様式 1-4-2 (2022-1)

令和４年度研究開発成果概要書

採 択 番 号　　22901
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

（1）研究開発の目的

複数の銀行から提供される顧客データ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。

（2）研究開発期間

令和４年度から令和５年度（２年間）

（3）受託者

国立大学法人神戸大学＜代表研究者＞
ＥＡＧＬＹＳ株式会社

（4）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）

（5）研究開発項目と担当

研究開発項目１　DeepProtect の高度化に関する研究
1-1　実運用を模擬したテスト環境の開発と性能評価（神戸大学）
1-2　DeepProtect の継続学習化（ＥＡＧＬＹＳ）
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（神戸大学）

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
2-1　データパイプライン構築モジュールの開発（ＥＡＧＬＹＳ）
2-2　DeepProtect 学習実施モジュールの開発（ＥＡＧＬＹＳ）

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
3-1　不正送金検知特徴量と不正判定基準の標準化（神戸大学）
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（ＥＡＧＬＹＳ）

（6）特許出願、外部発表等

		累計（件）	当該年度（件）
特許出願	国内出願	0	0
	外国出願	0	0
外部発表等	研究論文	0	0
	その他研究発表	0	0
	標準化提案・採択	0	0
	プレスリリース・報道	0	0
	展示会	0	0
	受賞・表彰	0	0

（7）具体的な実施内容と成果

研究開発項目１　DeepProtect の高度化に関する研究

1-1. DeepProtect の高度化を効率的に行うテスト環境を構築するため、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た。項目データと時系列データを生成できる Python ライブラリーである SDV（The Synthetic Data Vault）を用い、生成された疑似口座情報に基づいて、その個人プロフィールに基づいた取引疑似データが生成されることを確認した。但し、実際の顧客取引データと比較して不自然な取引もあり、次年度において、疑似データ生成方法の改善を継続する。

1-2. 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した。銀行取引のオープンデータセットを用いて擬似データを生成し、犯罪手口が変容するシナリオで検証を行い、直近のデータのみで連合学習するモデルを提案した。全期間のデータで連合学習するモデルと比較した結果、前者の提案モデルは後者と同等の性能を有しながらも、破滅的忘却を抑制しながら、安定した継続学習が可能であることが示された。

1-3. AI に対する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方法および対策について調査し、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した。しかしながら、DeepProtect コードと銀行データの提供が遅れており、攻撃への耐性評価実験は次年度に実施することとした。

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

2-1. 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・開発を実施した。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した。

2-2. 複数銀行間で 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の接合を行った。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確認を行った。

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

3-1　高精度な不正送金検知を行うため、顧客口座と口座取引の特徴量、さらに不正判定基準の標準化を試みる予定であったが、データ提供先が決定されていないため、本項目は実施せず、次年度において実施することとした。

3-2. 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」

（8）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年で検討した連合学習アルゴリズムを協力銀行データに適用し、精度検証を実施する。

・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

・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