プレスリリース
2022 年 3 月 10 日

国立研究開発法人情報通信研究機構
国立大学法人神戸大学
株式会社エルテス

プライバシー保護連合学習技術を活用した不正送金検知の実証実験を実施

～被害取引の検知精度向上や不正口座の早期検知を確認～

【ポイント】

- 銀行 5 行と不正送金検知の実証実験を実施し、被害取引の検知精度向上や不正口座の早期検知を確認
- プライバシー保護連合学習技術「DeepProtect」を活用し、取引データを互いに開示することなく不正送金の検知精度を向上
- 今後も、より高い検知精度を達成するために実証実験を継続し、金融機関での実運用を目指す

国立研究開発法人情報通信研究機構（ＮＩＣＴ（エヌアイシーティ）、理事長: 徳田 英幸）サイバーセキュリティ研究所セキュリティ基盤研究室、国立大学法人神戸大学（学長: 藤澤 正人）及び株式会社エルテス（代表取締役: 菅原 貴弘）は、プライバシー保護連合学習技術[*1]「DeepProtect」[*2] 等を利用し、千葉銀行、三菱 UFJ 銀行、中国銀行、三井住友信託銀行及び伊予銀行と連携して、不正送金検知の実証実験を実施し、目標としていた不正送金の検知精度 80%以上を達成するとともに、一銀行では検知できなかった不正送金の被害に遭った取引（被害取引）の検知や、不正送金に悪用された口座（不正口座）の早期検知を確認しました。

今回は「被害取引の検知」と「不正口座の検知」という二つの目的に対し、複数銀行が連合して「DeepProtect」を活用した学習モデルを構築し、不正検知をする実証実験を行いました。その結果、目標としていた検知精度 80%以上を達成するとともに、前者では、一銀行では検知できなかった不正取引が検知されるケースが確認され、後者では、不正口座として実際に凍結されるより大幅に早く検知することに成功しました。

本実証実験で得られた成果と課題を踏まえ、「DeepProtect」等で検知精度を更に向上させ、銀行における不正送金検知業務に適用するための実証実験を進め、不正送金検知システムの社会実装を目指します。

【背景】

マネー・ローンダリング、不正送金、振り込め詐欺などの金融犯罪手法は複雑化・巧妙化しており、早急な対策が求められています。中でも、振り込め詐欺等の特殊詐欺による2021年の全国の被害金額は278億円を超え（暫定値、警察庁発表）、依然として深刻な社会問題[*3]となっています。そこで、AIを用いた不正取引の自動検知システムの導入が検討されていますが、単独の金融機関では十分な量の学習データを用意することが難しく、また、個人情報を含む金融取引データを各金融機関外に持ち出すことができないため、複数の金融機関で協力して学習することもできず、同システムの普及は進んでいませんでした。

このような状況下で、NICT、神戸大学及びエルテスは、データを外部に開示することなく機密性を保ったまま機械学習[*1]を行うNICT独自開発のプライバシー保護連合学習技術「DeepProtect」等を活用し、複数の金融機関（千葉銀行、三菱UFJ銀行、中国銀行、三井住友信託銀行、伊予銀行）と連携して不正送金等を自動検知するシステムの実現を目指し、実証実験に取り組んできました。

【今回の成果】

各銀行の検知目的に応じて、被害取引の検知を目的とする「被害検知グループ」（図 1 参照）と、不正口座の検知を目的とする「加害検知グループ」（図 3 参照）とで、それぞれ実証実験を行いました。

被害検知グループの実証実験には 2 行の銀行が参加し、「単独組織のデータのみを用いた通常の機械学習モデル（個別学習モデル）による検知精度」と「2 行のデータを用いた DeepProtect モデル（連合学習モデル）による検知精度」の比較を行いました。図 2 の例に示すように、連合学習モデルにより検知精度が向上し、1 日当たりの被害取引のアラート件数を 600 件としたときの検知精度は、当初目標としていた 80%以上を達成しました。また、個別学習モデルでは検知できなかった不正取引が検知された事例も確認しました。なお、検知率は、検知漏れの少なさを表す指標である再現率で示しています。

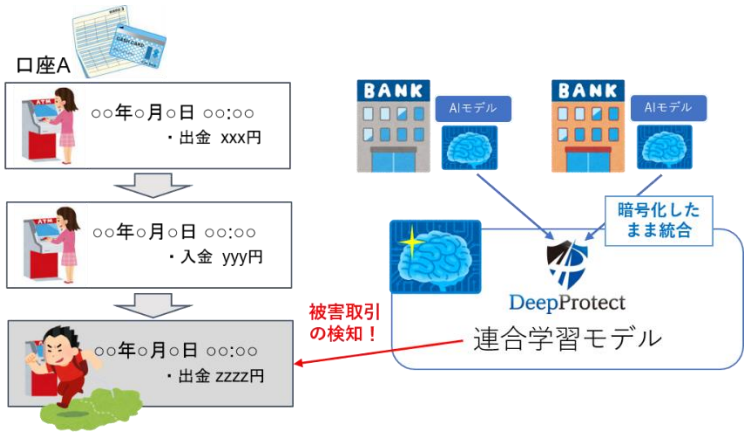


図 1　被害取引の検知

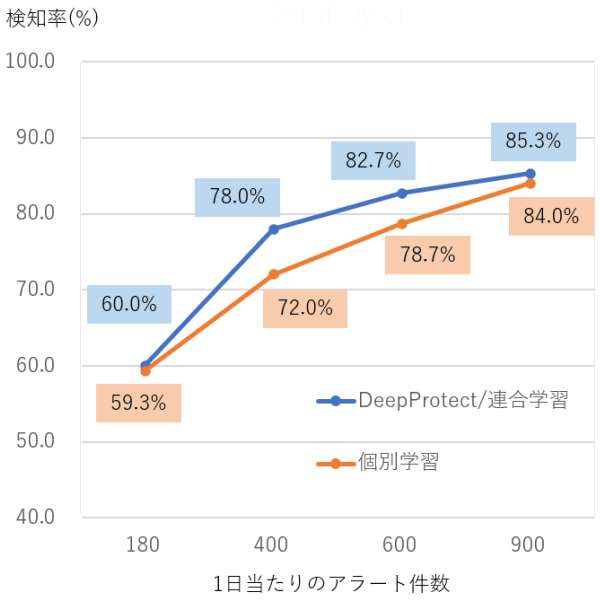


図 2　被害検知グループの検知率（一例）
（実データを基にした検知結果を銀行の許可を得て表示）

加害検知グループの実証実験には 4 行の銀行が参加し、「通常取引に使われていた口座が、あるタイミングから犯罪に悪用され、取引停止・凍結されたもの」を不正口座と定義し、不正口座を検知することとしました。個別学習モデルとハイブリッドモデル（個別学習モデルと連合学習モデルの組合せ）で比較し、モデルの性能を検知率に加えて、不正口座に対して凍結時点より何週前に検知できるかという検知タイミングで評価しました。その結果、図 4 の例に示すように、個別学習モデルと比較してハイブリッドモデルが高い性能を示し、検知率 80%以上を達成しました。さらに、実データでの不正口座凍結よりも 20～50 週程度の早期検知が可能であることが確認されました。

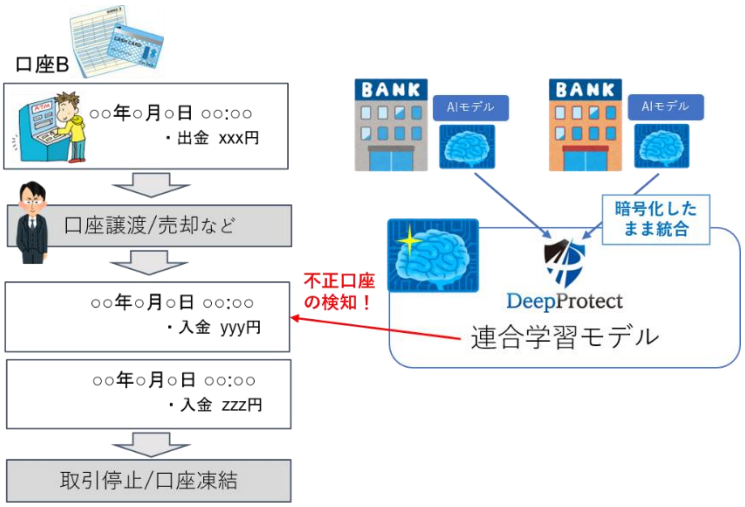


図 3　不正口座の検知

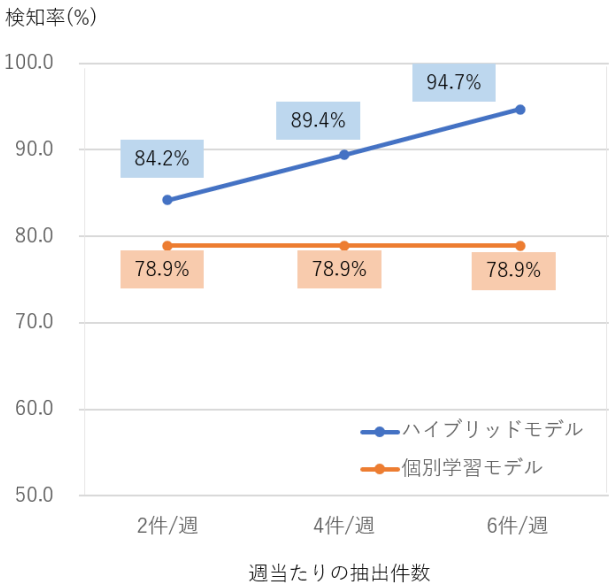


図 4　加害検知グループの検知率（一例）
（実データを基にした検知結果を銀行の許可を得て表示）

【今後の予定】

本実証実験は、2019 年度から JST CREST「イノベーション創発に資する人工知能基盤技術の創出と統合化」の加速フェーズ研究課題として採択された研究課題「プライバシー保護データ解析技術の社会実装」（課題番号 JPMJCR19F6）の下で実施されてきましたが、2022 年度から開始される JST AIP 加速課題 研究課題「秘匿計算による安全な組織間データ連携技術の社会実装」として採択され、今後は、銀行における不正送金業務への実運用に向け、更なる検知性能の向上やシステム実装に取り組みます。

＜関連する過去のプレスリリース＞

・2019 年 2 月 1 日 プライバシー保護深層学習技術で 不正送金の検知精度向上に向けた実証実験を開始
https://www.nict.go.jp/press/2019/02/01-2.html

・2020 年 5 月 19 日 プライバシー保護深層学習技術を活用した不正送金検知の実証実験において金融機関 5 行との連携を開始
https://www.nict.go.jp/press/2020/05/19-1.html

＜用語解説＞

*1 プライバシー保護連合学習技術

データ自体を一か所に集約せず、分散した状態で連合して機械学習（AI の一種で、あるデータの中からコンピューターが一定の規則を発見し、その規則に基づいて未知のデータに対する推測・予測等を実現する技術）を行う技術であり、データを持つ複数の法人や個人がそれぞれ独自に機械学習を行い、学習結果の一部の情報のみを集約することによって学習済みモデルを更新することができる。あたかもデータを一か所に集約して機械学習を適用したような効果を安全に得られる技術として期待が集まっている。

*2 DeepProtect

連合学習技術に暗号技術を融合することによって、NICT が独自に開発したプライバシー保護連合学習技術である。まず、各組織で持つデータを基に深層学習を行う際に、学習中のパラメータ（勾配情報）を暗号化して中央サーバに送り、中央サーバでは、暗号化したまま学習モデルのパラメータ（重み）の更新を行う。次に、更新されたこの学習モデルのパラメータを各組織においてダウンロードすることで、より精度の高い分析が可能になる。DeepProtect は、各組織から中央サーバにデータそのものを送ることなく、学習中のパラメータのみを暗号化して送信するが、このパラメータは複数のデータを集計した統計情報とすることによって個人を識別できない状態にすることが可能であり、さらに、暗号化を施すため、データの外部への漏えいを防ぐことができる。本技術により、パーソナルデータのような機密性の高いデータを外部に開示することなく、複数組織で連携して多くのデータを基にした機械学習が可能となる。

本技術は、下記ジャーナルに採択・掲載されている。

L. T. Phong, Y. Aono, T. Hayashi, L. Wang, and S. Moriai, "Privacy-Preserving Deep Learning via Additively Homomorphic Encryption", IEEE Transactions on Information Forensics and Security, Vol.13, No.5, pp.1333-1345, 2018.
L. T. Phong and T. T. Phuong, "Privacy-Preserving Deep Learning via Weight Transmission", IEEE Transactions on Information Forensics and Security, Vol.14, No.11, pp 3003-3015, 2019

*3 深刻な社会問題

不正送金等の金融犯罪に対して、現状、多くの金融機関は、それぞれが保有する金融取引データに対し、ルールベースのモニタリングツールを用いて、人手で不正取引を検出しているが、これには担当者の経験等への依存やコストの課題がある。先般公表された FATF（金融活動作業部会）による第 4 次対日相互審査報告書においても、我が国の金融機関で導入されている取引モニタリングシステムでは、誤検知の平均比率は最大 99%に上っており、不十分であることが指摘されている。

「金融庁 FATF（金融活動作業部会）による第 4 次対日相互審査報告書の公表について」
https://www.fsa.go.jp/inter/etc/20210830/20210830.html

日本が優先して取り組むべき行動として、報告書に以下が記載されている。

「リスクベースでの AML/CFT 監督を強化する。これには、特定事業者において実施されている予防的措置の評価のためのオフサイト・モニタリングとオンサイト検査の組み合わせについて、その頻度及び包括性を強化することや、金融機関、 DNFBPs、暗号資産交換業者による義務履行における肯定的な効果を確保するために、抑止力のある行政処分と是正措置が適用されることを含む。」

補足資料

加害検知グループの不正口座検知実証実験の方法

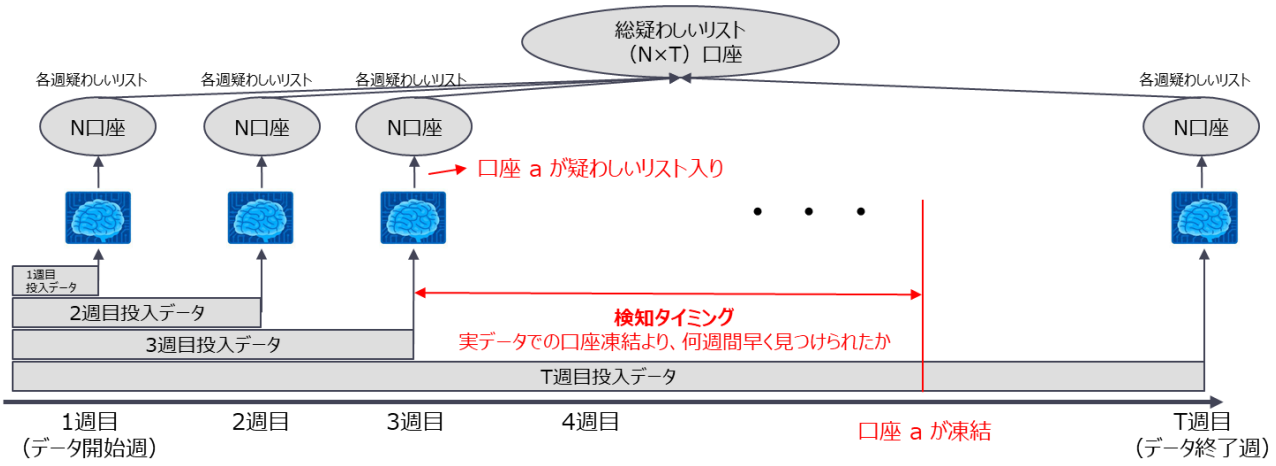
加害検知グループの不正口座検知実証実験では、取引停止や凍結に至った口座を不正案件として、以下の基準に基づき、性能評価を行いました。

① 不正口座と正しく検知できた率（再現率）

② 不正口座と正しく検知できたとき、取引停止・凍結時点よりも早期検知した時点（検知タイミング）

本実証実験では、週次で口座の不正判定を行うこととし、個別学習モデルや連合学習モデル（DeepProtect）が出した口座ごとの不正確信度に基づいて、最も疑わしい取引を行っている N 口座を週ごとに抽出して、上記 2 つの基準で性能を評価しました。

以下の図を使って、具体的に説明します。取引データの開始週で全ての口座に対して特徴量を作成して、個別学習モデルや連合学習モデル（DeepProtect）に判定させます。このとき、個別学習モデルや連合学習モデル（DeepProtect）が各口座に対して出力する不正確信度の大きいものから N 個の口座を抽出し、「疑わしいリスト」に入れます。この処理を第 2 週目以降、最後の T 週目まで続けます。①再現率は、取引停止・凍結された口座の全体に対する（N×T）口座に取引停止・凍結口座が含まれる率で求めます。②検知タイミングは、口座 a が 3 週目に「疑わしいリスト」に含まれたとして、実際に口座凍結された時点より、何週間早く見つけられたかで評価します。



＜ 本件に関する問合せ先 ＞

国立研究開発法人情報通信研究機構
サイバーセキュリティ研究所
セキュリティ基盤研究室
野島　良
E-mail: crest-ppdm-info@ml.nict.go.jp

国立大学法人神戸大学
数理・データサイエンスセンター
大学院　工学研究科　電気電子工学専攻
教授　小澤　誠一
Tel: 078-803-6466
E-mail: ozawasei@kobe-u.ac.jp

株式会社エルテス
DX 推進
部長　近藤　浩行
Tel: 03-6550-9280
E-mail: hiroyuki-kondo@eltes.co.jp

＜ 広報 （取材受付） ＞

国立研究開発法人情報通信研究機構
広報部　報道室
Tel: 042-327-6923
E-mail: publicity@nict.go.jp

国立大学法人神戸大学
総務部　広報課
Tel: 078-803-5453
E-mail: ppr-kouhoushitsu@office.kobe-u.ac.jp

株式会社エルテス
広報事業推進部
Tel: 03-6550-9280
E-mail: pr@eltes.co.jp