2025 法遵科技與電腦稽核專題競賽企劃書

聯邦式學習與圖神經網路的結合應用
——以反洗錢為例

中華民國114年12月06日

目錄

壹、前言

一、作品主題

本研究運用聯邦式學習（Federated Learning, FL）和圖神經網路（Graph Neural Networks, GNN）技術，建立新型反洗錢偵測模型，旨在克服反洗錢模型在數據隱私和數據共享的挑戰。基於洗錢行為的網絡特徵，本研究使用GNN建構交易圖，提取潛在的交易模式，以提高模型在偵測複雜洗錢行為中的準確性。而為了保護數據隱私，採用聯邦式學習技術，在保護數據隱私的同時實現數據共享。

經過測試，本模型的預測準確度上較單一客戶端訓練的模型提升逾20%，並相較傳統集中式訓練模型有更優異的表現，成功達成提升預測精準度與保護數據隱私的目標。根據研究結果，本研究證明FL和GNN整合應用的可行性，能在反洗錢模型中突破傳統技術限制，提供應對洗錢威脅的解決方案。

二、研究的背景和動機、目的

在當前金融體系中，洗錢活動已成為全球金融犯罪的一大挑戰，特別是在金融技術快速發展的背景下，洗錢手法變得更加隱蔽和複雜。根據國際貨幣基金組織（IMF）的估計，全球每年因洗錢而流失的金額約佔全球GDP的2%至5%，即約8000億至2兆美元[1]。並且隨著金融交易數量和複雜度的增加，舊有的基於規則引擎和監管報告的傳統反洗錢方法，其有效性逐漸下降，難以應對日益增多且多變的洗錢行為和各種欺詐手段[2]。

因此，當前的反洗錢趨勢逐漸向數位化與智能化發展，越來越多的金融機構依賴大數據和人工智慧技術來提升偵測效率[3]。智慧機器人、模型訓練及自動化工具被廣泛應用於交易監測系統，這些技術能夠在交易中自動挖掘異常模式和隱蔽的交易路徑，減少人力需求並加快反應速度[4]。然而，當前的反洗錢工具面臨許多挑戰，包括大量數據的處理以及資料不平衡的影響[5]，其中資料

不平衡的問題極大地影響模型的偵測精度, 因為洗錢樣本在所有交易中的比例極低, 導致訓練出的模型容易偏向正常樣本, 難以偵測到異常樣本。另外一方面, 反洗錢模型的訓練需要大量的交易數據來提升偵測的精準度, 但是由於原始數據常包含敏感的用戶資訊, 多數金融機構不願也無法共享數據, 造成所謂的“數據孤島”現象, 此困境限制了可用的數據量, 並進而影響模型的偵測精準度[6]。

為應對上述挑戰, 本研究結合了圖神經網絡 (Graph Neural Network, GNN) 與聯邦式學習 (Federated Learning, FL) 技術去建構新型反洗錢模型, 目的在於提高洗錢行為的偵測效果。圖神經網路特別適用於金融反洗錢, 因為其能夠根據金融交易的網絡結構, 如多方之間的轉帳關係等, 去自動學習異常模式, 增加模型對複雜交易路徑的理解與適應能力, 捕捉複數交易間的模式與關聯性。而透過與聯邦式學習技術的結合, 本研究支持跨機構的模型共享, 允許多個金融機構在不共享原始數據的情況下進行模型協作訓練, 在優化偵測效果的同時保護數據隱私, 增加金融機構合作的可能性。此外本研究針對原始資料採用了下採樣技術, 以解決資料不平衡的問題, 從而改善模型因不平衡數據集而產生偏誤的問題。

三、作品特色

本研究在應對金融體系中日益嚴峻的洗錢挑戰時, 提出使用圖神經網絡（GNN）和聯邦式學習（FL）的創新應用來偵測異常交易行為, 本專題有以下特色：

- 提高洗錢偵測準確性:本研究利用GNN捕捉複雜交易之間的模式與關聯性, 使模型能夠自動學習並識別異常交易, 有助於揭示傳統模型難覺察的隱蔽行為, 提高偵測的準確性和靈敏度。
- 數據隱私保護和跨機構協作:透過將GNN與FL技術結合, 本研究允許多家金融機構在不共享原始數據的前提下進行協同模型訓練, 解決數據孤島問題, 並在保護數據隱私的同時提高了模型的偵測效果。透過跨機構合作增強模型的泛化能力與協同偵測能力, 從而提升整體的反洗錢效能。

- 資料不平衡的處理:針對少數異常樣本比例極低的狀況, 本研究採用了下採樣技術來平衡數據分佈, 使模型能更準確地識別出少數類別樣本, 減少模型在不平衡數據上的偏誤, 同時提高對異常交易的敏感度, 增強模型的穩定性和適用性。

總的來說, 本研究為當前金融體系提供了一個新穎且有效的洗錢偵測框架, 藉由GNN和FL的結合應用, 本研究在兼顧數據隱私的情況下實現了跨機構協作, 並有效提升了對少數異常樣本的識別能力, 為金融監控體系提供了一個靈活且具擴展性的反洗錢模型的訓練方案。這些貢獻不僅能降低洗錢帶來的金融風險, 還有助於維護金融市場的穩定性和公信力。

貳、成果

一、摘要

本研究中進行了兩項主要實驗，以驗證聯邦學習（Federated Learning, FL）結合圖神經網絡（Graph Neural Network, GNN）對於異常交易偵測的效能提升情況。此外本研究在資料預處理階段引入下採樣技術以平衡資料，使模型能更有效地識別少數類別。以下將分別對資料不平衡處理、單一客戶端比較、集中式模型比較這三項實驗的結果進行分析。

首先是資料不平衡處理實驗。經測試，在處理後的數據集中，模型的精確率提升約374%，召回率提升約 150%，而F1分數提升了約 78%。這些結果表明下採樣技術在處理資料不平衡問題上具有顯著的效果，有效增強了模型對異常交易的辨識能力，使其在實際應用中能更準確地識別潛在異常行為。

第二個實驗比較單一客戶端模型與聯邦學習模型的偵測能力。本實驗中，客戶端各自基於本地數據進行訓練，並得到準確率、精確率、召回率和F1分數，接著將各自的結果與FL模型的全局性能進行對比。根據實驗數據，FL模型的全局性能相較於單一客戶端模型有顯著提升。準確率和召回率均提升了約 22%，精確率提升了約 23%，而 F1 分數提升了約 30%。這些提升表明，通過 FL 模型的協同學習，系統能夠有效地綜合不同客戶端的數據特徵，從而提高的偵測能力。

第三個實驗則將 FL + GNN 模型與集中式模型進行比較。在實驗中，兩組皆使用相同的交易數據集進行訓練，目的是檢測在數據不共享的情況下，FL+ GNN是否能達到或超越集中式模型的偵測效能。從實驗中發現，FL+ GNN模型相較於集中式模型，表現略有優勢。具體來說，FL+GNN在準確率和召回率上均略高，分別為 0.5937 和 0.5937，相較於集中式模型的提升了約 4%。儘管兩者的精確率相差不大，FL + GNN 模型的 F1 分數（0.5361）相比集中式模型（0.5192）提升了約 3.3%。

綜上所述，本研究在保護數據隱私的環境下實現了高度準確的異常交易偵

測, 還成功應對了資料不平衡問題, 這些結果證明了 FL + GNN 結合應用於反洗錢領域的巨大潛力, 為金融機構提供了一種高效且安全的解決方案, 提升異常交易偵測的準確性和效率。

二、應用產業與對象

本研究適用於上市上櫃公司及各類機構, 特別針對因數據機密性限制而無法共享原始數據的行業具有顯著應用價值, 例如金融業與醫療產業等。研究中採用的技術, 能在保障數據隱私的前提下實現多方協作分析, 從而突破數據孤島的限制。

隨著研究的不斷深入與技術的持續優化, 本研究的應用範圍有望進一步拓展至其他數據敏感性高的領域, 如保險業、電子商務、公共衛生以及智慧城市等產業, 為多樣化場景提供高效且安全的數據協作解決框架。

三、亮點

1. 留出法

異常行為偵測模型的目標是識別和檢測數據中的異常行為, 但若偵測目標具有高度隱蔽性, 模型在訓練過程中容易受到特定數據模式的影響, 導致過擬合現象, 使模型在訓練數據上的表現非常優異, 但在新數據上的準確性卻有所降低。為應對這一問題, 留出法(Hold-out Method)是一種解決方案, 將原始數據集分割為訓練集和測試集, 在模型訓練時僅時使用訓練集, 並在訓練完成後使用測試集評估模型在陌生數據上的表現, 模擬模型應對真實環境的能力。

在此研究中, 原始資料集按8:2的比例切分為訓練集與測試集, 其中80%用於訓練模型, 20%用於測試模型[7]。留出法的使用能應對模型過擬合的風險, 同時增強模型的泛化能力, 使其更精確地模擬在面對複雜且動態的反洗錢情境下的實際應用需求。

2. 資料不平衡的處理

訓練反洗錢模型時, 原始資料的不平衡是一大問題, 因為在實際情境中, 洗錢行為在大量正常交易數據中僅占極小比例, 導致洗錢標記資料量相較於正常交易數據量非常有限, 以本專題使用的資料庫為例, 洗錢行為僅占全體資料的0.12%。在資料平衡處理中, 上採樣和下採樣是常用方法, 上採樣通過增加少數類別樣本數量來提升該類別的特徵權重, 而下採樣則通過減少正常類別樣本數量, 使模型更專注於少數類別樣本, 提升辨識能力。在反洗錢模型的應用中, 下採樣更為合適, 因為下採樣通過減少正常類別樣本數量, 使正常和異常樣本的數量達到相對平衡, 避免了引入重複樣本所帶來的偏擬合風險, 保持更高的泛化能力。同時下採樣減少了模型訓練所需的數據量, 可以降低模型的訓練成本與資源需求, 確保訓練資源的有效利用。

基於上述考量, 在訓練以及測試資料集的比例設計上, 本研究經由下採樣將異常交易(洗錢行為)與正常交易的比例調整為1:2, 以提升模型在異常行為檢測中的靈敏度和辨識效果。選擇1:2比例, 旨在提升模型對異常行為的檢測靈敏度和準確性, 使模型能從充足的異常樣本中學習, 而不會因樣本稀少而導致學習不足, 進而增強模型對洗錢行為的辨識準確度與表現性。

3. 應用圖神經網路提升洗錢偵測準確性

傳統的反洗錢方法主要依賴規則引擎和人工審查, 規則引擎透過預先設定的規則, 對交易金額、匯款頻率等資料進行篩選和監控, 辨別可疑活動; 人工審查則由專業人員對可疑交易進行深入分析和判斷。然而隨著金融交易數量和複雜度的增加, 這些方法面臨諸多挑戰如效率低下、無法應對新型洗錢模式等等。

為應對上述挑戰, 應用圖神經網絡(Graph Neural Network, GNN)成為反洗錢研究的重點[10]。相較傳統的反洗錢方法通常基於單一交易的特徵, GNN能夠從整體網絡的角度出發, 分析交易網絡中的多層次關係。GNN將每筆交易和帳戶視作圖中的節點, 並根據交易和資金流向在節點間建立連接, 形成一個圖結構。通過此結構, GNN能夠識別潛在的複雜交易模式, 例如洗錢行為中可能

出現的多層次轉帳和複雜資金流動等, 揭示洗錢網絡中隱蔽的關聯。這些多層次的關聯分析讓模型能夠檢測出可能的異常行為, 即便其單一交易特徵並不顯眼。同時GNN能夠自動學習圖結構中的特徵, 無需人工手動設計特徵。GNN透過多層神經網絡對圖節點進行嵌入表示, 使得交易節點的特徵能夠包含周圍節點的資訊及其關係, 這種自動特徵學習方式讓GNN能夠根據圖結構和交易的特性, 自主識別出異常模式, 從而有效識別新的、隱蔽的洗錢手法。

綜上所述, GNN在反洗錢領域具有顯著的優勢, 能有效解決傳統方法的不足, 提升洗錢行為的偵測能力, 因此本研究透過採用GNN技術計畫提高反洗錢系統的偵測準確性, 並提升系統的適應性和效率, 有助於金融機構更有效地應對不斷演變的洗錢手法, 保障金融體系的安全與穩定。

4. 聯邦式學習達成跨機構協作

聯邦式學習(Federated Learning, FL)是一種分散式機器學習方法, 允許多個參與方在無需共享原始數據的前提下協同訓練模型。聯邦式學習的過程中, 各參與方僅需在本地執行模型訓練, 並傳送模型參數或梯度, 而非原始數據, 從而構建出一個綜合多個本地模型特徵的全局模型。這不僅能保障數據隱私, 符合數據隱私保護的相關規範, 還能藉由多方協同的方式顯著提升模型的預測精準度和泛化能力。

反洗錢模型的訓練過程中, 為了提升模型的預測精準度及泛化能力, 通常需要使用大量原始資料進行訓練。然而這在實際應用中遭遇許多困難, 因訓練資料往往包含大量敏感的個人和企業財務信息, 造成對數據隱私與安全性有嚴格要求, 同時傳統的集中式訓練模式往往需將數據集中至單一伺服器, 這可能導致數據洩露和隱私侵犯的風險, 使得在多機構之間協作訓練模型變得困難[12], 這些阻礙限制了金融機構間資料共享的意願與能力, 而聯邦式學習能統合異構數據資源, 有效解決這些問題。

本研究在聯邦式學習中使用了 Flower (FL) 框架, 是一個支援聯邦式學習的 Python 函式庫, 具備高度彈性和可擴充性, 能讓不同端點共同參與模型訓練, 同時保留各端的數據隱私。而在各個模型的權重調整上, 本研究使用聯邦平

均策略（Federated Averaging, FedAvg）作為聚合方式, FedAvg 是聯邦式學習中最經典的模型聚合方式, 通過對各個客戶端的模型權重進行平均, 生成全局模型, 確保協作訓練過程的有效性和穩定性。

5. 模型性能提升與實驗結果

在本研究中, 本組進行了兩項主要實驗, 以驗證聯邦學習（Federated Learning, FL）結合圖神經網絡（Graph Neural Network, GNN）對於異常交易偵測的效能提升情況。第一個實驗比較了單一客戶端模型與聯邦學習模型的異常偵測能力, 以觀察多機構協作學習是否能有效提升模型的表現;第二個實驗則將 FL + GNN 模型與集中式模型進行比較, 以檢測在數據不共享的情境下, FL + GNN 模型的偵測效果是否能達到或接近集中式模型的水準。

此外, 由於數據集中的異常交易樣本數量稀少, 為了增強模型的異常偵測能力, 我們在資料預處理階段引入下採樣技術, 使模型能更有效地識別少數類別。以下將分別對這三項實驗的結果進行分析。

5.1. 資料不平衡處理

本研究使用下採樣技術來平衡資料中的異常與正常交易比例, 以提升模型對少數類別樣本的偵測效果。經實驗測試, 結果顯示下採樣顯著提升了模型的異常偵測性能。在處理後的數據集中, 模型的精確率、召回率和F1分數均有顯著提升。

根據實驗結果, 下採樣後模型的精確率提升約 374%, 召回率提升約 150%, 而 F1 分數提升了約 78%。這些結果表明下採樣技術在處理資料不平衡問題上具有顯著的效果, 有效增強了模型對異常交易的辨識能力, 使其在實際應用中能更準確地識別潛在異常行為。

5.2. 客戶端與聯邦學習（FL）模型比較結果

在本實驗中, 本組比較了單一客戶端模型與聯邦學習（FL）模型在異常交易偵測上的效能。兩個客戶端各自基於本地數據進行訓練, 並分別測試其準確率、精確率、召回率和 F1 分數, 然後將結果與 FL 模型的全局性能進行對比。

根據表一中的數據, 可以觀察到以下主要結果:

- 性能提升:FL 模型的全局性能相較於單一客戶端模型有顯著提升。準確率和召回率均提升了約 22%, 精確率提升了約 23%, 而 F1 分數提升了約 30%。這些提升表明, 通過 FL 模型的協同學習, 系統能夠有效地綜合來自不同客戶端的數據特徵, 從而提高異常交易的偵測能力。
- 協同學習的效果:單一客戶端模型僅基於其本地數據進行訓練, 因此其異常偵測性能較低, 尤其是在資料較少的情況下, 模型無法有效捕捉異常模式。相比之下, FL 模型通過聯邦平均策略(FedAvg)進行參數聚合, 整合了各客戶端的數據特徵, 這使得 FL 模型能夠在多機構合作下更準確地識別異常交易行為。
- 數據隱私的保護:值得注意的是, FL 模型在提升異常偵測效能的同時, 保持了數據隱私。各客戶端僅共享模型參數, 而不直接共享數據, 這使得 FL 模型在異常偵測上具有更高的應用價值。

(表一)

	Accuracy	Recall	Precision	F1
Client1	0.48575	0.48575	0.49110	0.41215
Client2	0.48526	0.48526	0.50172	0.41186
全局模型	0.59370	0.59370	0.61202	0.53612

總體而言, 聯邦學習(FL)模型在異常交易檢測中比單一客戶端模型表現更為優越。它能夠有效整合來自多機構的數據特徵, 提高了模型的準確性和泛化能力, 同時保持了數據隱私的保護。在未來的研究中, 進一步優化聯邦學習過程中的性能、客戶端選擇機制及數據分配問題, 將有助於進一步提升FL模型的效能和可擴展性。

5.3. FL + GNN 模型與集中式模型比較結果

在本實驗中, 我們比較了 FL + GNN 模型與集中式模型在異常交易偵測任務上的表現。兩者均使用相同的交易數據集進行訓練, 目的是評估在數據不共享的情況下, FL + GNN 是否能達到或超越集中式模型的偵測效能。

根據表二中的數據, 我們可以觀察到以下主要結果:

- FL + GNN 與集中式模型的性能比較:FL + GNN 模型相較於集中式模型, 表現略有優勢。具體來說, FL + GNN 在準確率(Accuracy)和召回率(Recall)上均略高, 分別為 0.5937 和 0.5937, 相較於集中式模型的提升了約 4%。儘管兩者的精確率相差不大, FL + GNN 模型的 F1 分數(0.5361)相比集中式模型(0.5192)提升了約 3.3%。
- FL + GNN 模型的優勢:FL + GNN 模型通過聯邦學習在多機構合作的框架下進行協作學習, 避免了數據的直接共享, 這使得模型能夠在保障隱私的同時, 充分利用不同客戶端的數據特徵, 進一步提升了對異常交易的識別能力。儘管集中式模型在全局訓練中可以使用完整的數據集, 但在數據隱私要求較高的環境下, FL + GNN 顯示出其獨特的應用價值。

(表二)

實驗結果				
	Accuracy	Recall	Precision	F1 Score
聯邦式	0.593703	0.593703	0.612023	0.53612
集中式	0.571157	0.571157	0.576457	0.51917

不僅如此, 我們還發現一個有趣的現象:儘管直觀上集中式模型應該表現更好, 但根據實驗結果顯示, 兩者的表現差距其實不大, 並且在某些情況下, FL 模型反而表現得更為優越。FL 模型的優勢主要來自於其分散式訓練方式。儘管

兩者使用相同的資料集, FL 模型能夠在不同客戶端進行本地訓練, 並透過聯邦平均策略進行參數聚合, 這使得 FL 模型能夠更好地捕捉來自多機構數據的多樣性與變異性, 進而提升模型的準確性和泛化能力。相比之下, 集中式模型將所有數據集中進行訓練, 可能會出現過度擬合的情況, 尤其是在訓練資料中某些模式過於顯著時, 這會使得集中式模型過度依賴訓練數據, 進而影響其在未知數據上的表現。因此, FL 模型通過協同學習有效減少過度擬合的風險, 並提升其在異常檢測任務中的效能。

總體而言, FL + GNN 模型在異常交易檢測中表現與集中式模型相媲美, 並且在隱私保護上提供了顯著的優勢。儘管兩者在某些指標上的表現差異不大, FL + GNN 模型在實際應用中能夠更好地應對數據隱私保護需求, 顯示出其在分散式學習架構下的潛力。未來的研究可進一步優化 FL + GNN 的協作學習策略, 以提升其在多機構環境中的效能和可擴展性。

綜上所述, 本研究在保護數據隱私的環境下實現了高度準確的異常交易偵測, 還成功應對了資料不平衡問題, 這些結果證明了 FL + GNN 結合應用於反洗錢領域的巨大潛力, 為金融機構提供了一種高效且安全的解決方案, 有助於實現多機構協同合作, 提升異常交易偵測的準確性和效率。

參、結論

本研究成功地將聯邦式學習（Federated Learning, FL）與圖神經網絡（Graph Neural Network, GNN）結合，應用於反洗錢領域，為金融機構提供了一種在保護數據隱私的前提下提升異常交易偵測能力的創新解決方案。透過 GNN 的圖結構學習能力，我們有效地捕捉了金融交易網絡中的複雜關聯和潛在異常模式；而 FL 的引入則克服了傳統集中式訓練對數據隱私的挑戰，實現了多機構間的協同學習。這一研究的應用不僅提供了技術創新，還在滿足數據隱私法規的基礎上，為多機構間的數據協同和模型共享提供了可行的解決方案。這對於提升反洗錢的效率和準確性具有重要意義，能夠幫助及時發現並阻斷洗錢活動，進而維護金融市場的穩定與公信力。

然而，這項研究也存在一些局限性需要重視。首先，使用的數據集規模相對較小，且為模擬數據，這可能導致模型在實際應用中的表現與預期存在差距。此外在實際應用中，聯邦式學習可能面臨通訊成本高和客戶端計算資源有限等挑戰，這對技術的普及性和實用性提出了要求。隨著洗錢手法變得日益複雜和多變，模型的泛化能力仍需進一步提升，以應對更為嚴峻的挑戰。

未來的研究方向可以聚焦於幾個方面的優化，進一步提升技術的應用價值。首先，可以擴展模型的適用性，將其應用於更大規模的數據集和更多的金融機構，以驗證模型的可擴展性和實際效果。其次，將其他數據來源，例如用戶行為數據和社交網絡關係等納入分析，從而提升異常交易偵測的準確性和全面性。此外，進一步引入差分隱私和安全多方計算等技術，可以加強數據在聯邦學習過程中的隱私保護，降低數據洩露的風險。

綜上所述，本研究為反洗錢領域提供了一條嶄新的技術路徑，成功結合了聯邦式學習與圖神經網絡的優勢，實現了在保護數據隱私的前提下提升異常交易偵測能力的目標。儘管仍有優化的空間，但這一研究為金融機構在反洗錢領域提供了重要的參考，並期望未來能夠促進行業間的多方合作，共同應對日益複雜的洗錢手法，進一步維護金融體系的安全與穩定。

肆、參考資料

[1] Kasawnih, A. A. (2020). Evolution or Regress of the Size and Scope of Global Phenomenon of Money Laundering. *Proceedings of the International Conference on Business Excellence*, 963–974.

[2] Husnaningtyas, N., Hanin, G. F., Dewayanto, T., & Malik, M. F. (2023). A Systematic Review of Anti-Money Laundering Systems Literature: Exploring the Efficacy of Machine Learning and Deep Learning Integration. *JEMA Jurnal Ilmiah Bidang Akuntansi Dan Manajemen*, 91–116.

[3] Phua, C., Lee, V. V., Miles, K. S., & Gayler, R. W. (2010). A Comprehensive Survey of Data Mining-Based Fraud Detection Research. *Artificial Intelligence Review*.

[4] Peña, L. S., & Quijaite, J. jesus S. (2024). *Machine Learning Models for Money Laundering Detection in Financial Institutions. A Systematic Literature Review*.

[5] He, H., & Garcia, E. A. (2009). "Learning from Imbalanced Data." *IEEE Transactions on Knowledge and Data Engineering*.

[6] Kairouz, P., Mcmahan, H. brendan, & Avent, B. (2019). *Advances and Open Problems in Federated Learning*.

[7] Géron, A. (2019). *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow*. O'Reilly.

[8] Fernández, A., García, S., & Galar, M. (2019). *Learning from Imbalanced Data Sets*.

[9] Wei, M. (2022). *Advances and Open Problems in Federated Learning*.

[10] Wu, Z., Pan, S., Chen, F., & Yu, P. S. (2019). *A Comprehensive Survey on Graph Neural Networks*.

[11] Hu, W., Leskovec, J., & Jegelka, S. (2018). *How Powerful Are Graph Neural Networks?*

[12] Chishti, S., Bartoletti, I., Leslie, A., & Millie, S. M. (2020). *Automated Machine Learning and Federated Learning*.