令和5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6年度（3年間）
◆研究開発予算（契約額）　令和4年度から令和6年度までの総額56百万円（令和5年度16百万円）

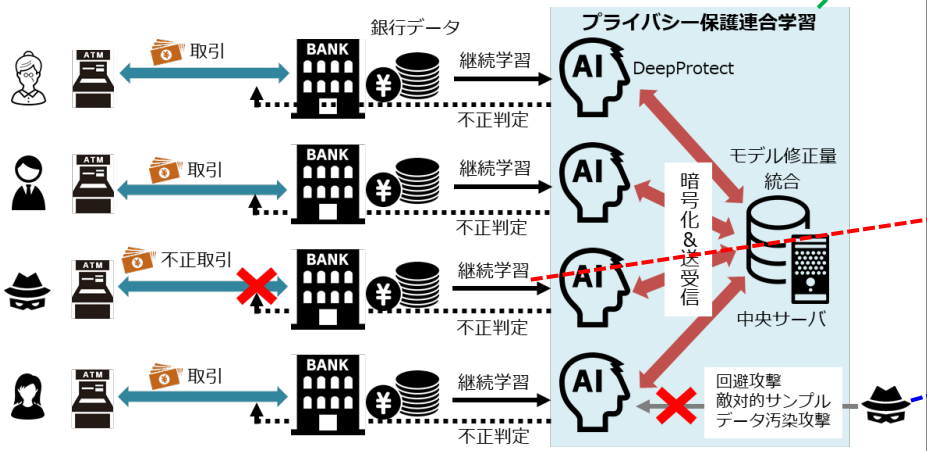
2．研究開発の目標

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

DeepProtectなどの連合学習AIによる検知を回避する攻撃への対策を行った上で、不正検知再現率90%以上を継続学習で達成するため、不正取引を模擬する疑似データ生成の手法を確立し、それを汎化性能の向上に利用できるテスト環境を開発する。また、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で、模擬データを使った評価実験を行い攻撃への耐性を評価する。



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

口座取引を模擬する時系列生成モデルを定式化し、モデルパラメータと口座の潜在変数（入出金特性）を推定して疑似データを生成する手法を開発。ベンチマークデータで動作検証し、多様な不正取引の生成を確認。

生成された疑似残高データ

研究開発項目1-2：DeepProtectの継続学習化

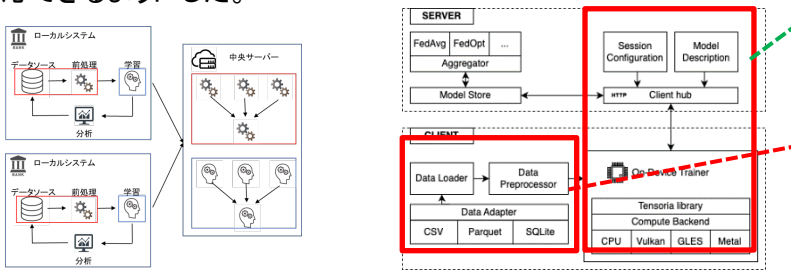
令和4年度に考案した連合学習・継続学習を同時に達成するアルゴリズムの詳細検証として、継続学習に関する種々の問題への精度検証を実施するとともに、対外発表を行うための精査を実施。また、銀行取引データに適用し検証を実施する前準備を実施。前準備として、データ提供された1銀行の取引データについて、特徴量の精査、分析モデルの構築・検証を実施。次年度においては、追加される他行のデータ解析を進め、継続学習の有効性を検証。

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

不正送金検知連合学習AIに対し、不正・不審取引を正常取引に誤判定させる回避攻撃が可能かをベンチマークデータにより検証。

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

実銀行環境への導入時に課題となりうる事項について、システムのチューニングを実施。ドッカー上にシステムをホストするバージョンに加え、ネイティブアプリとして提供可能なバージョンのを用意することで、様々な運用に対応できるようにした。



連合学習モジュールのシステムアーキテクチャ図
ネイティブアプリ化や計算フレームワークの改良により、種々の環境に対応可能なようにチューニングを進めた

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

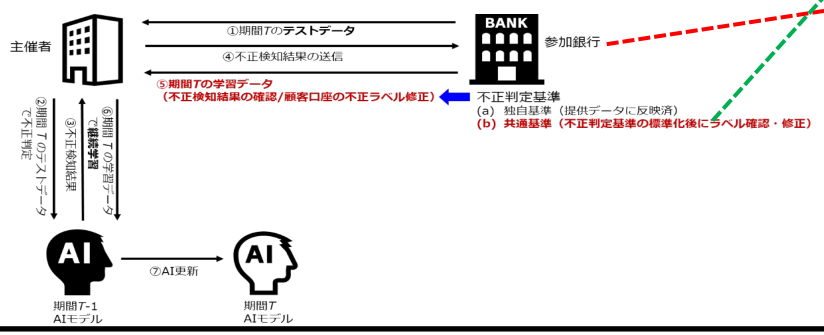
プロトタイプシステムを導入する銀行の運用体制に柔軟に対応するため、特徴量抽出を実施するETLモジュールのネイティブアプリケーション化を進め、種々の運用体制に対応可能となるように改良を実施した。

研究開発項目2-2：DeepProtect 学習実施モジュールの開発

ローカル環境に用意される種々の計算機に対応できるように、特定の計算環境に依存しないように改良を実施した。これにより、様々な計算環境を有する参加者出会っても、安定した連合学習の実施が可能となると考えられる。

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

NICTが指定する4行以上の銀行において、日々処理される顧客取引データを継続的に正常・不正の判定を実行し、疑わしい取引と判定されたものについては、その犯罪性の有無を取引パターンや顧客情報などから判定する。令和4年度に実施したデータ属性と不正ラベルの標準化を令和5年度でも継続し、不正判定基準に関する情報共有を支援するツールのプロトタイプ設計と開発を行う。

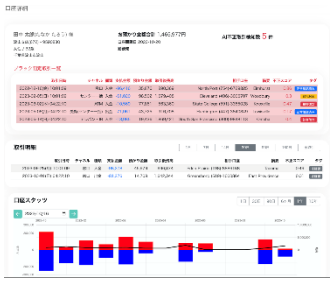


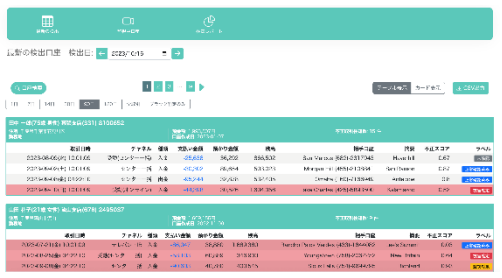
研究開発項目3-1：不正送金検知特徴量と不正判定基準の標準化

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

参加銀行にプロトタイプシステムの説明およびヒアリングが実施できるよう、説明用動画の撮影、ヒアリング項目の取りまとめを実施。開発したアプリケーションのネイティブアプリケーション化を実施。





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)	3 (2)	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

令和6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可能な限り行い、データ項目が完全に一定していなくても、疑似データ生成による欠損補完を行い、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。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～9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～9月）
  令和5年度に開発した状態空間モデルを使った疑似データ生成の性能評価を銀行から提供された実データで行い、実データに近い疑似データが得られるよう改善を行う。これにより、銀行不正送金検知の実運用を模擬したテスト環境を完成する。
  研究開発項目1-2　DeepProtectの継続学習化（4月～9月）
  令和5年度に提供された銀行データに加え、追加される他行のデータ解析を進め、開発アルゴリズムに関する継続学習の検証を行い、銀行取引データに関する連合学習システムの継続学習化の検証を行う。
  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～9月）
  連合学習AIに対して、不正・不審取引の検知回避につながる敵対的サンプル生成のアルゴリズムを完成し、その攻撃に対する耐性を有するDeepProtectの継続学習アルゴリズムを完成する。また、銀行から提供された実データを使った性能評価実験を実施し、攻撃への耐性能力を定量的に評価する。

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～9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～9月）
  開発したモジュールの導入に関するヒアリングを実施し、参加銀行の環境に合わせたシステムチューニングを実施する。
  研究開発項目2-2　DeepProtect 学習実施モジュールの開発（4月～9月）
  開発したモジュールの導入に関するヒアリングを実施し、参加銀行の環境に合わせたシステムチューニングを実施する。

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～9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～9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～9月）
  開発した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を行う。