令和6年度研究開発成果概要図 （目標・成果と今後の成果展開）

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6年度14百万円）

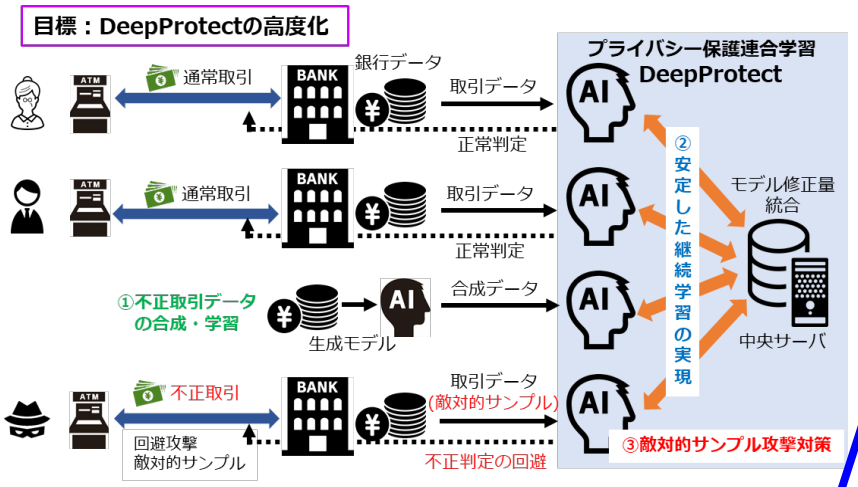
2．研究開発の目標

DeepProtectの継続学習を実現し、４行以上の銀行から提供されるデータに対して、データ項目と不正判定基準の標準化を行い、不正検知の再現率が90%以上維持されることを目標とする。さらに、DeepProtectの高度化として、不正送金検知におけるAIの回避攻撃を想定し、銀行口座の実取引に対して攻撃が成立することを実証する。さらに、銀行業務に転用可能な継続学習オペレーションとするため、判定分析・可視化、判定フラグ修正などを容易にするフロントエンド開発し、銀行ローカルシステムと連携し、不正検知モニタリングモデルを連合学習で学習するプロトタイプシステムを開発。

3．研究開発の成果

① DeepProtectの高度化

取引モニタリングシステム構築に向けたDeepProtectの高度化

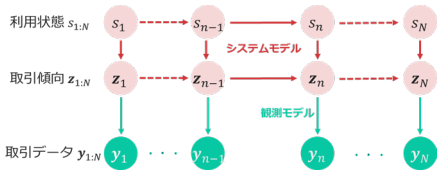


① 不正取引データの合成による不均衡データ問題の緩和
② 安定した継続学習の実現
③ AIの不正判定を回避する敵対的サンプル攻撃への防御

研究開発成果：①不均衡データ問題の緩和

不正取引は通常取引に比べて極めて少ない → 機械学習における**不均衡データ問題**

- 状態空間モデルを応用して不正取引データを合成し、これを学習に使って不均衡データ問題を緩和
- 人工データを状態空間モデルで学習し、合成データを訓練データに加えて、DeepProtect の性能が大幅改善。



研究開発成果：②安定した継続学習の実現

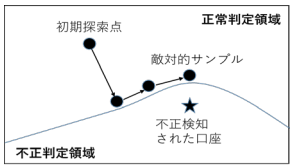
不正取引モニタリングシステムの安定した長期運用に必須

- DeepProtectのマハラノビスアンサンブルモデルによる破滅的忘却を回避
- 6つの分析アルゴリズムの徹底的な比較から生まれた、学習効率のよく、シンプルな継続学習アルゴリズム

研究開発成果：③ 敵対的サンプル攻撃への防御

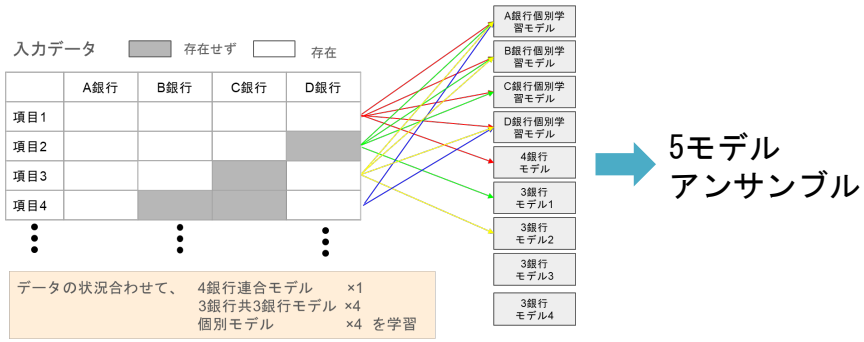
不正検知AIの判定を回避させる摂動を探索

- 状態空間モデルによる不正取引データの合成と敵対的攻撃アルゴリズムを組み合わせて、AIに検知されない不正取引を探索



② DeepProtectによる高精度不正検知モニタリング

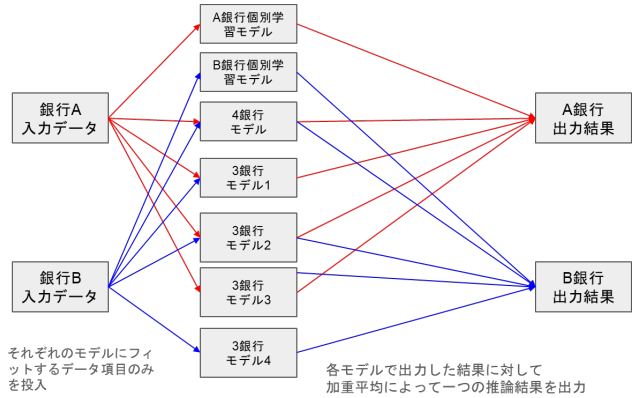
目標： ４銀行の連合学習で不正取引を再現率90%以上で検知



5モデル
アンサンブル

アンサンブルで銀行間データ項目のばらつきを標準化して高性能化を目指す

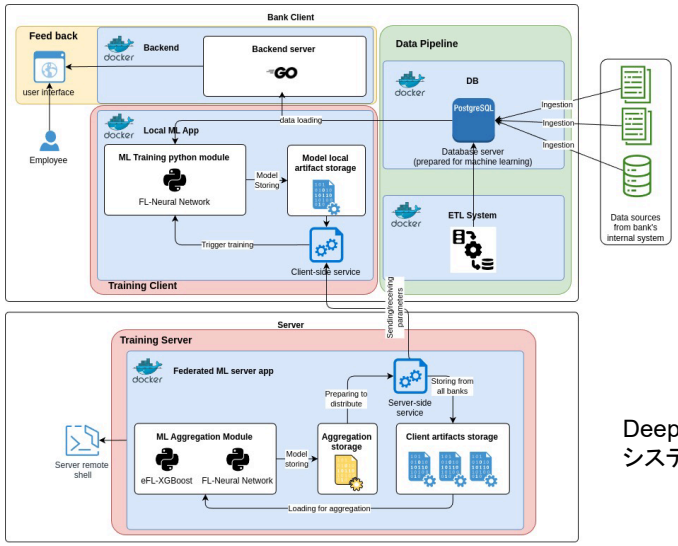
研究開発成果：4銀行に対し不正取引検知を再現率90%以上達成



特徴量の標準化と不正検知の高性能化を同時に実現！

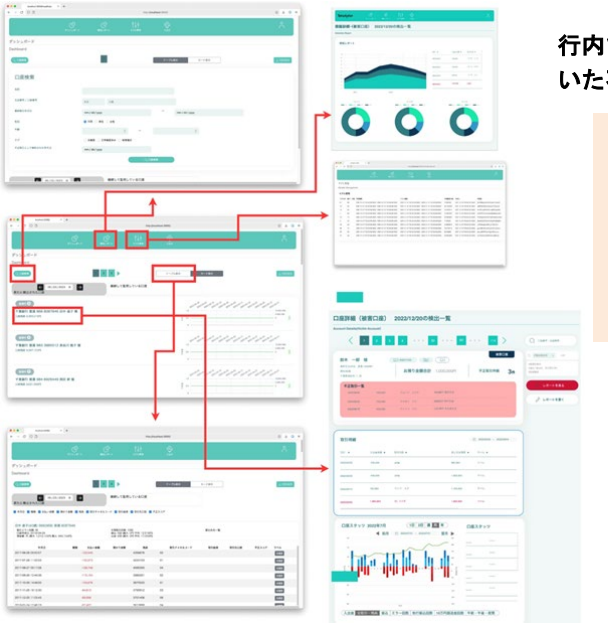
③継続実運用に資するDeepProtectプロトタイプシステム

- 銀行における不正検知モニタリングの実施現場で学習実施を行うデータパイプラインと学習・推論を行うプロトタイプシステムを開発
- 銀行ローカルシステムと連携し、不正検知モニタリングモデルを連合学習で学習するプロトタイプシステムを開発



DeepProtectプロトタイプシステムの全体構成図

研究開発成果：不正取引監視者を支援するフロントエンドを開発



行内で継続的にDeepProtectを用いた不正検知や連合学習が実現

フロントエンドシステム
①「ダッシュボードページ」
②「口座詳細ページ」
③「検知レポートページ」
④「モデル管理ページ」
⑤「入出力」ページ

銀行間での不正判定基準の標準化

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)	9 (5)	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。

■ ～高度通信・放送研究開発委託研究～ 「プライバシー保護連合学習の高度化に関する研究開発」

『連合学習ワークショップ』

第1回 2024年9月26日、オンライン

第2回 2025年2月28日、オンライン

概要：プライバシー保護連合学習モデルDeepProtectを用いた不正送金検知エンジンを開発し，4銀行（りそな銀行、他3行）が互いにデータを直接共有しなくても，連合学習により高精度な不正取引検知が可能かを検証する。

5. 研究開発成果の展開・普及等に向けた計画・展望

1）計画

DeepProtectは不正取引のデータ量が限られていても機械学習ベースの不正検知モニタリングが実施可能という利点があり、研究開発開始時には地銀など小規模な銀行におけるAIの導入への期待は大きかった。しかし、そこから2年以上が経過し、システムベンダーによるAIモニタリングサービスの導入検討が進んでいる。そこで、小規模な組織が多い信用金庫を初期顧客ターゲットとして、以下の活動を通じて本研究成果の社会導入を行う予定である。

（a）フェーズ1：プロダクトマーケットフィットを目的としたPOCフェーズ （～2026年度末）

信用金庫をターゲットとした、不正取引モニタリングの状況調査、および現場で生じるニーズのヒアリングを実施する。ヒアリングを通じて、本研究開発で開発したプロトタイプの紹介を行い、POCの実施提案を行う。2026年末までに5件のPOC実施を目標とする。POCにおける運用を通じて、不正検知モニタリングサービス、およびシステム要件整理を行う。またPOC実施金融機関で勘定系システムや既存のモニタリングシステムを提供するシステムベンダーにヒアリングを実施し、既存のモニタリングシステムへの組み込みを模索することで、ビジネスモデル仮説を構築する。

（b）フェーズ2：商用プロダクト開発 （～2027年度末）

POCを通じて得られた要件をもとに、サービス提供として必要な機能要件、非機能要件を元としたプロダクト開発を行う。プロダクトは他のシステムと独立して不正検知モニタリングを行うシステム開発を主として実施するとともに、データパイプライン・学習実施モジュールを既存のモニタリングシステムに組み込むことが可能なバージョンの2種類のシステム開発を開発候補とし、開発内容はPOCフェーズの活動を通じて決定する。

（c）フェーズ3：プロダクト展開 （～2029年度末）

プロダクトの展開を行うためにチャネル開拓を実施する。チャネル開拓の活動候補は、プロダクト開発内容に応じて実施する。