令和4年度研究開発成果概要図 （目標・成果と今後の研究計画）

1．研究課題・受託者・研究開発期間・研究開発予算

◆研究開発課題名　プライバシー保護連合学習の高度化に関する研究開発
◆副題　継続実運用に資する不正取引モニタリングに向けたプライバシー保護連合学習の高度化
◆受託者　国立大学法人神戸大学、ＥＡＧＬＹＳ株式会社
◆研究開発期間　令和4年度～令和5年度（2年間）
◆研究開発予算（契約額）　令和4年度から令和5年度までの総額56百万円（令和4年度26百万円）

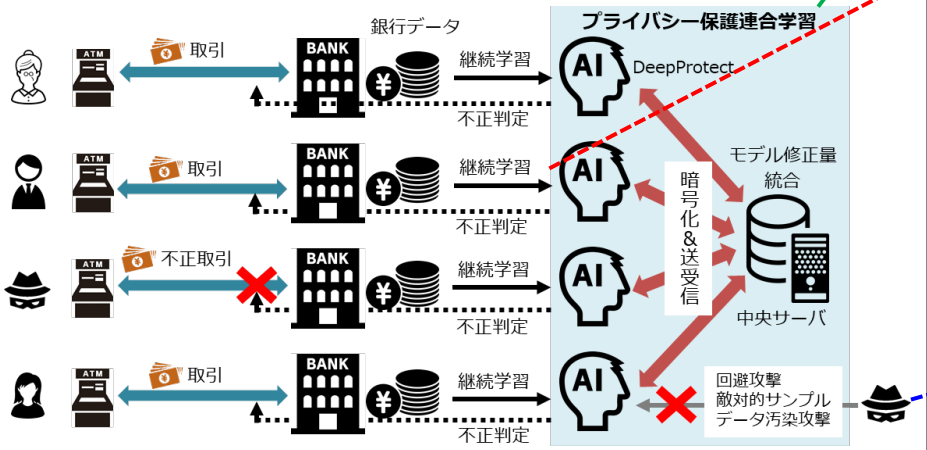
2．研究開発の目標

複数の銀行から提供される顧客口座データで求められる特徴量と犯罪フラグの共通化を行い、参加金融機関数に増減があっても、安定した継続学習を行える学習アルゴリズムの開発とそのロバスト性向上のための銀行取引履歴と口座情報の疑似データ生成アルゴリズムを提案する。さらに、不正送金検知におけるAIの回避攻撃とデータ汚染攻撃のシナリオを数種類想定し、その対策に関する先行研究を調査する。

3．研究開発の成果

研究開発項目1　DeepProtectの高度化に関する研究

プライバシー保護連合学習技術「DeepProtect」の実運用を想定した場合、安定した継続学習の実現と回避攻撃を意図した敵対的サンプルやデータ汚染攻撃への耐性向上が必要となる。そこで、これら2点についてDeepProtectの機能を高度化し、実運用に資する取引モニタリング共同システムを実現するプライバシー保護連合学習モデルに改良する。



研究開発項目1-1：実運用を模擬したテスト環境の開発と性能評価

顧客口座情報と口座取引を模擬する疑似データの生成方法を検討し、個人プロフィールに基づいた取引疑似データが生成可能であることを確認。

研究開発項目1-2：DeepProtectの継続学習化

DeepProtectを使ったシステムの長期運用を見据え、破壊的忘却を回避する継続学習と連合学習を同時に達成する学習アルゴリズムの調査検討と精度検証を実施。直近のデータのみで連合学習するモデルであっても、全期間のデータで連合学習するモデルに匹敵する性能を有するものがあることを確認。

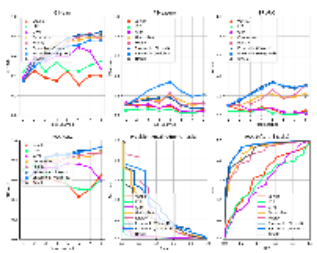


図1. 継続学習精度

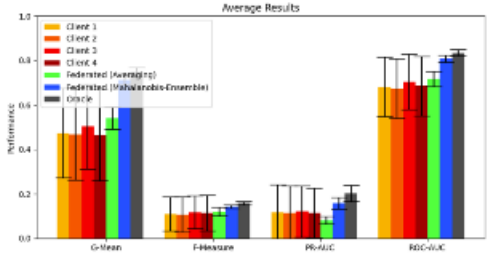


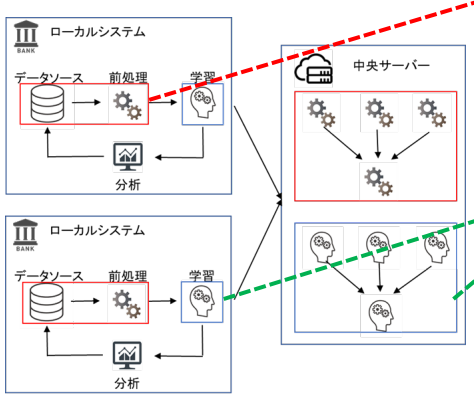
図2. 連合学習時の継続学習精度

研究開発項目1-3　DeepProtectの敵対的サンプルとデータ汚染攻撃への耐性向上

複数銀行が連合学習を行うシナリオにおいて、不正取引検知AIに対する敵対的サンプルの生成方法とデータ汚染方法を検討。

研究開発項目2　DeepProtectを用いた不正取引検知エンジンの開発

実運用を想定したDeepProtectの継続的な学習実施のためには、前処理や学習実施を行うシステムの安定的な稼働が必要不可欠であり、その要件が明確化される必要がある。そこで、銀行ー中央サーバー間でデータ前処理・学習のプロセスを完遂するシステムを開発し、これらのシステムを協力銀行データに合わせて擬似的に導入・運用することで本番運用に資するシステムの全体構成を明らかとする。



研究開発項目2-1：データパイプライン構築モジュールの開発

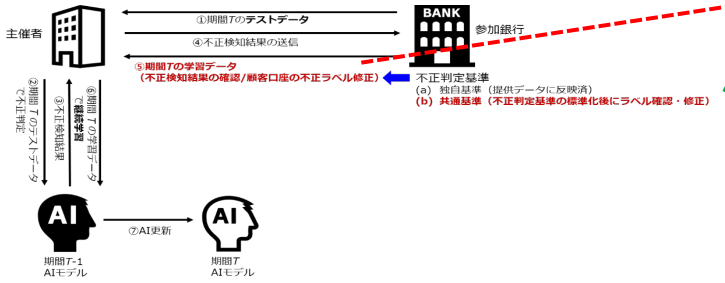
DeepProtectを実銀行で運用するために銀行勘定系システムから取引情報や顧客情報を加工し、機械学習システムへ送信するELTシステムの設計・開発を実施。DataVault2.0モデリング手法を用いて構築し、今後の開発・チューニング過程で生じうる新しい特徴量の追加や、銀行間のデータフォーマットの違いに対して柔軟に対応できるよう開発。

研究開発項目2-2：DeepProtect 学習実施モジュールの開発

数銀行間でDeepProtectによる連合学習を実施するモジュールを開発し、2-1のデータパイプラインモジュールとを接合。データパイプラインモジュールにおいて、銀行勘定系から出力される素データを日次で学習用フォーマットに成形し、学習実施モジュールが中央サーバーと通信しながらデータ正規化とDeepProtectの連合学習を実行する動作を確認。

研究開発項目3　継続実運用を想定した不正送金検知実証実験環境の整備

機構が提供する4銀行以上の取引・顧客口座データを、その取引時期に応じて、「初期学習フェーズ」と「テスト・継続学習フェーズ」に分け、継続実運用を想定した不正送金検知実証実験の環境を整備する。その際、継続学習を銀行業務に転用可能なオペレーションとするため、不正判定パターンの分析・可視化、不正判定フラグの修正・追加などを行うユーザインタフェースを開発し、低コストで高精度な不正検知を実現する有効特徴量の探索および不正判定基準の標準化などを容易にする支援ツールを備える。



研究開発項目3-1：不正送金検知特徴量と不正判定基準の標準化

データ提供先が決定されていないため、次年度において実施。

研究開発項目3-2：人間系フィードバックを容易とする支援ツールの開発

銀行内で継続的にDeepProtectを用いた不正検知や連合学習を行えるよう、不正送金監視者による運用を見据えて、人間系のフィードバックの取り込みを容易とする支援ツールの開発を実施した。本支援ツールとして、以下で構成されるシステムを開発した。

① 直近の検知結果を表示し、そのリンクを提供する「ダッシュボードページ」
② 疑わしい口座として検知された口座の取引推移、取引場所、その他の特徴などをグラフ・表で可視化し、結果の修正が可能な「口座詳細ページ」
③ 継続的に学習されるモデルやDeepProtect・個別学習など種々のモデルの検知成績を確認する「検知レポートページ」
④ 検知に使用するモデルを選択する「モデル管理ページ」

4. 特許出願、論文発表等、及びトピックス

国内出願	外国出願	研究論文	その他研究発表	標準化提案・採択	プレスリリース 報道	展示会	受賞・表彰
0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)

※成果数は累計件数、( )内は当該年度の件数です。

5. 今後の研究開発計画

R5年度では、実運用に資するDeepProtectの継続学習方式を実現し、複数の銀行から提供される顧客口座データを用いて、データ提供期間における不正検知の再現率が90%以上維持されることを実証する。また、銀行が提供するデータ項目から求められる特徴量と犯罪フラグの共通化を行い、参加金融機関数に増減があっても、安定した継続学習が実現可能であることを実証する。さらに、不正送金検知におけるAIの回避攻撃とデータ汚染攻撃のシナリオを数種類想定し、その対策を施しても、なお不正検知の再現率90%の目標が達成できるかを検証する。また、本委託研究で得られた研究成果は、参加銀行との秘密保持契約に違反しない範囲で学術雑誌への投稿、国際会議での発表、プレス発表などのかたちで公開する。具体的な実施項目を以下に示す。

研究開発項目1　DeepProtectの高度化に関する研究（4月～3月）

- **研究開発項目1-1**　実運用を模擬したテスト環境の開発と性能評価（4月～12月）
  令和4年で検討を行った疑似データ生成手法を実装し、銀行不正送金検知の実運用を模擬したテスト環境を開発する。
- **研究開発項目1-2**　DeepProtectの継続学習化（4月～3月）
  令和4年で検討したアルゴリズムに関して、協力銀行データへの適応を実施し、精度検証を実施
- **研究開発項目1-3**　DeepProtectの敵対的サンプルとデータ汚染攻撃への耐性向上（4月～3月）
  回避攻撃につながる敵対的サンプルとデータ汚染攻撃のシナリオを各1件以上ずつ検討し、耐性向上機能をDeepProtectの継続学習アルゴリズムに付加。模擬データおよび実データを使った評価実験を実施し、攻撃への耐性について定量的に評価する

研究開発項目2　DeepProtectを用いた不正取引検知エンジンの開発**（4月～3月）**

- **研究開発項目2-1**　データパイプライン構築モジュールの開発（4月～3月）
  令和4年で開発したモジュールをデータ提供銀行のシステムに合わせてチューニングを実施。導入時問題点の策定と改善を実施
- **研究開発項目2-2**　DeepProtect 学習実施モジュールの開発（4月～3月）
  令和4年で開発したモジュールをデータ提供銀行のデータに合わせて学習を実施。問題点の策定と改善を実施

研究開発項目3　継続実運用を想定した不正送金検知実証実験環境の整備（4月～3月）

- **研究開発項目3-1**　不正送金検知特徴量と不正判定基準の標準化（4月～3月）
  参加金融機関が提供するデータ属性値の標準化と不正判定基準の標準化を行う。
- **研究開発項目3-2**　人間系フィードバックを容易とする支援ツールの開発（4月～3月）
  令和4年度開発の支援ツールに関して、参加銀行に対するヒアリングを実施。ヒアリングに基づくフィードバック支援ツールの改良