事前検証では、通常人工データや公開データ、もしくは 1 地点のデータのみを用いたシミュレーションで実施することが多い

02

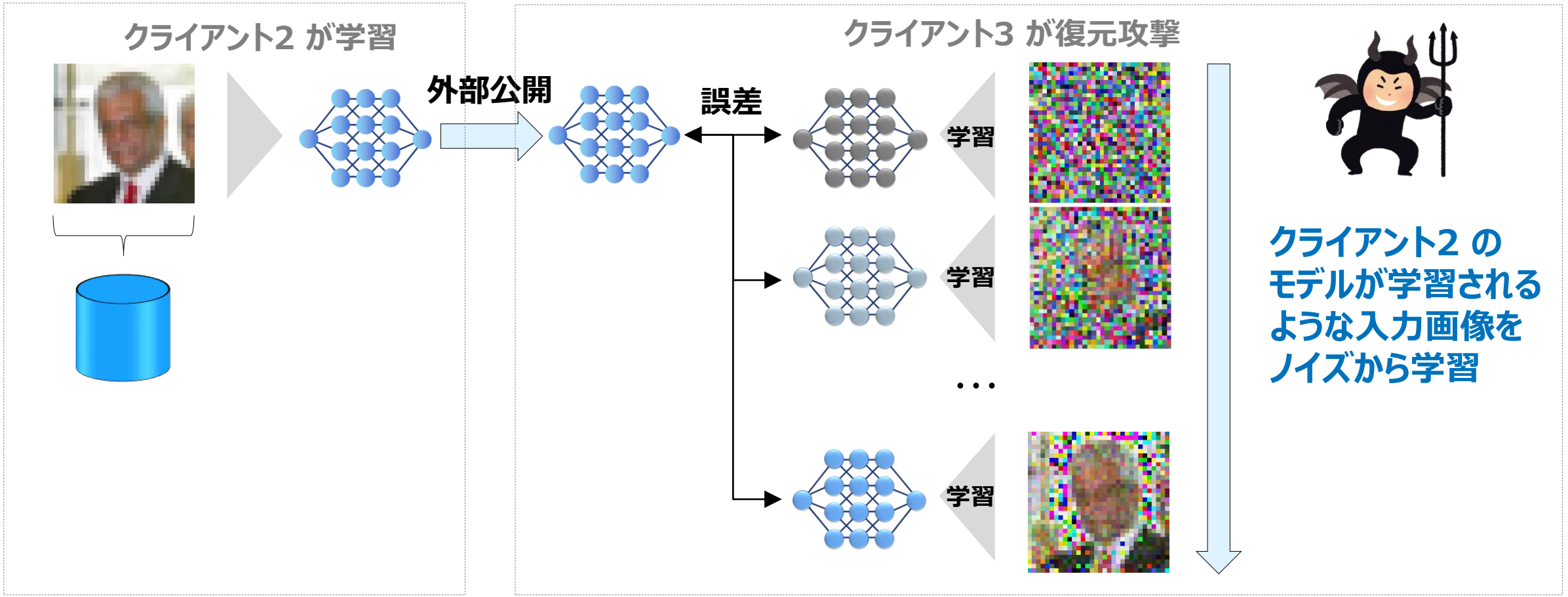
差分プライバシーに基づくプライバシー保護



2.1 連合学習におけるプライバシー保護の必要性

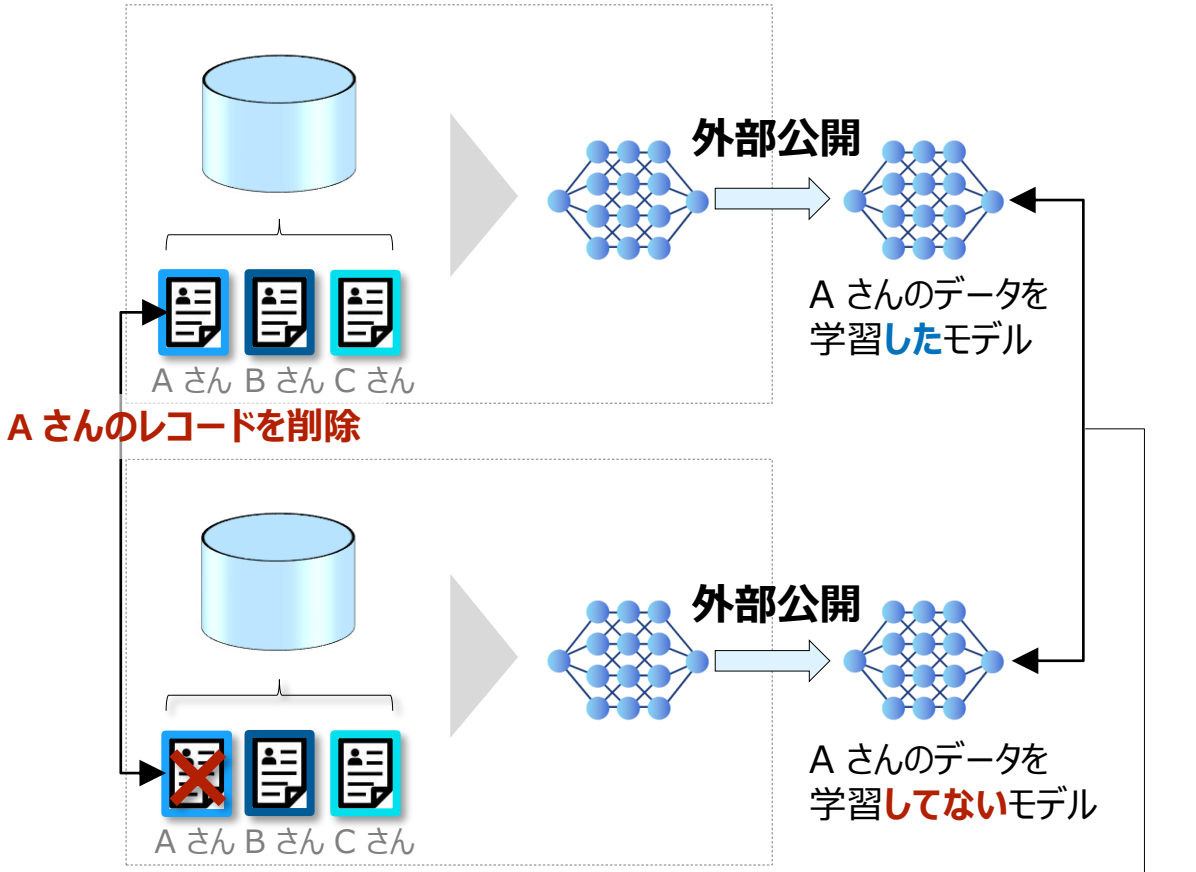
連合学習では学習データの共有を禁じているが、実は **AI モデルから学習データが復元可能**。

* Zhu et al., 2019, Deep Leakage from Gradients



2.2 差分プライバシーとそれに基づくプライバシー保証

差分プライバシー: **外部公開情報から、学習元となる個別データがどれだけ流出しやすいか** を定量的に評価。



2 モデルに (非確率的な) 差があればそれは A さんのレコードに起因
⇒ モデル差を見ることで A さんの情報が漏洩しうる

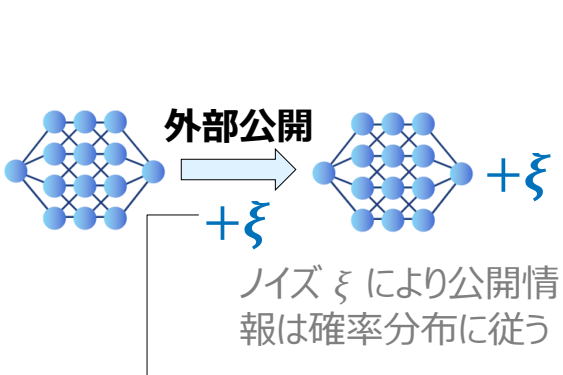
差分プライバシーに基づくプライバシー保護:

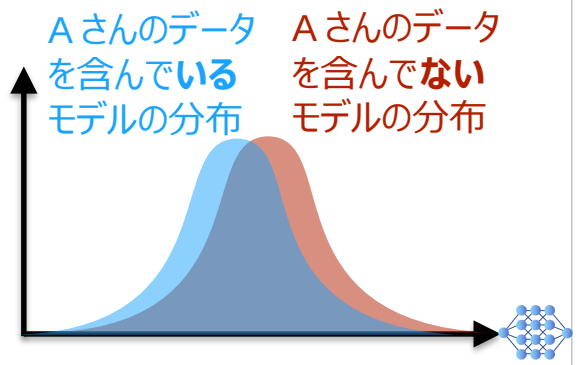
モデルにノイズを足してから外部に公開

⇒ 2 つのモデルはノイズの影響で確率的に変動。

2 モデルの確率分布の近さをプライバシー保護強度と定義。

(ε, δ)-差分プライベートアルゴ: 2 つの確率分布がプライバシーパラメータ ε, δ で規定されるある近さになるようなアルゴのこと。

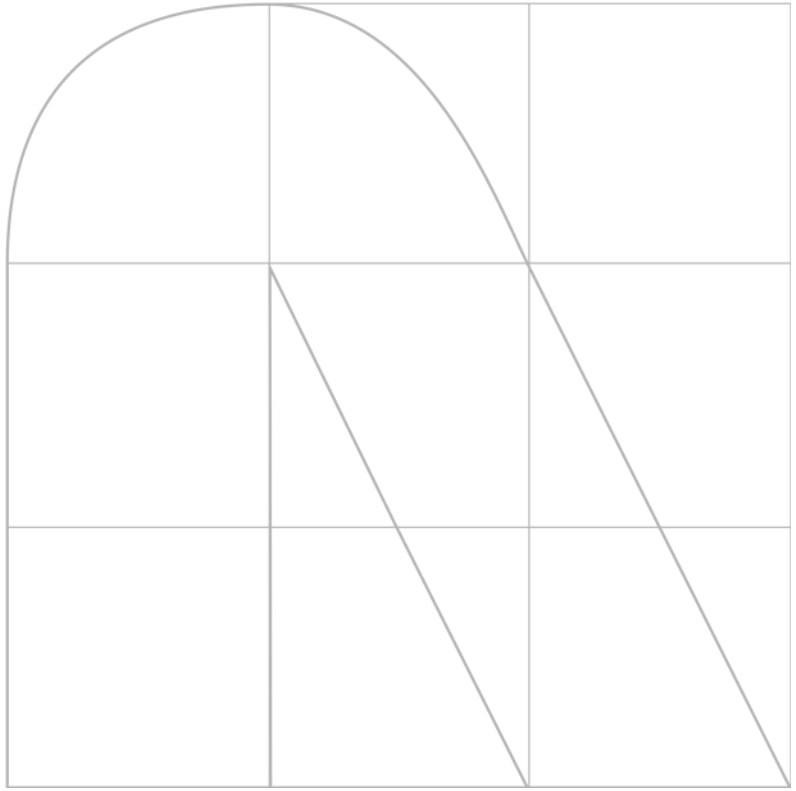




ξ は通常 $N(0, \sigma)$ から取る。σ はサンプル数から (ε, δ)-差分プライバシーを満たすように**数学的に決定可能**

03

プライバシー保護つき連合学習の共同研究の紹介



3.1 共同研究 (DP-Norm) の紹介1: 概要とポイント

差分プライベートなアルゴはプライバシーを保護するが、モデルに**ノイズを足すためモデル精度が低下**する傾向。

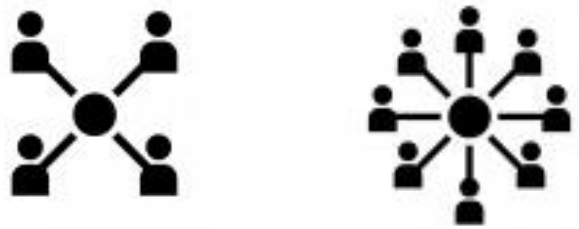
⇒ **精度劣化を緩和**する**デノイジング機構**を**主双対法**に取り入れたアルゴを提案。

* DP-Norm: Differential Privacy Primal-Dual Algorithm for Decentralized Federated Learning, Takumi Fukami, Tomoya Murata, Kenta Niwa, Iifan Tyou, IEEE Transactions on Information Forensics and Security (2024)

ポイント:

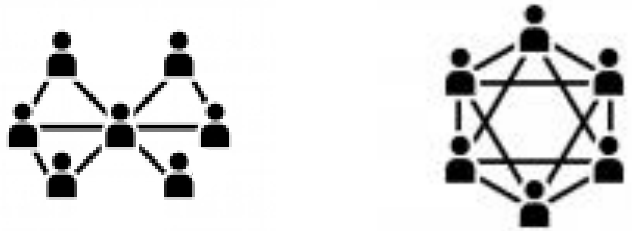
- **非中央集権型**の連合学習
- **主双対法**を利用
- **デノイジング機構**の導入

中央集権型



中央サーバーがクライアント間の通信を仲立ち

非中央集権型



中央サーバーなしで隣接クライアントのみと通信

NTT DATA

3.2 共同研究 (DP-Norm) の紹介2: 主双対法とデノイジング機構

主双対法は、**非中央集権型**の連合学習において**主双対問題**に基づく学習手法。
クライアントの所有するデータセット間の**不均一性に対する頑健性**があることが知られている。

主問題

$$\min_{x} \sum_{i} f_i(x)$$

x: 学習パラメータ (主変数)
f_i: クライアント i の目的関数

変形

主双対問題

$$\min_{X} \max_{\Lambda} \sum_{i} f_i(x_i) + \sum_{i,j:\text{隣接}} \langle \lambda_{ij}, x_i - x_j \rangle + \rho \|x_i - x_j\|^2$$

λ_{ij}: 学習パラメータ(双対変数)
ρ: ハイパーパラメータ

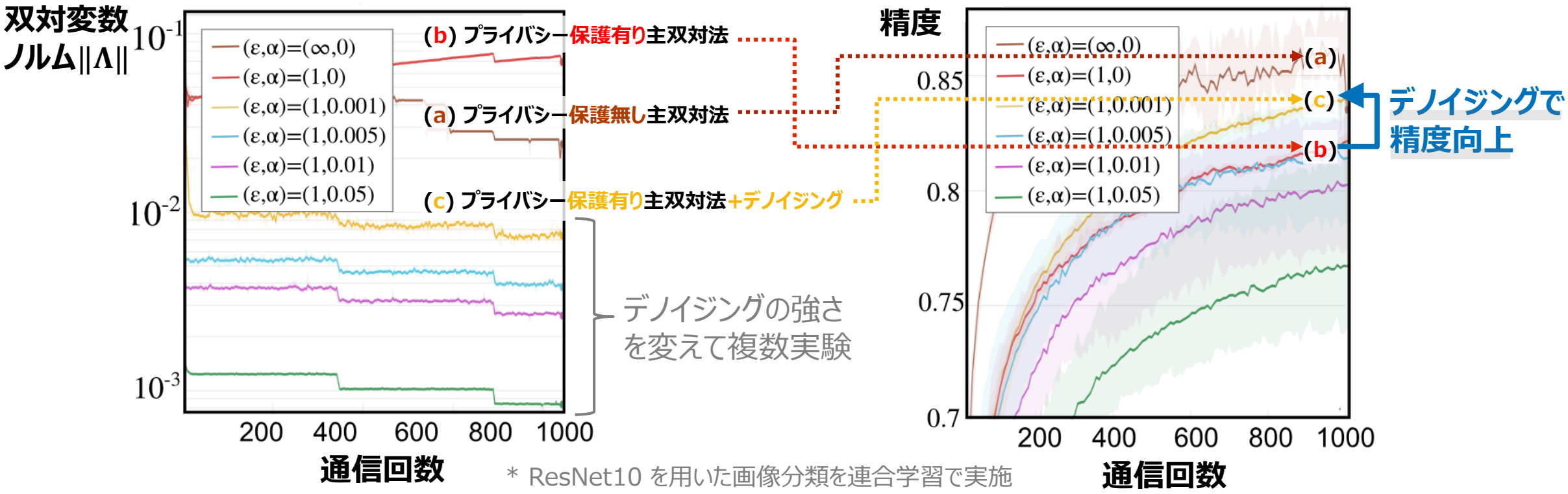
主双対法と差分プライバシーを単純に組み合わせると、**ノイズにより双対変数 Λ が大きくなる問題**が発生
⇒ **双対変数に関する L2 正則化**を付加することでこの問題を解消

主双対問題 + デノイジング機構

$$\min_{X} \max_{\Lambda} \sum_{i} f_i(x_i) + \sum_{i,j:\text{隣接}} \langle \lambda_{ij}, x_i - x_j \rangle + \rho \|x_i - x_j\|^2 - \alpha \|\Lambda\|^2$$

3.3 共同研究 (DP-Norm) の紹介3: 検証実験結果

ε: プライバシーパラメータ (∞ は保護無しに対応)
α: デノイジングの強さ (0 はデノイジング無しに対応)



＊ ResNet10 を用いた画像分類を連合学習で実施

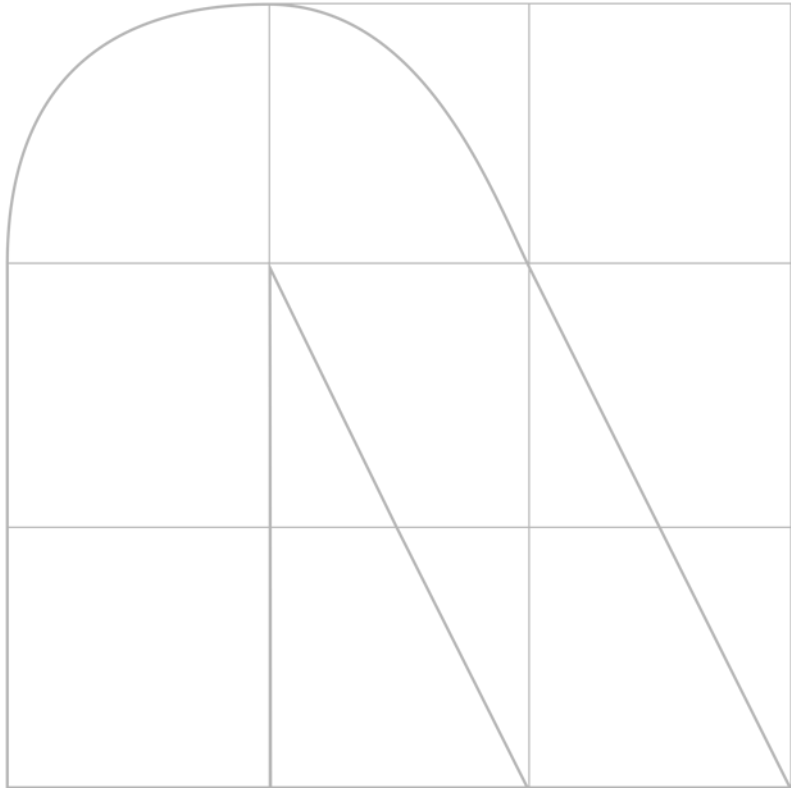
- (a): 双対変数ノルム 10^{-2} 程度
- (b): 双対変数ノルム 10^{-1} 程度と (a) より上昇
- (c): デノイジングにより**双対変数ノルム**を 10^{-2} まで**抑制**

- 単純なプライバシー保護有り (b) はプライバシー保護無し手法 (a) と比べ大きな精度劣化
- **適切なデノイジング** (c) により単純なプライバシー保護有り手法 (b) の**精度劣化を緩和**

NTT DATA

04

前半パートのまとめ



4.1 前半パートのまとめ

01 連合学習の基礎

- 連合学習: 分散したデータを直接共有せずにモデルを共有することで学習を実現
- スマートフォン等を活用したクロスデバイス学習、病院や銀行間でのクロスサイロ学習などの活用事例
- 標準アルゴ: FedAvg
- 三つの課題: 通信コスト、プライバシー保護、頑健性
- 連合学習活用に向けて事前に検討・検証すべきこと

02 差分プライバシーに基づくプライバシー保護

- 復元攻撃によるデータ窃取 ⇒ 追加のプライバシー保護の必要性
- データ漏洩のしやすさを定量的に評価する差分プライバシー
- モデルにノイズを付加してから外部に公開 ⇒ 差分プライバシーに基づくプライバシー保護

03 プライバシー保護つき連合学習の共同研究の紹介

- ノイズ付加に伴う精度劣化を緩和する「デノイジング機構」を主双対法に取り入れる
- デノイジングは双対変数に関する L2 正則化により実現
- 実験的にデノイジングにより双対変数ノルムの増大を緩和 & 精度向上を観察

目次

前半パート (NTTデータ数理システム村田)

1. 連合学習の基礎
2. 差分プライバシーに基づくプライバシー保証
3. プライバシー保護つき連合学習の共同研究の紹介
4. 前半パートのまとめ

後半パート（NTT社会情報研究所深見）

1. 連合学習の取り組み紹介
   - 1.1. 医療支援
   - 1.2. 悪性通信検知
   - 1.3. 言語処理
2. 後半パートのまとめ

1. 連合学習の取り組み紹介

我々の連合学習の取り組み

01　医療支援

各病院が異なるレントゲンデータを持つことを想定し, データを共有せずに連合学習を用いて共同で診断モデルを学習

02　悪性通信検知

ネットワーク悪性通信情報が分散保持されていることを想定し, データを共有せずに連合学習を用いて悪性通信検知モデルを学習

03　言語処理

分散保持された文書理解モデル構築のため, 文書画像とそれに対するQAを連合学習を用いて学習

1.1. 医療支援における連合学習の取り組み

医療支援AIとは

AIを活用して患者の診断や治療を支援

- AIの支援による診断精度の向上
- 個人向けの予測分析
- ウェアラブルデバイスを活用した健康予測

過去の膨大なデータを学習したAIに患者のデータを入力し，検査結果を予測する

予測結果を考慮し，診断・治療

医療支援AIの課題

医療支援AIの実現において，以下の課題が存在

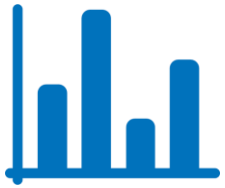
学習データ不足

希少疾患など特定のデータが少ない

データプライバシー

医療情報の共有はできない

データの偏り



組織ごとに持つデータの性質・数が異なる

医療支援AIの最新の取り組み

- **メラノーマ判定**

German Cancer Research Center(DKFZ)は, 6大学病院で取得されたWSIを用いて
メラノーマ判定モデルを学習

- **医療必要性判定**

日本国内病院は, 患者データを用いて, 追加の医療行為が必要有無を判断する
モデルを学習

- **創薬**

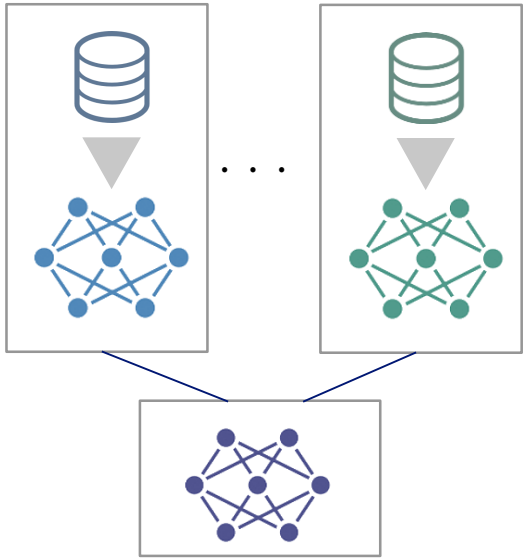
研究機関と製薬企業は, 創薬データを用いて, 統合創薬モデルを学習

医療支援AIへの我々の取り組み

我々の連合学習を用いて，胸部X線写真診察モデルを学習

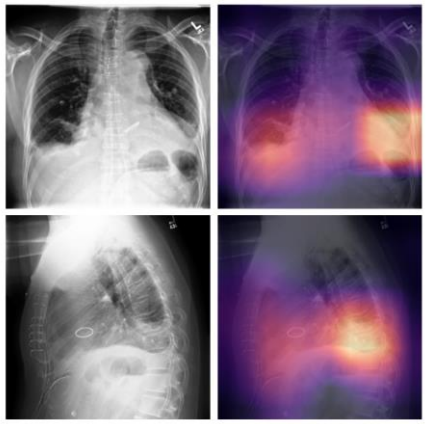
連合学習構成

クライアント7台



学習データについて

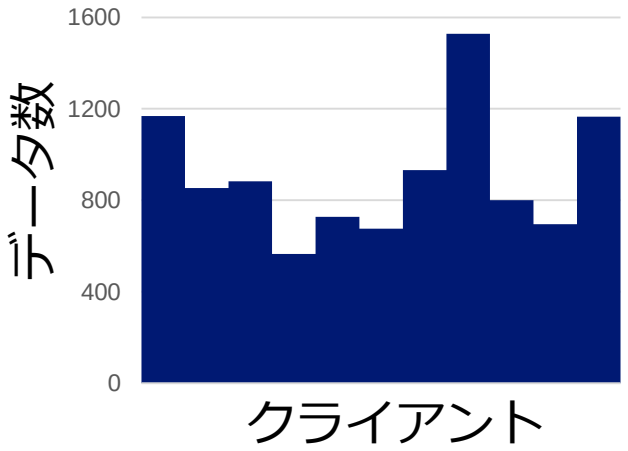
Chexpert small データセットを使用



Irvin, Jeremy, et al. "Chexpert: A large chest radiograph dataset with uncertainty labels and expert comparison." *Proceedings of the AAAI conference on artificial intelligence*. Vol. 33. No. 01. 2019.

データの分散方法

Japan Medical Image Datasetのような医療情報連携を想定し，以下のようにデータをクライアントに分割

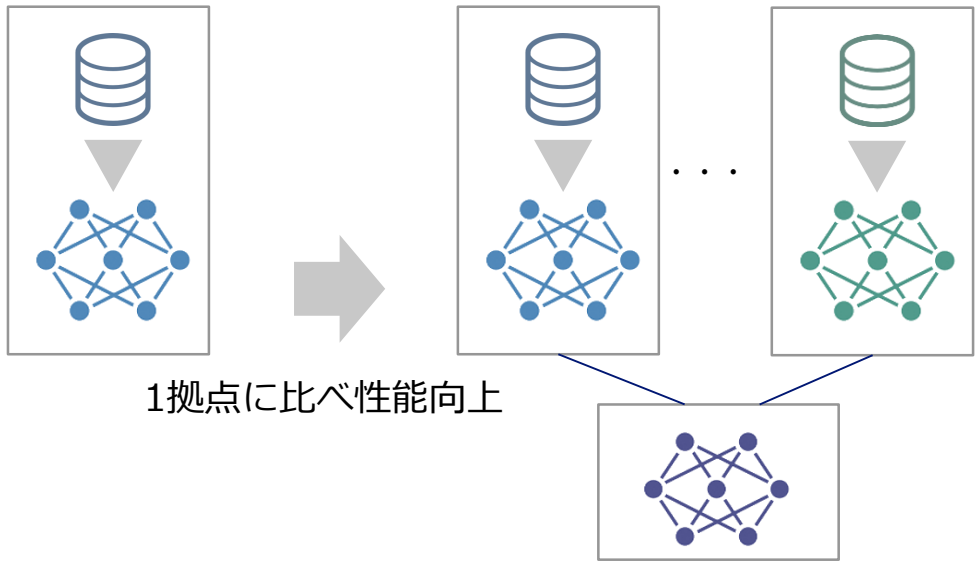


医療支援AIの性能

連合学習を用いることにより，1拠点で学習したモデルより性能向上

一拠点での学習　　連合学習

複数病院のデータを学習可能



	手法	AUROC
1拠点	–	0.736
FL	FedAvg	0.747
	SCAFFOLD	0.772
	Ours	**0.797**

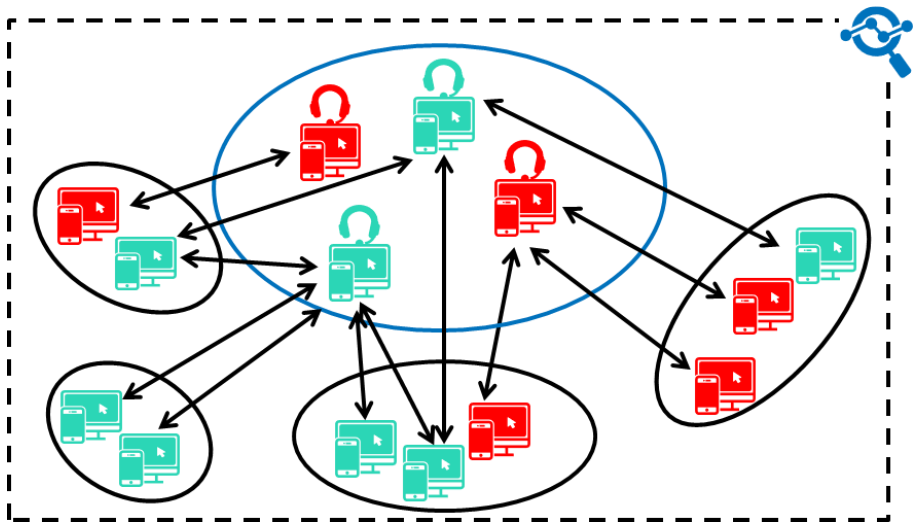
1.2. 悪性通信検知における連合学習の取り組み

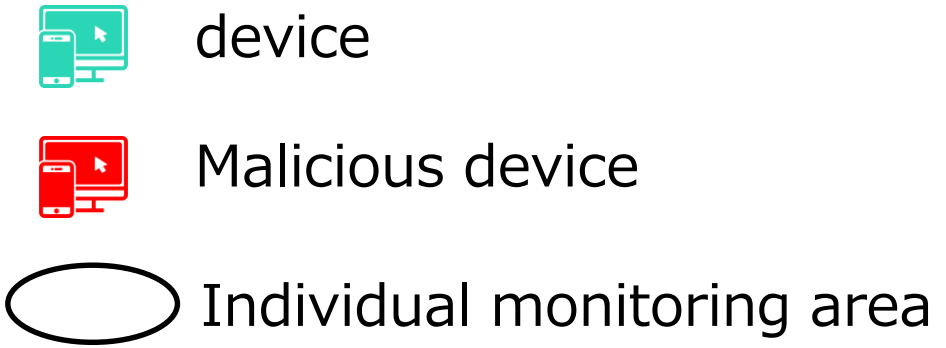
悪性通信検知とは

悪性通信やそれを行うマシンを検知

悪性通信を含む大量のデータを用い，既知の通信の類似性をとらえることが重要

- ネットワークの悪性通信検知
- 銀行不正送金検知



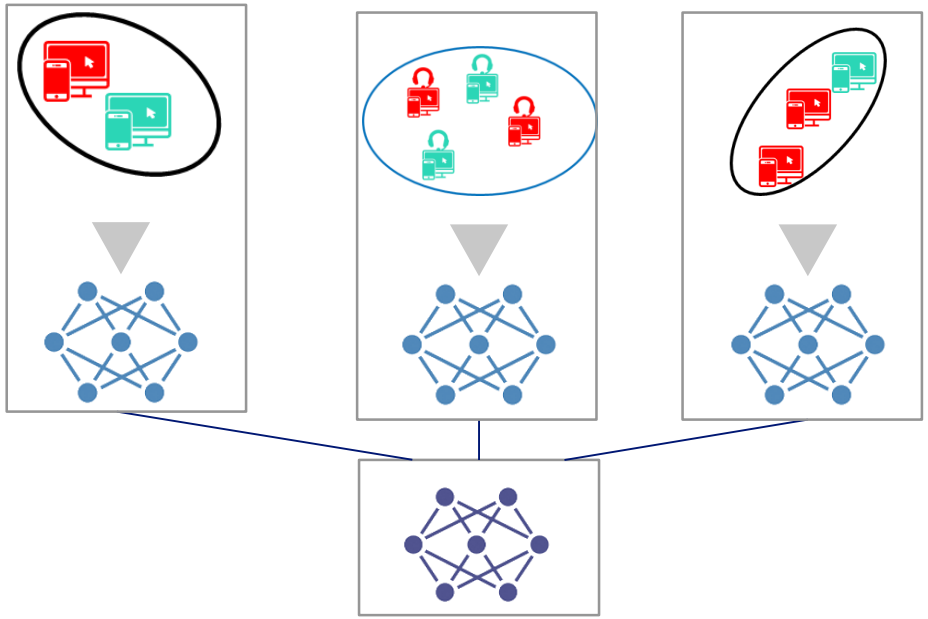


各企業が持つ**ブロックリストを共有・連携**できると良いが
データの機微性を理由に，データ共有には積極的でないことが多い

悪性通信検知への我々の取り組み

我々の連合学習を用いて, ネットワーク悪性検知モデルを学習

大規模ISPから収集した通信ログを5クライアントに分割

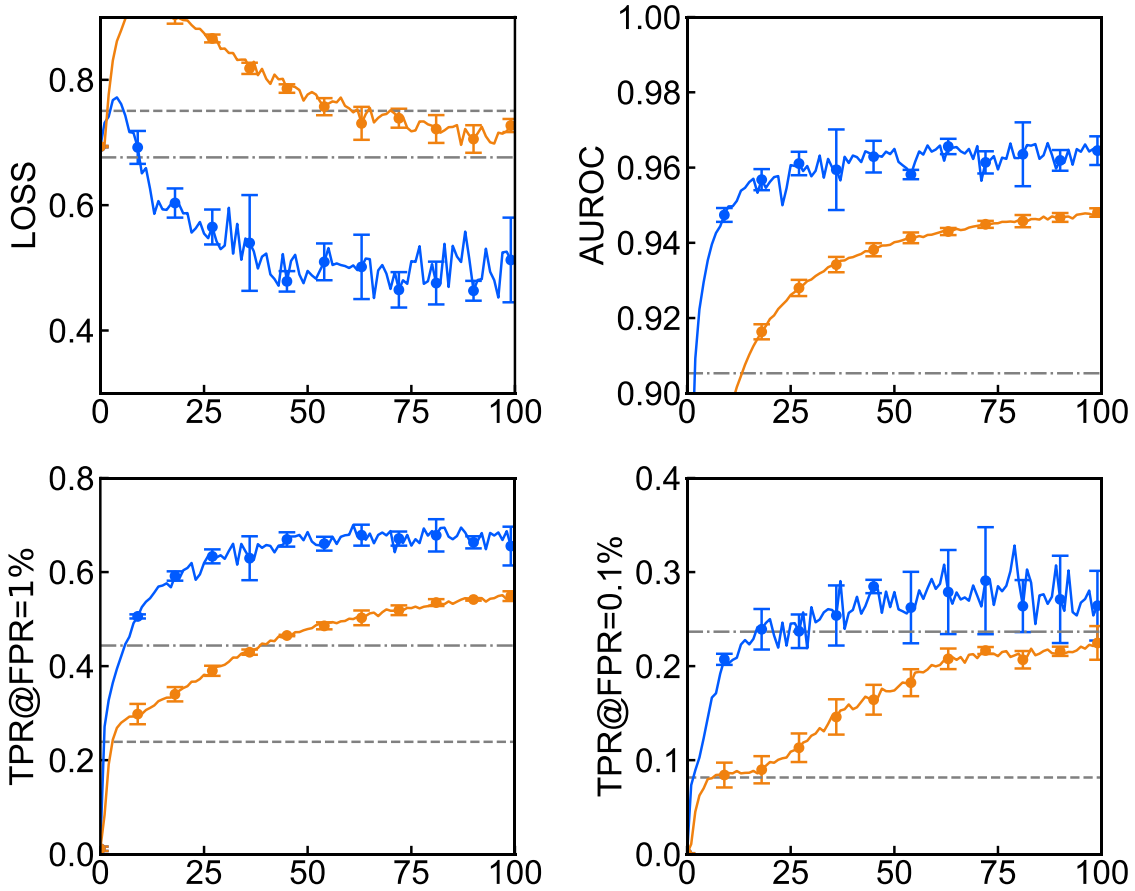


IPアドレスをMaxmind社のGeoIP Lite DBに問い合わせて得られた国毎に分割．フローの総数で上位5カ国のフローデータを各Client のデータとし分配

Client Ind.	1	2	3	4	5
benign	495785	61454	37499	27832	26504
malicious	4126	35	1661	1156	1238

悪性通信検知の性能

全ての指標で連合学習を用いることでSoloよりも性能を改善

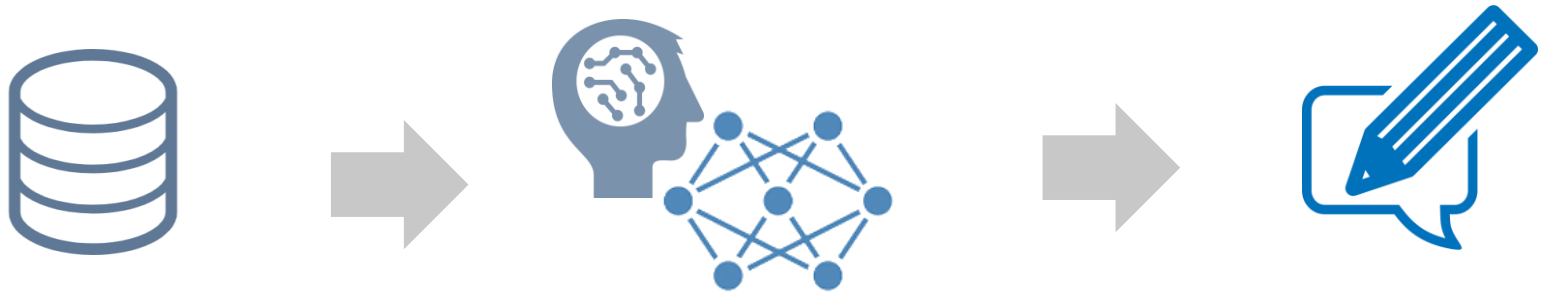


1.3. 言語処理における連合学習の取り組み

言語処理とは

人間が日常的に用いている言語を理解・学習する

- 対話エージェント
- 情報抽出
- 文章分析



LLM（大規模言語モデル）は大量のデータを
学習する大規模な自然言語処理モデル

大量のデータで事前学習した後に, 各々のデータを用いてファインチューニング
することでタスクに特化して性能を向上する

言語処理の最新の取り組み

- **事前学習大規模言語モデル**

事前学習に連合学習を用いることで, 様々なデータと計算リソースを活用

- **AIアシスタントのパーソナライゼーション**

各ユーザのデータを活用することで, テキストの次単語予測や絵文字の提案などを実現

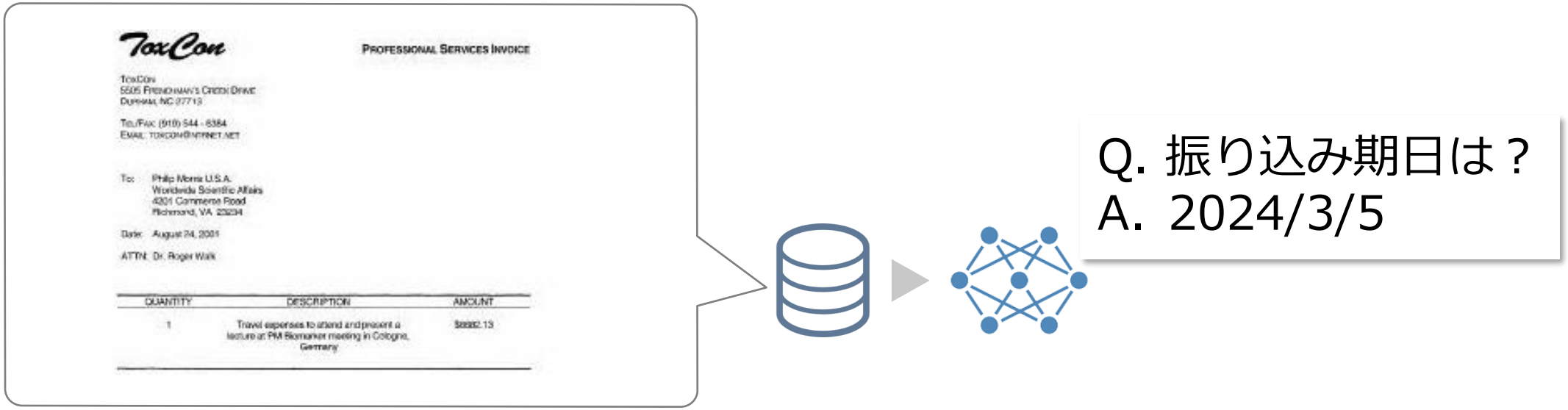
- **文書解析**

各企業の契約書などの文書を学習し, 文書レビューや内容要約モデルを学習

DocVQAとは

Document Visual Question Answeringは画像内の文書に関連する質問応答するタスク

- 画像内の文章や文書の位置を学習
- 画像内の文章を用いて回答



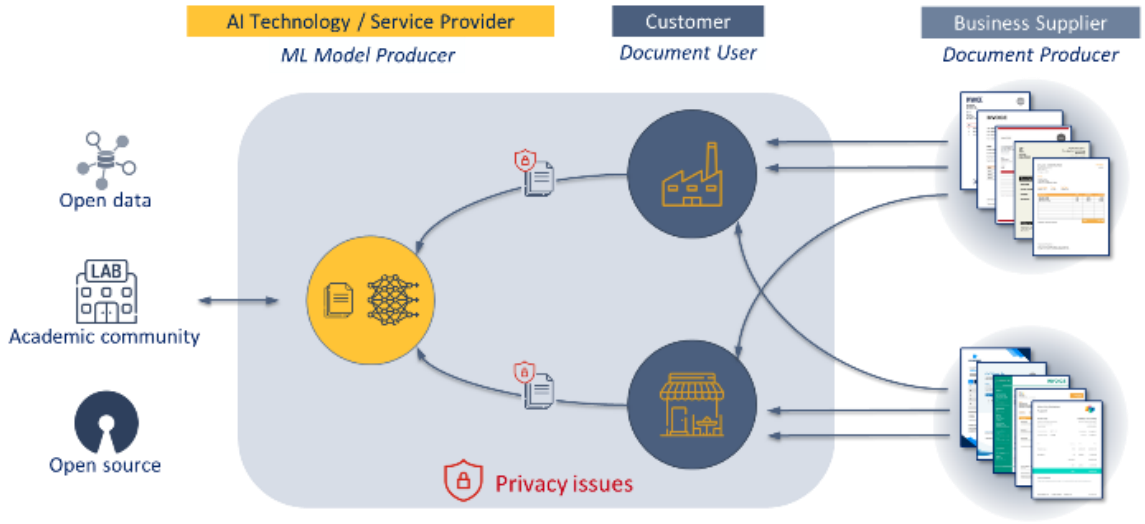
金融, 法律, 保険などの分野において特に使われうるタスクであり,
同時に文書には機密情報が含まれる可能性がある

DocVQAへの我々の取り組み

我々の差分プライバシーを考慮した連合学習手法を用いてDocVQAを学習

差分プライバシー(DP): 「**外部公開情報から、学習元となる個別データがどれだけ流出しやすいか**」 を定量的に評価

DocVQAの学習



https://sites.google.com/view/pfldocvqa-neurips-23/home

Date	Method	ANLS	Accuracy
2024-03-27	DP-CLGECL-momentum	0.6411	57.0347
2023-10-25	Differentially Private Federated Learning with LoRA	0.6066	53.5952
2023-10-27	DP-CLGECL	0.5925	51.9113
2023-10-24	(Baseline) FedAvg + DP Baseline	0.4996	43.3637

Ours

PFL-DocVQAといった安全性を考慮しFLを用いたDocVQAのコンペなども開催されている

2. 後半パートのまとめ

後半パートまとめ

01　医療支援

- 医療支援について
- 医療支援AIの課題
- 医療支援に関する最新の取り組み
- 胸部X線写真診察を用いた我々の取り組み

02　悪性通信検知

- 悪性通信検知について
- ネットワーク悪性検知への我々の取り組み

03　言語処理

- 言語処理について
- 言語処理に関する最新の取り組み
- DocVQAを用いた我々の取り組み


NTT


NTT DATA