東海大學統計學系

碩士論文

指導教授：王榮琮博士

Credit Card Fraud Detection using Hidden Markov Model

隱藏式馬可夫模型應用於信用卡欺詐偵測

研究生：陳雅馨

中華民國一百零八年三月

東海大學碩士班研究生

論文口試委員審定書

統計學系碩士班陳雅馨君所提之論文

隱藏式馬可夫模型應用於信用卡欺詐偵測

經本委員會審議，認為符合碩士資格標準。

論文口試委員召集人 陳主華 （簽章）

委員 江彌邵

王紫岑

中華民國 108 年 03 月 25 日

隱藏式馬可夫模型應用於信用卡欺詐偵測

中文摘要

由於互聯網的便利性和電子商務技術的創新，信用卡支付數量的增長比以往還要更加强勢。信用卡交易量的急劇增加，同時詐欺交易的數量也同時增加。對金融業者來說，信用卡欺詐所造成的成本會使機構蒙受巨大損失，因此如何識別欺詐交易，甚至建立欺詐偵測系統 (FDS) 已經成為金融業者的主要問題之一。持卡人表現出特定的消費行為。每一位持卡人都可以用一組典型的購買類別模式表示，例如上次購買的時間，以及消費的金額等，但是該模式的偏差會對欺詐偵測系統產生潛在威脅。本論文探討如何將隱藏式馬可夫模型 (HMM) 運用到信用卡交易過程的序列中和偵測信用卡欺詐。應用 HMM 的優點在於它能夠即時偵測以及有效的增加準確率。本文提出一個可以同時檢測以單一持卡人為中心和以機構 (銀行或者是商家) 為中心的詐欺交易之框架。我們使用 BankSim 的資料進行單一持卡人與機構的詐欺偵測，透過 K-means 聚類演算法，將交易金額分為幾個觀察符號。先用訓練資料估計模型參數，再將所估計的參數帶入 HMM 測試資料。以機構為中心的交易，由於交易數量龐大，使用非重疊窗口偵測欺詐交易。單一持卡人為中心的交易，交易數量相對前者小許多，使用重疊的窗口來偵測欺詐交易。為衡量 FDS 的效能，文章中使用真陽率、偽陽率等指標進行效能判斷，並與其他分析方法比較，結果顯示採用 HMM 捕追詐欺的能力佳，但在準確率上略為失色。

關鍵字：隱藏式馬可夫模型、信用卡偵測系統、雙重隨機過程。

Credit Card Fraud Detection using Hidden Markov Model

Abstract

Credit card payment has strongly growing due to easy access to internet and innovations in e-commerce technology. The number of transactions has increased dramatically not only on regular credit card consumption but also on fraud events. Identifying fraudulent transactions and establishing an efficient fraud detection system (FDS) have become major issues in the financial industry as costs of credit card can bring substantial losses for financial industry. By regarding the true-fraud transactions as hidden states, this thesis applies a hidden Markov model (HMM) to analyze the sequence of operation in credit card transaction processing and shows how it can be used for detecting fraud. An HMM is a doubly stochastic process with an underlying Markov process that is not directly observable but can be inferred by analyzing another set of stochastic process which produces the sequence of observations. In practice, testing FDS is difficult as banks usually do not agree to share their data with researchers as well as no benchmark data set are available. To evaluate the efficiency of the HMM-based FDS, we consider three criteria, namely, true positive rate, false positive rate and overall accuracy. Simulation results show that the proposed HMM-based FDS performs more efficiently in terms of true positive rate compared with false positive rate. Overall, the HMM-based FDS performs well in terms of overall accuracy.

Keyword : Hidden Markov Model, Credit card fraud detection system, Doubly stochastic process.

目 錄

圖 錄

表 錄

第一章　緒論

第一節　研究動機與背景

在近幾年來，金融市場和電子商務 (例如淘寶、亞馬遜和京東) 逐漸走向全球化，隨著商業市場的發展，以信用卡作為交易使用媒介的行為逐漸增加，The Nilson Report (no.1121) 指出，全球在 2016 年的信用卡發卡量成長 4.7 %，已發行近 121 億張晶片卡 (晶片卡包含信用卡、借記卡和預付卡)，產生 2571.7 億筆交易，較前一年成長近 13.3 % (約 302.1 億美元)，信用卡作為交易使用媒介的行為在消費市場扮演重要角色。在台灣，根據聯合信用卡處理中心 105 年度年報，銀行發卡數量亦逐年成長，在 2016 年的信用卡發卡量成長 5.67 %，簽帳金額達新台幣 2.42 兆元，年成長 8.53 %。全球每年的信用卡累積發卡量與簽帳金額均呈現快速成長，信用卡業務已逐漸成為銀行重要的收入來源之一。

在信用卡市場快度成長的同時，因為信用卡衍生而來的信用卡詐欺金額也逐年攀升，The Nilson Report (no.1118) 指出，全球在 2016 年晶片卡詐欺金額達到 20.18 億美元，占所有交易金額的 0.0775 %。聯合信用卡處理中心亦顯示，在 2016 年處理中心會員通報詐欺金額為 12.7 億 (臺幣)，其中以 MOTO/ EC (郵購、電購和線上購物等未經持卡人授權之非面對面交易) 占最大比例，高達 88.6 % 。

由於信用卡用戶遍布世界各地，因此信用卡交易處理的數量龐大，再加上持卡人往往沒有意識到，他的信用卡資訊已經遭到竊取。實務上，發卡銀行會針對持卡人消費行為的時間、筆數、頻率、金額等，作為觀察的要點，判斷持卡人的使用行為是否有異常。透過詐欺偵測系統 (Fraud Detection System, FDS)，將某一段時間間隔中高刷卡次數，或者金額有所異常的卡片做預防動作，以避免發卡銀行蒙受損失以及影響持卡人的權益。為了降低發卡銀行所面臨的信用風險和詐欺風險，並減少信用卡所導致的金融損失，建立一個能夠即時偵測出詐欺，並且做出有效決策的信用卡詐欺偵測系統，已成為金融業的一項重要挑戰。

過去幾年，多位學者已經提出了一些偵測信用卡欺詐的相關研究，這些研究主要是基於決策樹 (Decision Tree)、類神經網絡 (Artificial Neural Network)、遺傳演算法 (Genetic Algorithm)，分類 (Classification)、分群 (Clustering) 等，所謂的機器學習技術 (Machine Learning)。機器學習是讓機器具有學習能力，從資料中自動學習規則，並利用規則對新的資料進行預測。大多數已提出的方法所面臨的問題是，它們的分類器 (Classifier) 需要對真實與詐欺這兩種交易作標記，才能被訓練。然而，真實情況是，獲取真實的詐欺數據本身就是信用卡詐欺偵測相關的最大問題之一。並且，這些方法無法偵測到無標記數據的新型詐欺，換句話說，這些方法對於一筆新交易的到來，缺乏動態的偵測能力。

本論文的研究目的是提出一個，基於隱藏式馬可夫模型 (Hidden Markov Model, HMM)，具有即時性以及高準確率的信用卡 FDS，並且透過模擬資料對比其他技術。HMM 的優點在於不需要欺詐標記，只需透過持卡人的消費習慣便能從中偵測欺詐行為，並具有對新交易的即時動態分析能力。本論文並進一步由資料集中，同時進行單一持卡人與機構兩種不同的詐欺偵測。

本論文內容共分為五章。第一章為緒論，說明本論文對於信用卡詐欺偵測的研究動機與背景、研究目的、信用卡交易流程、信用卡詐欺種類、與近期偵測信用卡詐欺方法。第二章為研究方法，論述本論文所使用之 HMM，包含 HMM 的架構、HMM 的基本問題與其對應的演算法、並給一個生活示例。第三章則說明如何運用 HMM 至 FDS，包含對應到 FDS 中的觀察符號、隱藏狀態、和機率矩陣，如何進行詐欺偵測，及系統效能評估。第四章為資料模擬，使用 BankSim 的資料進行單一持卡人與機構的詐欺偵測，分析結果顯示 HMM 捕追詐欺的能力佳，但在準確率上略為失色。第五章為結論。

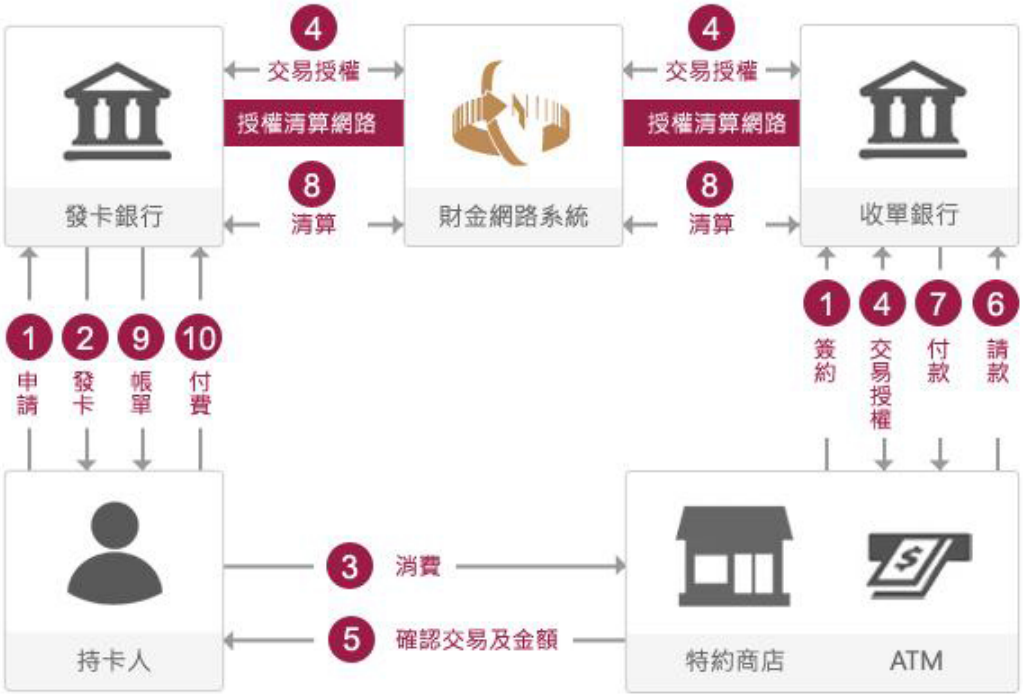
第二節　信用卡交易流程

1. 信用卡與借記卡

發卡銀行依照持卡人的信用狀況及經濟能力，核發一張尺寸大小符合 ISO-7810 國際標準的卡片，銀行核卡時會發給持卡人一張有額度上限的信用卡。除了實體信用卡外，虛擬信用卡也日趨風行。信用卡可以在卡號內沒錢的情況下，先消費、預支現金、或分期購物，可以延遲還款直到繳款截止日再把錢還給銀行。無法還款時，發卡銀行會收取高額年利率的循環利息。

信用卡不同於借計卡 (Debit Card)。借記卡是持卡人先把錢存進帳戶裡，然後再持卡消費。借記卡在 Mastercard 稱為轉帳卡，在 Visa 稱為 Visa 金融卡，是先存款後消費 (或是取現)，可以在網路或 POS (Point of sales) 消費，或者透過 ATM 轉帳和提款，消費或提款時資金直接從儲蓄帳戶扣款，不能透支。借計卡是連結活存帳戶的支付工具，持卡人必須同時在發卡銀行開立活存帳戶，表面上是信用卡，實質上是提款卡，可提領可刷卡，但不能預支，也不一定能辦分期購物。和信用卡一樣，借記卡卡片的尺寸大小也是採用 ISO-7810 國際標準。

過去信用卡交易採用磁條和簽名的方式，由於磁條和簽名易受側錄，詐欺者便能獲取持卡者的信用卡訊息製作偽卡。現今發卡銀行已全面改用晶片密碼卡，減少側錄所造成的信用卡詐欺。信用卡的正面有發卡銀行名稱、信用卡別、EMV 晶片、卡號、持卡人姓名和有效日期，卡片背面有卡片磁條、持卡人簽名欄和信用卡安全號碼。信用卡授權請求的通訊協定格式是根據 ISO-8583 國際標準 (Marshall, 2007)。



圖一、信用卡交易流程圖 (財金資訊股份有限公司)

2. 信用卡交易流程

一個完整信用卡交易之運作流程，由財金網路系統 (國際信用卡組織/ 國內清算中心/ 授權轉接中心)、發卡銀行、收單銀行、特約商店與持卡人五個主體參與其中。申請人先向發卡銀行提出信用卡申請後，發卡銀行會先審核申請者之背景資料 (包含性別、年齡、所得、資產等項目)，待徵信審核通過之後，即核發信用卡，成為持卡人。持卡人可持信用卡至任一特約商店刷卡消費，特約商店將持卡人卡片連線到發卡銀行，獲得發卡銀行授權後，即接受持卡人消費，並提供持卡人銷售的物品或服務。收單銀行透過財金網路系統向國內外發卡銀行進行帳單帳務清算，收單銀行並依規定與特約商店約定之付款期限撥款予特約商店。而發卡銀行於每月結帳日後，向持卡人寄送對帳單並要求付款，持卡人需於該月最後繳款日前完成付款。這五個主體便構成一個完整的信用卡交易流程，信用卡交易流程如圖一所示。(財金資訊季刊第 67 期 2011/ 07/ 07)

第三節　信用卡詐欺類別和現行發展與挑戰

1. 信用卡詐欺

信用卡詐欺風險是指不法分子藉由惡意透支、騙領、冒用、使用偽造或作廢的信用卡及特約商店詐騙，讓銀行造成經濟損失的可能性，也容易產生消費者與發卡銀行之間的消費糾紛與責任歸屬問題。詐欺風險形式多樣，隱蔽性强，是信用卡業務中最直接、最難追索的風險。聯合信用卡處理中心將信用卡詐欺類型分成八類：

(一) 遺失卡 (Lost) — 卡片因為遺失而被盜用。

(二) 偽卡 (Counterfeit) — 使用未經發卡機構授權製作之卡片。

(三) 未達卡 (Mail Non-receipt) — 卡片在寄送過程中被他人截獲並開卡進行交易。

(四) 被竊卡 (Stolen) — 卡片被竊而發生盜用。

(五) MOTO/ EC — 通過電話、郵件和網際網路等不需要出示真實卡片的管道使用信用卡。犯罪分子只需提供非法取得的他人信用卡卡號、持卡人姓名、信用卡有效期限和信用卡安全號碼等資訊，即可進行詐欺性交易。

(六) 冒用申請卡 (Account Take-over) — 詐欺者獲取了部分或全部真實持卡人資訊，並假冒真實持卡人對卡帳戶的資訊進行變更，要求信用卡公司把郵件送到新的通信地址，然後向信用卡公司謊報信用卡丟失了，要求把新的信用卡寄到詐欺分子指定的通信地址。

(七) 多刷帳單之詐欺 (Merchant Fraud) — 商店詐欺包含惡意倒閉、虛假商店、洗單、信函、電話、網路行銷詐欺、商戶套現、側錄、卡號測試詐欺等。

(八) 其他 — 不屬於以上詐欺類型者。

2. 信用卡詐欺偵測系統 (FDS)

由於信用卡詐欺手法不斷翻新，傳統之作業方式已無法掌握時效，因此各銀行均積極進行即時詐欺偵測，以期能達到有效控管、降低損失、節省人力、及提高服務品質。一組有效率的詐欺偵測系統 (FDS)，能即時偵測信用卡交易是否出現異常，若系統判定異常，當下停止交易授權給特約商店，通知持卡人信用卡有遭盜用的疑慮，達到降低詐欺風險的目的。大多數的信用卡詐欺偵測系統是使用異常檢測 (Anomaly or outlier detection)，即從衆多數據找出有所差異的數據，例如針對持卡人的消費行為做識別，判斷與過去相比是否有無異常。

由於詐欺檢測存在高度的複雜性，目前的研究重點是盡量提高模型對詐欺交易的預測，減少將正常的交易分類為詐欺交易的比例，並且希望模型能夠迅速地進行檢測。對於詐欺檢測目前已有許多的研究，部分研究是有關通訊、模式識別、金融商品等領域的詐欺檢測，而部分的技術是針對信用卡相關的詐欺檢測。早期 Ghosh and Reilly (1994) 使用類神經網路 (Artificial Neural Network) 到信用卡詐欺偵測研究。利用財務報表中的信息作為類神經網路模型中的欺詐信號，Aleskerov et al. (1997) 提出 CARDWATCH 用於信用審批，破產預測，股票選擇和自動交易。Phua et al. (2007) 將應用數據挖掘技術的 FDS 進行了廣泛的調查。Chen et al. (2005) 則是提出一個用於信用卡欺詐檢測的個性化方法，這的方法同時採用支持向量機和類神經網路。早期的技術如類神經網路，屬於監督視學習，必須對信用卡正常或詐欺交易作標籤，才能進行分類，因此 Srivastava et al. (2008) 提出信用卡交易處理序列由 HMM 建模，HMM 屬於非監督視學習。

在第一章第四節，進一步介紹這幾種用於建構信用卡詐欺偵測系統的方法，第二章則是進一步介紹 HMM，其架構、基本問題與其對應的演算法則。我們引用 Prakash and Chandrasekar (2012) 的話，他們認為 HMM 或許是應用於詐欺偵測領域中成功的方法之一。

3. 信用卡詐欺系統的挑戰

從機器學習的觀點來看，基於異常檢測的 FDS 可分成監督式學習 (Supervised) 和非監督式學習 (Unsupervised)，這兩種學習的差異在於監督式學習知道該筆數據那些為異常值—例如信用卡交易數據有標籤 (label) 顯示哪些數據為詐欺，而非間監督式學習是不需要標籤的，它可以自行學習進行異常值分類。

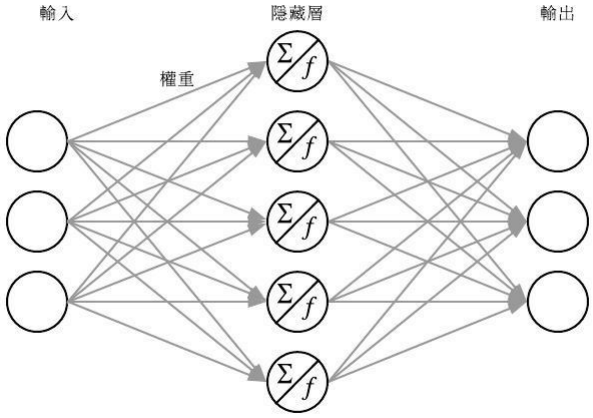
異常檢測主要會面臨三個問題：概念漂移、數據不平衡和即時偵測。

- 概念漂移的發生基於異常檢測主要依據持卡人的消費行為建立模型，當持卡人因特定情況所產生的的消費行為 (例如生日大餐和禮物)，此時詐欺偵測可能會將此次消費列入異常值，視為詐欺交易。
- 不平衡資料為信用卡詐欺只占所有交易比例的小部分，由於大部分演算法建立在數據平衡的假設下運作，在面臨實際資料可能出現將詐欺交易誤判成正常交易。
- 由於交易筆數龐大詐欺交易隱藏於其中且隨時可能發生，詐欺偵測系統若能即時偵測到詐欺交易預警發卡銀行，發卡銀行便可以進行後續控卡動作，降低損失。

發卡銀行即便對詐欺偵測系統顯示異常者採取控管，由於詐欺手段多元且與時俱進，持卡人的消費觀也並非一成不變。透過過去交易所建立的異常檢測模型只能檢驗出和去過去形似的詐欺交易，因此建立一個完整且能夠不斷修正調整以及即時偵測回復的系統顯得更為重要，有效遏止發卡公司和持卡人蒙受損失。

第四節　近期偵測信用卡詐欺方法

1. 類神經網路 (Artificial Neural Networks, ANN)

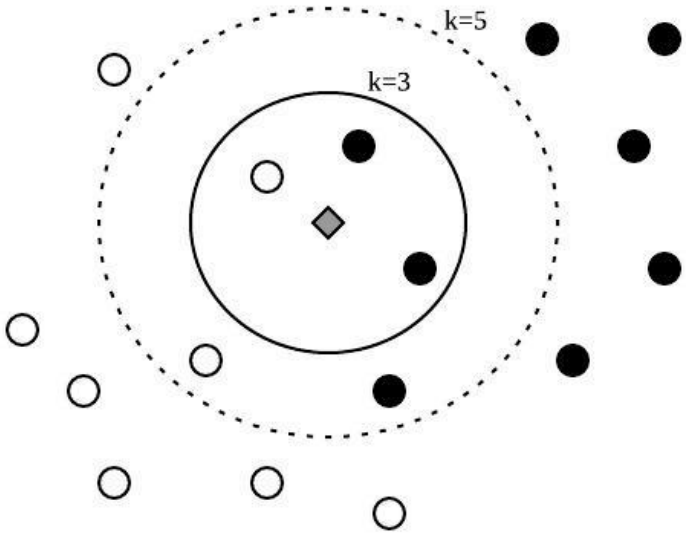


圖二、類神經網路

ANN 核心想法為模擬神經元細胞對外在刺激的反應，它模擬神經元細胞將之連結形成網路狀模型，將輸入的符號不停地在這個網路上傳遞下去。這個結構與運作啓發 David Hubel and Torsten Wiesel 於 1981 年對於動物視覺系統的發現。倒傳遞網路 (Back Propagation Network) 是 ANN 學習模型中應用最普遍的模型，它是利用最陡坡降法 (Gradient Steepest Descent Method) 將誤差函數最小化。

ANN 主要有三個要素，分別為輸入值、轉換函數和輸出值。將輸入值輸入神經元，神經元代表給定的轉換函數，而輸入值代入神經元會產生輸出值，再將這輸出值作為輸入值再代入下一層的神經元，直到最後一層形成一預測結果停止，換言之，上一層神經元的輸出結果，作為輸入值代入下一層的神經元，圖形表示如圖二。由於每一個輸入值都會連接一個權重，因此需要透過不斷的學習修正權重，使最後的輸出值盡可能對應到實際值。ANN 屬於監督學習的一種，隱藏層多且須不斷的修正權重，ANN 的優勢在隨著時間的演進，不斷的修改學習，使結果接近實際值，甚至預測下一步。ANN 雖然能使模型誤差越來越小，但它可能會面臨過度學習和收斂速度過慢，或是無法收斂的情況。

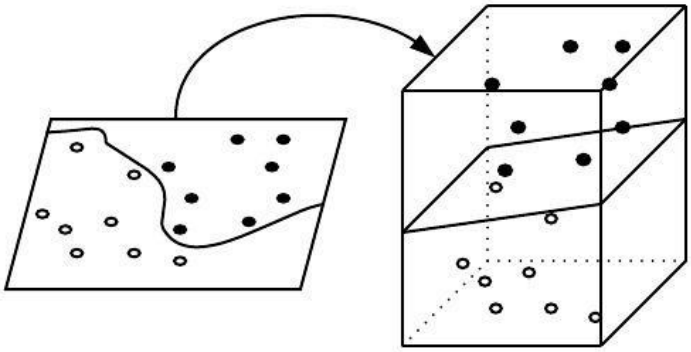
2. k-最近鄰居法 (k-Nearest Neighbor, KNN)



圖三、最近鄰居法

KNN 為基於分類和迴歸的演算法，是 Cover and Hart 於 1968 年提出。KNN 主要的核心有三個要素，分別為 k 值選擇、距離演算法選擇、和分類決策規則。基本的想法為根據已定類別的訓練資料找出對測試資料點鄰近的 k 個訓練資料點，通常以多數決的形式，選擇這 k 個鄰近點中出現最多次的類別進行分類。圖三為 k 值分別為 3 和 5 時，決定中間菱形點被歸類至黑球或白球的呈現。該方法沒有明顯的學習過程，僅透過訓練資料所得的特徵向量空間進行劃分，但需事前知道訓練集分類標籤，為監督式學習的一種。KNN 一個非常好的性質是訓練資料的分布不需要事前假設，然而 KNN 在某一類樣本遠不同於其他類，容易導致分類錯誤。演算法過程需計算所有點的距離，排序求得最近的 k 個鄰近值，因此需要龐大的電腦計算量。

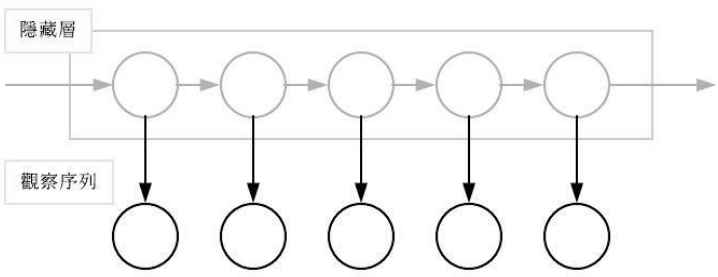
3. 支持向量機 (Support Vector Machine, SVM)



圖四、支持向量機

SVM 為 Cortes and Vapnik 在 1995 年提出，核心想法為使特徵空間中分類間隔為最大的超平面，即透過超平面將不同類別區隔開來，使超平面的兩側邊界為最大。圖四為資料投影至高維度，使用超平面分割資料點的圖示。即便資料為線性可分，但真實資料大部分仍難以找到一個超平面將不同類別完全分開。根據超平面兩側是否允許資料點包含其中，區分為軟邊界或硬邊界的支持向量機。由於本身內含核函數，不論在線性或者非線性分類都可運作。使用 SVM 事先需要知道訓練資料的分類標籤，屬於監督式學習的一種。SVM 在分類上具有良好的特性，但在樣本過大或特徵太多時，會面臨時間運算複雜度高而難以處理。

4. 隱藏式馬可夫模型 (Hidden Markov Model, HMM)



圖五、隱藏式馬可夫模型

HMM 最早運用在語音辨識研究 (Rabiner, 1989)，Srivastava et al. (2008) 推廣到信用卡詐欺，考慮持卡人的消費習慣，建立信用卡交易過程序列為一個隨機的 HMM，基於 FDS 在運行時發卡銀行無法得知持卡人所購買物品的細節，所以他們將持卡人所購買的物品歸於隱藏的有限馬可夫鏈狀態，而購買的金額為可以觀察的序列，如圖五。HMM 更詳細的方法介紹會呈現在第二章。Robinson and Aria (2018) 提出有別於 Srivastava et al. (2008) 的方法，將信用卡詐欺交易的關注點聚焦在持卡人消費行為的異常改變，根據商家的信用卡交易來建立HMM，檢測信用卡交易序列上的異常。Santos and Ocampo (2018) 則透過馬可夫調控普瓦松過程 (Markov Modulated Poisson Process) 進行信用卡交易模擬，認為正常交易和詐欺交易各為一個常態分配，彼此間存在一個馬可夫鏈，會相互轉移。

第二章　研究方法

第一節　隱藏式馬可夫模型 (Hidden Markov Model)

Baum 和 Petrie 於 1966 年推廣隱馬可夫模型 (HMM)，在當時他們稱之為馬可夫鏈的機率函數，而 Ferguson (1980) 則是第一個使用 HMM 這名詞的學者 (Ephraim and Merhav，2002）。語音識別研究 (Speech Recognition) 普及了 HMM 的理論，自此 HMM 不僅廣泛應用到手寫識別、手勢識別等一般的模式識別 (Pattern Recognition)，並且外溢至許多應用中，例如生物訊息科學 (Bioinformatics)、機器翻譯 (Machine Translation)、基因預測 (Gene Prediction) 等領域，甚至是强化學習 (Reinforcement Learning) 以及人工智能 (Artificial Intelligence) 的其他領域。近年來已有學者將 HMM 運用在信用卡詐欺偵測上，例如，Srivastava et al. (2008)、Robinson and Aria (2018)、Santos and Ocampo (2018) 等。

馬可夫鏈是一個具備無記憶性質 (Memoryless Property)、獨立增量 (Independent Increments) 和平穩狀態 (Stationary States) 的隨機過程。無記憶性質為未來事件的狀態只和現在狀態有關和過去狀態無關，也就是在時間點 t+1 的狀態只和時間點 t 的狀態有關，和時間點 t-1 以及之前的時間點的狀態無關。例如，明天是否會下雨只和今天的天氣有關和昨天及以前的天氣無關。獨立增量是指在兩個不重疊的時間區間，其事件發生的次數多寡彼此互相獨立。平穩狀態則是指轉移機率矩陣在經過一段夠長的時間後，會收斂至一個穩定狀態。

HMM 擴展了馬可夫模型的概念，觀察值是狀態的機率函數，因此 HMM 可被視為一個雙重隨機過程，也就是在馬可夫鏈的架構裡，再增加一層輸出層。第一個隨機過程為馬可夫鏈，描述模型中狀態轉移的機率，此時的狀態序列為隱藏，無法直接觀察；能被觀察到的是第二個隨機過程，它是每一個時間點下的狀態所產生的觀察值。換句話說，我們考慮兩層結構：轉移狀態 (Transition States) 和觀察符號 (Observation Symbols)。本論文假設輸出具有獨立性，這也是一般 HMM 的假設，就是不同時間點下所產生的觀

察符號彼此獨立。例如，時間點 t 的狀態所輸出的觀察符號只和時間點 t 的狀態有關，與其他時間點的狀態跟其輸出的觀察值均無關。

1. HMM 架構

一個 HMM 可經由以下七點來說明 (參見Rabiner，1989)：

(1) 模型具備 N 個有限狀態 (states)，狀態的集合為 S，且 $S = \{S_1,\ S_2, \ldots,\ S_N\}$。

(2) 在時間點 t 的狀態被記為 q_t。

(3) 模型具備 M 個有限和可能觀察到的觀察值 (observation)，為在不同時間點所產生的各個狀態，會機率輸出的觀察值。觀察值的集合為 V，且 $V = \{V_1,\ V_2, \ldots,\ V_M\}$。

(4) 在時間點 t 的觀察值被記為 O_t，觀察值的序列為 O，$O = \{O_1,\ O_2, \ldots,\ O_T\}$，T 代表實際會看到的觀察值個數。

(5) 狀態之間有一個狀態轉移機率矩陣 (state transition probability matrix) 表示為 A。$A = [a_{ij}]$，且

$$a_{ij} = P(q_{t+1} = S_j \mid q_t = S_i), \quad a_{ij} \geq 0, \quad 1 \leq i, j \leq N \tag{1}$$

為一個 $N \times N$ 的矩陣，給定在時間點 t 的狀態為 S_i 在時間點 $t+1$ 轉移至狀態 S_j 的機率，其中 $\sum_{j=1}^{N} a_{ij} = 1, \quad 1 \leq i \leq N$。

(6) 依據不同狀態所產生觀察值為一個輸出矩陣 (Observation symbol probability) B。$B = [b_j(k)]$，且

$$b_j(k) = P(O_t = V_k \mid q_t = S_j), \quad b_j(k) \geq 0, \quad 1 \leq j \leq N, \quad 1 \leq k \leq M \tag{2}$$

為一個 $N \times M$ 的矩陣，給定在時間點 t 的狀態為 S_j 的情況下，觀察到實際觀察序列理第 k 個觀察符號。其中 $\sum_{k=1}^{M} b_k = 1, \quad 1 \leq j \leq N$。

(7) 在時間 $t = 1$ 時，各狀態都有一個初始機率 (initial probability) π_i 的初始狀態機率向量 (initial state probability vector) π，且

$$\pi_i = P(q_1 = S_i), \quad \pi_i \geq 0, \quad 1 \leq i \leq N \tag{3}$$

為一個 $N \times 1$ 的向量，其中 $\sum_{i=1}^{N} \pi_i = 1$。

一個完整的 HMM 包含三個機率分佈：轉移機率矩陣 (A)、輸出矩陣 (B) 和初始機率向量 (π)。HMM的完整參數集可以通過以下方式表徵符號 $\lambda = \{A,\ B,\ \pi\}$，其中 A，B 隱含地包括參數 N 和 M。

2. HMM 關鍵問題與演算法

建構出一個 HMM 後，需面對的三個關鍵問題 (Rabiner, 1989; Dymarski, 2011)：

(1) 評估問題：假設參數 λ 已知，計算 $P(O \mid \lambda)$ 。

(2) 編碼問題：假設參數 λ 和觀察值序列 O 已知，透過計算最大的 $P(Q \mid O,\ \lambda)$，求出 HMM 背後馬可夫鏈最有可能的狀態轉移序列 $Q = \{q_1,\ q_2, \ldots,\ q_T\}$。

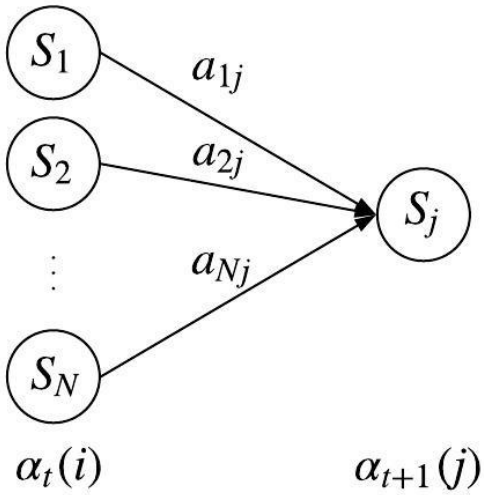
(3) 學習問題：建構完整的 HMM 模型 $\lambda = \{A,\ B,\ \pi\}$ 。根據已知的觀察值序列 O，透過計算最大的 $P(O \mid \lambda)$，找出最佳的 λ 參數模型。

在實務上，第一步驟會先利用 (3) 學習問題，以訓練資料來建構 HMM 模型，第二步驟再將第一步驟所訓練的 HMM 模型帶入 (1) 評估問題，根據測試資料的結果來判斷模型的好壞。對於詐欺偵測系統，我們關心系統是否能夠有效地找出詐欺交易和降低錯誤預警的機率，因此我們討論關鍵問題 (1) 和 (3)，也就是評估和學習問題。

2A 評估問題

評估問題可以用向前演算法 (forward algorithm) 和向後演算法 (backward algorithm) 解決。先給定觀察序列 $O = \{O_1,\ O_2, \ldots,\ O_T\}$ 和狀態序列 $Q = \{q_1,\ q_2, \ldots,\ q_T\}$，假設模型參數 λ 已知，目標為找出 $P(O \mid \lambda)$，且 $P(O \mid \lambda) = \sum_{\forall Q} P(O \mid Q,\ \lambda)$。

向前演算法



圖六、向前變數

定義向前變數 (forward variable) $\alpha_t(i)$，為給定 HMM 模型下，求在時間點 t 所發生的狀態為 S_i 且時間點 t 以前所看到的觀察值序列為 $O = \{O_1, O_2, \ldots, O_t\}$ 的機率，表示為

$$\alpha_t(i) = P(O_{1:t},\ q_t = S_i \mid \lambda) \tag{4}$$

- 初始化 (Initialization)：

$$\alpha_1(i) = P(O_1,\ q_1 = S_i \mid \lambda) = \pi_i b_i(O_1), \quad 1 \le i \le N \tag{4.1}$$

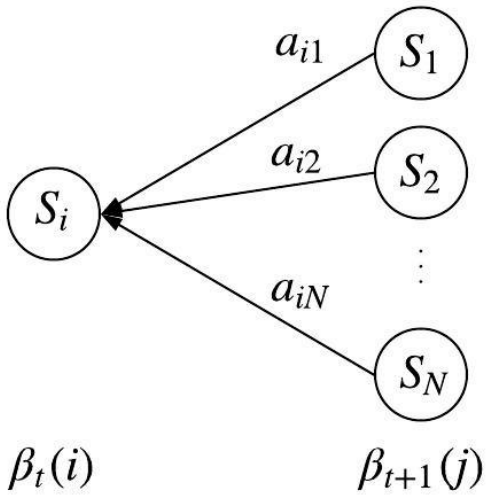
- 遞歸 (Recursion)：

$$\alpha_{t+1}(j) = P(O_{1:t+1},\ q_{t+1} = S_j \mid \lambda) = \sum_{i=1}^{N} b_j(O_{t+1}) a_{ij} \alpha_t(i)$$

$$1 \le i \le N, \quad 1 \le t \le T-1 \tag{4.2}$$

$$P(O \mid \lambda) = \sum_{i=1}^{N} \alpha_T(i) \tag{5}$$

向後演算法



圖七、向後變數

定義向後變數 (backward variable) $\beta_t(i)$，為給定 HMM 模型參數 λ 已知下且在時間點 t 所發生的狀態為 S_i，求時間點 t 以後所看到的觀察值序列為 $O = \{O_{t+1},\ O_{t+2}, \ldots,\ O_T\}$ 的機率，表示為

$$\beta_t(i) = P(O_{t+1:T} \mid q_t = S_i,\ \lambda) \tag{6}$$

- 初始化：

$$\beta_T(i) = 1, \quad 1 \le i \le N \tag{6.1}$$

- 遞歸：

$$\beta_t(i) = P(O_{t+1:T} \mid q_t = S_i,\ \lambda) = \sum_{j=1}^{N} b_{t+1}(j)\beta_{t+1}(j)a_{ij}$$

$$t = T-1,\ T-2, \ldots, 1, \quad 1 \le j \le N \tag{6.2}$$

$$P(O \mid \lambda) = \sum_{i=1}^{N} \alpha_t(i)\beta_t(i) \tag{7}$$

2B 學習問題

學習問題可以用 Baum-Welch 演算法 (也就是 EM 演算法) 解決。假設觀察序列 $O = \{O_{t+1},\ O_{t+2}, \ldots,\ O_T\}$已知，透過 $\max_{\lambda} P(O \mid \lambda)$，找出最佳化的 HMM 模型參數 $\lambda = \{A,\ B,\ \pi\}$。

Baum-Welch 演算法

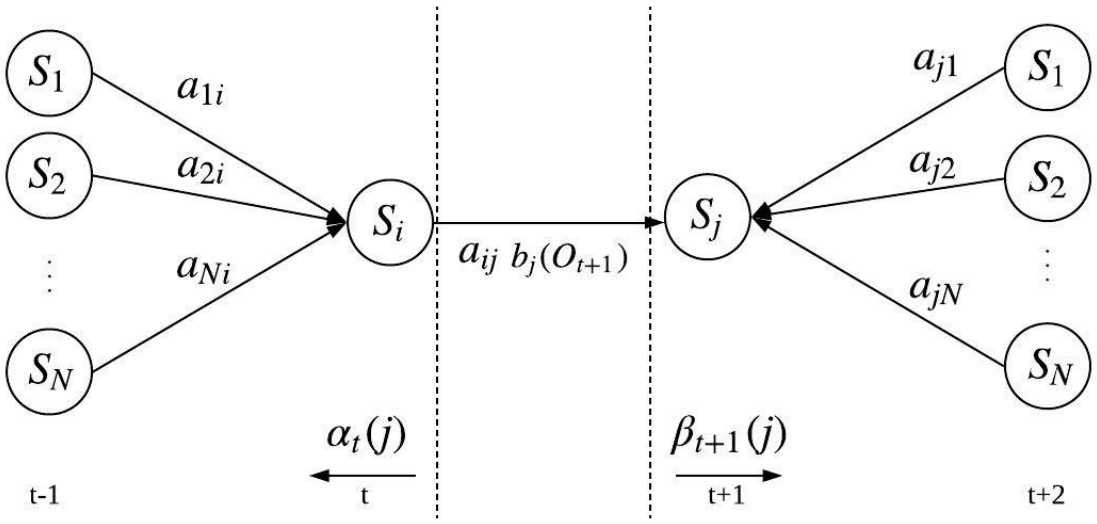
先定義兩個變數分別為 $\gamma_t(i)$ 和 $\xi_t(i,j)$ (即向前向後變數)，$\gamma_t(i)$ 為給定 HMM 模型參數 λ 且 $O = \{O_{t+1},\ O_{t+2}, \ldots,\ O_T\}$ 已知下，求時間點 t 所發生的狀態為 S_i 的機率。表示為

$$\gamma_t(i) = P(q_t = S_i \mid O,\ \lambda), \quad 1 \le i \le N \tag{8}$$

可由前文定義的向前機率 $\alpha_t(i)$ 和向後機率 $\beta_t(i)$ 表示

$$\gamma_t(i) = \frac{\alpha_t(i)\beta_t(i)}{\sum_{i=1}^{N} \alpha_t(i)\beta_t(i)} \tag{9}$$

其中 $\sum_{i=1}^{N} \gamma_t(i) = 1$。



圖八、向前向後變數

$\xi_t(i,j)$ 同樣給定 HMM 模型參數 λ 且 $O = \{O_{t+1},\ O_{t+2}, \ldots,\ O_T\}$ 已知下，求時間點 t 所發生的狀態為 S_i 且時間點 t+1 所發生的狀態為 S_j 的機率。表示為

$$\xi_t(i,j) = P(q_t = S_i,\ q_{t+1} = S_j \mid O,\ \lambda), \quad 1 \le i,\ j \le N, \quad 1 \le t \le T-1 \tag{10}$$

同樣可由向前機率 $\alpha_t(i)$ 和向後機率 $\beta_t(i)$ 表示

$$\xi_t(i,j) = \frac{\alpha_t(i) a_{ij} b_j(O_{t+1}) \beta_{t+1}(j)}{\sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_t(i) a_{ij} b_j(O_{t+1}) \beta_{t+1}(j)} \tag{11}$$

其中 $\gamma_t(i) = \sum_{j=1}^{N} \xi_t(i,j)$。

將 HMM 的隱藏狀態序列 Q 作為無法觀察的數據，可以得到一個具有隱藏變數的機率模型 $P(O \mid \lambda) = \sum_{Q} P(Q \mid O,\ \lambda) P(O \mid \lambda)$。學習問題的目標為找到一個最佳化的 HMM 模型參數，因此令 λ 為當前 HMM 模型的參數，$\bar{\lambda}$ 為使 HMM 模型最佳化的參數。用式子表示 $P(O \mid \bar{\lambda}) \ge P(O \mid \lambda)$，代表最佳化的參數 $\bar{\lambda}$ 與當前的參數 λ 相比有較大的機率預估觀察值序列。下文為 EM 演算法在 HMM 模型參數上的具體實現。

E-步驟 (E-step)：

$$Q(\lambda,\ \bar{\lambda}) = \sum_{Q} P(Q \mid O,\ \lambda) log[P(O,\ Q \mid \bar{\lambda})] \tag{12}$$

M-步驟 (M-step)：

$$\bar{\lambda} = \arg\max_{\lambda} Q(\lambda,\ \bar{\lambda})$$

不斷重複上述兩個步驟，直到 $P(O \mid \bar{\lambda}) < P(O \mid \lambda)$ 停止。而極大化函數 $Q(\lambda, \bar{\lambda})$ 所得的 HMM 模型參數，分別可用 $\gamma_t(i)$ 和 $\xi_t(i,j)$ 表示，表示公式如下

$$\bar{a}_{ij} = \frac{\sum_{t=1}^{T-1} \xi_t(i,j)}{\sum_{t=1}^{T-1} \gamma_t(i)} \tag{13}$$

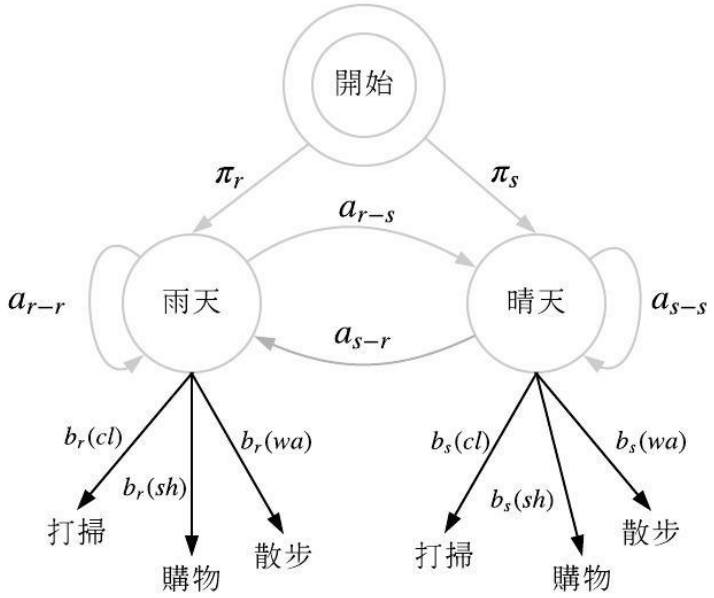
$$\bar{b}_j(O_k) = \frac{\sum_{t=1,\ O_t=v_k}^{T} \gamma_t(j)}{\sum_{t=1}^{T} \gamma_t(j)} \tag{14}$$

$$\bar{\pi}_i = \gamma_1(i) \tag{15}$$

3. 一個 HMM 的生活示例

假設你有一個住在遠方的朋友，你們每天都會聯絡，朋友會告訴你，他每天的的活動，他每天的活動可以歸納為三個基本行為：清潔房屋、外出購物和公園散步。你對朋友居住地方的天氣，並不了解，但是你認為天氣有兩個經典狀態，分別為雨天和晴天。因此你只能觀察到朋友每天的活動，卻不知道當天該地的天氣。 (故事背景節錄自網路)。

將天氣作為狀態，朋友每天的活動為一個觀察序列，由下面圖八表示，狀態集合為一個天氣序列 S ={雨天 (r)，晴天 (s)}；觀察值集合為一個朋友活動行為的觀察序列 V ={打掃 (cl)，購物 (sh)，散步 (wa)}，便構成一個完整的 HMM。



圖九、HMM 天氣行為示例

將以上狀態初始、轉移以及輸出給予機率所得參數 $\lambda = \{A,\ B,\ \pi\}$ 如下

- 初始機率分布：

$$\pi^{\mathrm{T}} = \begin{array}{c} \\ \end{array} \begin{array}{cc} r & s \\ \end{array} \begin{bmatrix} 0.6 & 0.4 \end{bmatrix}$$

- 轉移機率矩陣：

$$A = \begin{array}{c} r \\ s \end{array} \begin{array}{c} \begin{array}{cc} r & s \end{array} \\ \begin{bmatrix} 0.7 & 0.3 \\ 0.4 & 0.6 \end{bmatrix} \end{array}$$

- 輸出矩陣：

$$B = \begin{array}{c} r \\ s \end{array} \begin{array}{c} \begin{array}{ccc} wa & sh & cl \end{array} \\ \begin{bmatrix} 0.1 & 0.4 & 0.5 \\ 0.6 & 0.3 & 0.1 \end{bmatrix} \end{array}$$

第三章　應用隱藏式馬可夫模型至詐欺偵測系統

FDS 由信用卡發卡銀行進行運作，每筆收到的信用卡交易都會提交給 FDS 進行檢驗。FDS 於收到卡片信息跟交易金額後，立即檢驗交易是否正常。FDS 無法得知持卡人所購買的商品類別，它會根據持卡人的消費檔案、帳單地址和送貨地址等，試圖尋找出異常訊息。如果 FDS 確認某筆交易是惡意的，系統會發出警報，同時發卡銀行拒絕授權該筆交易，該筆交易的持卡人會被聯繫，提醒卡片可能已遭盜用。

本章解釋如何將 HMM 使用到信用卡欺詐檢測，以下分別就觀察符號、隱藏狀態、機率矩陣、詐欺偵測、及系統效能評估來說明。

第一節　觀察符號

FDS 必須從卡片信息、交易金額、消費檔案、帳單地址和送貨地址等資料，尋找出可能的異常訊息。在實體卡片被竊或遭側錄，虛擬卡片個資遭盜用的情況下，許多資訊是無法辨識真偽。然而，持卡人通常會有特定的消費行為，所以每一位持卡人都有個人專屬的消費檔案 (Spending Profile)，記錄著購買的時間、消費金額等。因此，將 HMM 應用到信用卡交易過程時，我們將交易金額視為 HMM 的觀察符號。為方便分析，進一步將交易金額量化成交易金額範圍。對於每一位持卡人，根據過去交易，我們使用適當的分類器 (例如，K-means 聚類演算法)，將交易金額劃分為幾個聚類，例如形成 M 個價格帶，$V_1,\ V_2, \ldots,\ V_M$。也就是說，每個觀察符號代表的交易金額範圍是根據該持卡人的交易習慣而架構出來的。當 FDS 收到一筆新交易後，便可以將該筆新交易金額對應到 M 個價格帶的其中一個。

Srivastava et al. (2008) 考慮 $M = 3$ 個交易金額範圍：低 (l)、中 (m) 和高 (h)。此時觀察符號集合記為 $= \{V_1 = l,\ V_2 = m,\ V_3 = h\}$ 。舉例來說，某一持卡人的三個金額範圍分別為：$l = (0, 150)$、$m = (150, 500)$ 和 $h = (500,$ 信用卡上限)，如果該名持卡人有一筆新的交易，金額為 210 元，那麼此筆交易其對應的觀察值符號為 $V_2 = m$。本論文參考 Srivastava et al. (2008) 的作法，將交易

金額範圍分為 $M = 3$ 及 $M = 4$。

第二節　隱藏狀態

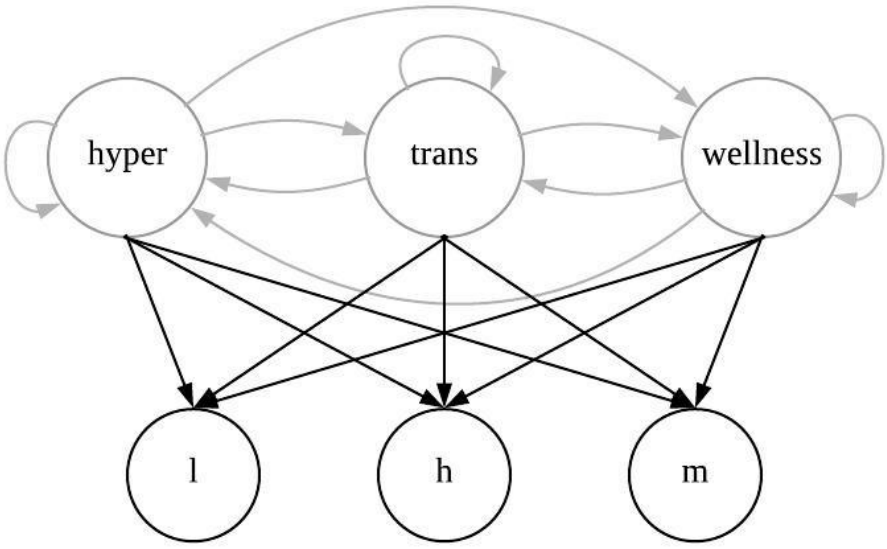
當交易金額量化為觀察符號 (即交易金額範圍) 後，持卡人所產生的交易序列，便以觀察符號序列來進行分析，尋求任何異常交易的線索。找出異常交易的一種可能性是尋找觀察符號序列中偏差的交易金額。Srivastava et al. (2008) 認為交易金額的產生源於持卡人所購買的商品類型，例如電子產品、雜貨、汽車零件等。相較於觀察符號序列，消費類別序列更加穩定，這是因為持卡者的消費行為，取決於持卡者在該段時間內對於不同商品類別的需求，進而生成一系列的交易金額。由於持卡者在交易時，發卡銀行並無法得知商家的業務範圍，也就是持卡人的消費類別對 FDS 是隱藏的，因此 Srivastava et al. (2008) 將消費類別的轉換視為 HMM 的狀態轉移。HMM 的隱藏狀態，由所有可能的消費類別，相當於所有可能的商家業務範圍之集合所形成的集合。此外，某些商家可能不僅提供單一類別的商品 (例如：誠品書店，好市多或小米銷售多種不同類別的商品)。任何關於消費類別已知的假設，運用在發卡銀行的 FDS 中，是不符和實際狀況，並且不具備有效性。

不同於 Srivastava et al. (2008)，他們假設隱藏狀態 $N =$ 消費類別，本論文假設隱藏狀態 $N = 2$：真實交易及詐欺交易。持卡人根據對不同商品的需求，也就產生出不同金額的交易，我們將真實交易金額範圍分類為 $M = 3$ 或 $M = 4$ 的觀察符號。對於欺詐交易，過去專家建議有兩種欺詐類型：一種類型是多次低價購買日常用品，另一種類型是單次高價購買電子產品。對於欺詐交易金額範圍的分類，一個可能性是採用 $M = 2$ 的觀察符號，然而，詐欺手法日新月異，分別詐欺交易與真實交易的難度增加，因此，對於詐欺交易與真實交易，我們採取一致的金額範圍的分類，在第四章中不論交易類型我們將交易金額範圍分類為 $M = 3$ 或 $M = 4$ 的觀察符號。

第三節　機率矩陣

在確定隱藏狀態和觀察符號後，下一步是決定 HMM 中三個機率矩陣：轉移機率矩陣 (A)、輸出矩陣 (B) 和初始機率 (π)。這三個模型參數是在訓練階段使用 Baum-Welch 演算法決定 (Baum et al, 1970; Rabiner, 1989)，其中參數的最初選擇可能會影響整體演算法的表現 (Srivastava et al., 2008)。我們假設隱藏的馬可夫鏈是遍歷的 (Ergodic)，就是模型中的每個狀態可以經由一步的轉移到達其他狀態。

圖十表示持卡人的消費概況，HMM 的隱藏狀態被分類為雜貨 (hyper)、交通 (trans) 和健康 (wellness)，並且分類器 (K-means，聚類演算法) 將交易價格分類為低價 (l)、中價 (m) 和高價 (h) 三種觀察符號。在每個隱藏狀態下，購買的產品價格會落入某一個觀察符號 (輸出概率)，經由持卡人的消費類別概況我們可以取得輸出矩陣 (B) 的起初估計。一個消費類別接著另一個消費類別的機率則是 HMM 的狀態轉移機率。此外，總是存在與購買類別相關聯的初始 (狀態) 機率，至少我們可以採用離散均勻分配。如此經由訓練，使用 Baum-Welch 演算法來確定 HMM 模型參數。一旦持卡人的 HMM 建立後，就可以使用它找出詐欺交易。



圖十、HMM 持卡人消費行為示例

第四節　詐欺偵測

在 HMM 參數學習後，我們採用了來自持卡人訓練資料中觀察符號，並且形成一個窗口大小 (window size) R 的基礎觀察符號序列，此序列記為 $O_1,\ O_2, \ldots,\ O_R$。這組序列其接受的機率為

$$\alpha_1 = P(O_1,\ O_2, \ldots,\ O_R \mid \lambda)$$

其中型 $\lambda = \{A,\ B,\ \pi\}$。當持卡人下一筆交易 O_{R+1} 產生，我們捨棄基礎序列中的第一筆觀察 O_1 並將這筆新交易 O_{R+1} 附加到序列中，形成新的觀察序列 $O_2,\ O_3, \ldots,\ O_{R+1}$，如此我們滑動窗口並維持窗口大小 R。將這個新序列 $O_2,\ O_3, \ldots,\ O_{R+1}$ 輸入到 HMM 並計算出它的接受機率

$$\alpha_2 = P(O_2,\ O_3, \ldots,\ O_{R+1} \mid \lambda)$$

檢視這兩組序列接受機率的差異 $\Delta\alpha = \alpha_1 - \alpha_2$，如果 $\Delta\alpha > 0$ 意味著新序列被 HMM 接受的機率低於前一個序列，因此最新的交易 O_{R+1} 是一個潛在的欺詐交易，在這種情況下，發卡銀行拒絕該筆交易，FDS 丟棄 O_{R+1} 。相反地，如果 $\Delta\alpha < 0$ 則意味著新序列被 HMM 接受機率較前一個序列高，因此最新的交易 O_{R+1} 極可能是真正的持卡人交易，在這種情況下，FDS 添加 O_{R+1} 到基礎序列並同時丟棄 O_1，新序列 $O_2,\ O_3, \ldots,\ O_{R+1}$ 成為基礎序列，用來確定下一筆交易的有效性。

當一筆真正的交易被添加到觀察序列 O，從而更新建模序列，能反映出 (或學習到) 持卡人不斷變化的消費行為。這個方法是假設單筆交易可能是詐欺，模型不識別欺詐序列，而是識別添加到有效序列的單筆欺詐交易。除了單一持卡人情形，在第四章我們也討論多筆欺詐交易的情形，也就是機構情形。

一筆新交易是否被確定為欺詐，可以根據前後兩組序列接受機率進一步量化。Robinson and Aria (2018) 使用 Kullback-Leibler 演算法，計算兩組序列接受機率的差異；Srivastava et al. (2008) 考慮機率百分比變化是否高於閥值 (threshold)，也就是選擇適當的閥值定出

$$\frac{\Delta\alpha}{\alpha_1} \geq Threshold$$

對於單一持卡人，窗口大小 R 一般設為 5 至 25 (Srivastava et al., 2008)。

第五節　系統效能評估

為衡量 FDS 的效能，我們採取三個常用的指標進行判斷：真陽率 (True Positive Rate, TPR)、偽陽率 (False Positive Rate, FPR)、及整體準確性 (Overall Accuracy, OA)。這裡的陽指的是詐欺。一個有效率的 FDS 通常應具有高真陽率和整體準確性，和低偽陽率。這些指標通常用於詐欺檢測研究，參見 Santos and Ocampo (2018)。TPR 和 FPR 指標的計算方法如下：

- TPR 真實交易為詐欺且分類為詐欺的比例

$$= \frac{TP}{\text{真實為詐欺的交易總數}} = \frac{\text{真實交易為詐欺且分類為詐欺的數量}}{\text{真實為詐欺的交易總數}}$$

- FPR 真實交易為正常但分類為詐欺的比例

$$= \frac{FP}{\text{真實為正常的交易總數}} = \frac{\text{真實交易為正常但分類為詐欺的數量}}{\text{真實為正常的交易總數}}$$

此外，真陰率 (True Negative Rate, TNR)、偽陰率 (False Negative Rate, FNR) 計算方法如下：

- TNR 真實交易為正常且分類為正常的比例

$$= \frac{TN}{\text{真實為正常的交易總數}} = \frac{\text{真實交易為正常且分類為正常的數量}}{\text{真實為正常的交易總數}}$$

- FNR 真實交易為詐欺但分類為正常的比例

$$= \frac{FN}{\text{真實為詐欺的交易總數}} = \frac{\text{真實交易為詐欺但分類為正常的數量}}{\text{真實為的詐欺交易總數}}$$

偽陰率高，便無法有效降低發卡銀行的詐欺交易損失 (詐欺抓不到)；偽陽率高，導致發卡銀行過度進行控卡 (到處是詐欺)，使持卡人感到麻煩，嚴重會影響發卡銀行商譽。

- OA

$$= \text{正確類的比例} = \frac{TP + TN}{\text{全部交易總數}}$$

$$= \frac{\text{真實交易為詐欺且分類為詐欺的數量+真實交易為正常且分類為正常的數量}}{\text{全部交易總數}}$$

例如：

100 筆真正的詐欺交易被分類為： 90 筆詐欺、10 筆正常

900 筆真正的正常交易被分類為：870 筆正常、30 筆詐欺

$$TPR = \frac{90}{100} = 0.9 \qquad FPR = \frac{30}{900} = 0.033$$

$$TNR = \frac{870}{900} = 0.967 \qquad FNR = \frac{10}{100} = 0.1$$

$$OA = \frac{90 + 870}{1000} = \frac{960}{1000} = 0.96$$

第四章　資料模擬

第一節　資料集

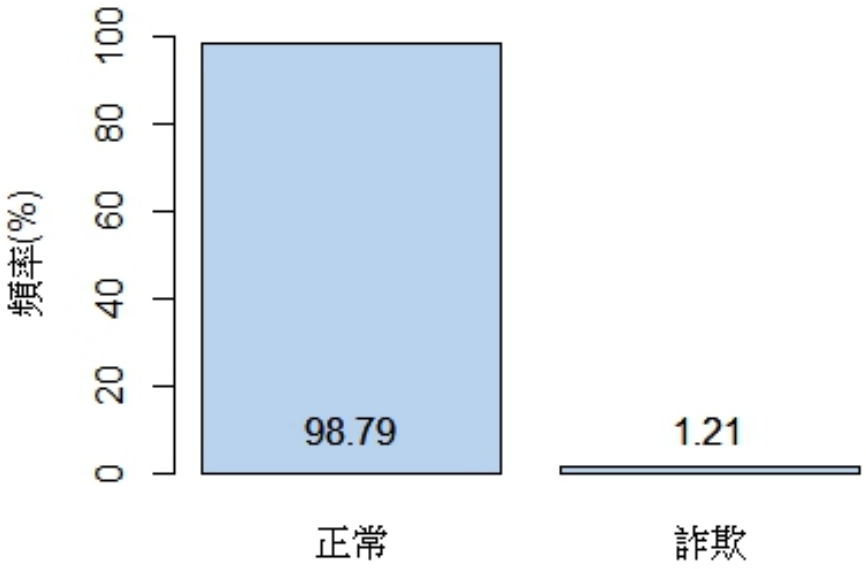
使用真實的數據來測試信用卡 FDS 是一項困難的任務。迄今文獻中雖然有不少關於信用卡詐欺偵測的文章，然而，一方面基於保護消費者資訊，金融單位不願意分享資訊，另一方面也無法得到可用於實驗的基準數據集 (Benchmark Dataset)，因此幾乎沒有文章是使用真實的數據來測試他們所提出的模型。

本論文使用 BankSim 支付模擬器所生成的數據集 (Lopez-Rojas and Axelsson, 2014)。BankSim 模擬器根據西班牙的某家銀行的信用卡刷卡交易數據模擬資料，該筆刷卡交易資料彙整了該家銀行的信用卡持卡人在 2012 年 11 月到 2013 年 4 月期間，於馬德里和巴塞隆納所進行的交易資料。BankSim 產生一組趨近於真實信用卡刷卡的模擬數據，這組數據共有 594,643 筆交易，其中 7,200 筆為詐欺交易。每筆交易都提供刷卡金額、交易是否為詐欺、原產地/ 來源的郵區編號位置、消費日期、商店代號、消費類別以及客戶編號、性別和年齡等資訊。詐欺交易佔所有交易的 1.21 %，是典型的不平衡數據 (圖十一)。消費類別共有 15 種，表一為 BankSim 持卡人各個消費類別占全部刷卡交易的百分比。

使用 BankSim 資料，我們驗證前面章節所描述的 FDS。由於 FDS 的主要目的是找出詐欺交易，在本章的第二節和第三節中我們將分別考慮兩種模型的詐欺偵測：(1) 單一持卡人模型，及 (2) 機構模型。稍後在第四節裡我們討論單一持卡人與機構詐欺交易之間的關係。

表一、BankSim 持卡人交易類別及百分比

編號	交易類別	頻率(%)	編號	交易類別	頻率(%)
1	barsandrestaurants	1.07	9	leisure	0.08
2	contents	0.15	10	otherservices	0.15
3	fashion	1.09	11	sportsandtoys	0.67
4	food	4.42	12	tech	0.4
5	health	2.71	13	transportation	84.94
6	home	0.33	14	travel	0.12
7	hotelservices	0.29	15	wellnessandbeauty	2.54
8	hyper	1.03			



圖十一、正常詐欺交易比例長條圖

第二節　詐欺偵測系統：單一持卡人

參數選擇：

(1) 隱藏狀態：$N = 2$ (真實交易、詐欺交易)。

(2) 觀察符號：使用 K-means 聚類演算法將交易金額範圍分為 $M = 3$ (低、中、高)、$M = 4$ (低、中低、中高、高)。

(3) 窗口大小(觀察值序列長度)：$R = 8$、10、12 筆交易

(4) 閥值：10 % 至 90 % (間隔 10 %)

對於單一持卡人分析步驟如下：

步驟 1. 首先由 BankSim 資料中篩選出單一持卡人的交易資訊，我們找出購買次數相對多數的持卡人，例如在 BankSim 資料中共有 166 位持卡人的交易次數為衆數 172 次，選出編號 C125481968 的持卡人的資料。將選出的每一位持卡人都建立一個 HMM。

步驟 2. 對於每一位持卡人，我們使用該持卡人的前 R 筆交易當作訓練資料，代入第二章的 HMM 中，計算出模型參數的初始值 $\lambda = \{A,\ B,\ \pi\}$ (其中π 是初始機率、A 是轉移機率矩陣、及 B 是輸出矩陣)，並且計算出起始觀察序列 $O_1,\ O_2, \ldots,\ O_R$ 的接受機率

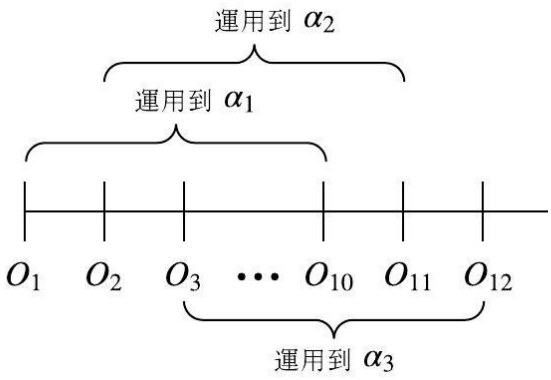
$$\alpha_1 = P(O_1,\ O_2, \ldots,\ O_R \mid \lambda)$$

步驟 3. 加入一筆新交易資料 O_{R+1}，計算出觀察序列 $O_2,\ O_3, \ldots,\ O_{R+1}$ 的接受機率

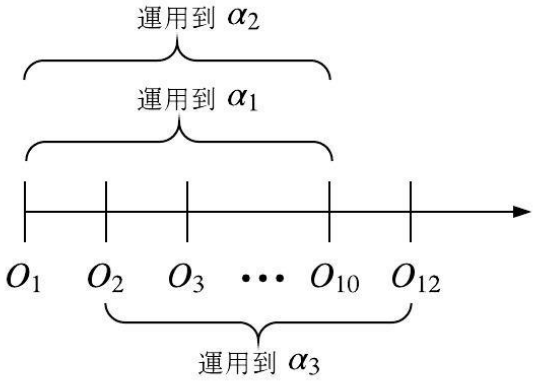
$$\alpha_2 = P(O_2,\ O_3, \ldots,\ O_{R+1} \mid \lambda)$$

，並且計算出 $\Delta\alpha = \alpha_1 - \alpha_2$，然後比較 $\frac{\Delta\alpha}{\alpha_1}$ 與閥值以分辨新交易 O_{R+1} 是否為詐欺交易。如果 O_{R+1} 是詐欺交易則刪除該筆資料，不再投入新的 HMM 模型中；如果 O_{R+1} 不是詐欺交易則加入觀察序列中，並且將第一筆資料 O_1 由觀察序列中去除。加入新交易到觀察序列中，能夠持續不斷的學習持卡人的使用習慣。圖十二及圖十三為 $R = 10$ 時的運作釋義圖。

步驟 4. 我們先進行觀察符號 $M = 3$ 的系統分析，再進行 $M = 4$ 的系分析。透過上面參數的設置，變化持卡者交易的窗口大小 R 值 (R= 8、10、12)，計算出不同閥值下的 TPR 和 FPR，將看到在變化持卡者交易窗口大小時，HMM 的表現。



圖十二、當 O_{11} 被視為正常交易，HMM 運作釋義圖



圖十三、當 O_{11} 被視為詐欺交易，HMM 運作釋義圖

表三~表七、表八~表十二和表十三~表十七分別為持卡人編號 C125481968、C944695695 和 C73919470 在不同閥值的 TPR、FPR 和 OA 的分析 (表格內的數字代表 %)。圖十四~圖十九則是這三位持卡人的 TPR 及 FPR 的折線圖。這三位持卡人的交易筆數和詐欺數量如下：

表二、三位持卡人的交易筆數和詐欺數量

持卡人編號	交易筆數	詐欺數量
C125481968	172	1
C944695695	171	4
C73919470	170	1

對於持卡人 C125481968 的分析，表三~表五分為 (M=3, R=10)，(M=3, R=8)，及 (M=3, R=12)。表六為 (M=3, R=10，間隔 2)，變換第三步驟的一次增加和刪除一筆資料，轉為一次增加和刪除兩筆資料，也就是由一次移動兩格取代一次移動一格，目的在測試改變移動格數是否能提高表三的效能。表七則是 (M=4, R=10)。

持卡人 C125481968 在不同參數下的 TPR、FPR 和 OA 表二~表六：

表三、C125481968 不同閥值的 TPR、FPR 和 OA ($M = 3,\ R = 10$)

閥值	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
TPR	100	100	100	100	100	100	100	100	100
FPR	29.193	29.193	29.193	29.193	29.193	29.193	29.193	29.193	17.391
OA	70.988	70.988	70.988	70.988	70.988	70.988	70.988	70.988	82.716

表四、C125481968 不同閥值的 TPR、FPR 和 OA ($M = 3,\ R = 8$)

閥值	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
TPR	100	100	100	100	100	100	100	100	100
FPR	30.675	30.675	30.675	30.675	30.675	30.675	28.221	28.221	20.245
OA	69.512	69.512	69.512	69.512	69.512	69.512	71.951	71.951	79.878

表五、C125481968 不同閥值的 TPR、FPR 和 OA ($M = 3,\ R = 12$)

閥值	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
TPR	100	100	100	100	100	0	0	0	0
FPR	60.377	59.748	59.748	59.748	59.748	49.686	49.686	10.692	8.805
OA	40	40.625	40.625	40.625	40.625	50	50	88.75	90.625

表六、C125481968 不同閥值的 TPR、FPR 和 OA ($M = 3,\ R = 10$，間隔 2)

閥值	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
TPR	100	100	100	100	100	100	100	100	100
FPR	46.583	44.099	44.099	44.099	42.857	41.615	41.614	41.614	37.267
OA	53.704	56.173	56.173	56.173	57.407	58.642	58.642	58.642	62.346

表七、C125481968 不同閥值的 TPR、FPR 和 OA ($M = 4,\ R = 10$)

閥值	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
TPR	100	100	100	100	100	100	100	100	100
FPR	98.137	90.683	90.683	90.683	93.789	93.789	73.913	86.957	47.826
OA	2.469	9.259	9.259	9.259	6.790	6.790	25.926	13.580	52.469

下面的討論中我們先找出高比例的 TPR，然後再找出低比例的 FPR，及高比例的 OA。

編號 C125481968 持卡人只有一筆詐欺交易，由表三至表七可知，FDS 在不同閥值下大多能有效偵測出該筆詐欺 (TPR=100 %)，除了表五 (M=3, R=12) 閥值 0.6 ∼0.9 時無法偵測出來。當閥值愈高 FPR 就愈低，也就是閥值愈高判定正常交易為詐欺交易的可能性愈低，其中表三 (M=3, R=10) 和表五 (M=3, R=12) 的 FPR 表現較好，表三 (M=3, R=10) 的 FPR 在閥值 0.9 時達到最低的 17.391 %。此外 OA 隨著閥值增加而增加，表三 (M=3, R=10) 的各個 OA 值均在 70 %以上表現最好，在閥值 0.9 時達到最高的 82.716 %。因此對編號 C125481968 持卡人而言，(M=3, R=10) 且閥值=0.9 是最佳的 FDS。

編號 C944695695 持卡人則有四筆詐欺交易，重複先前步驟，得到表八∼表十二。除了表十 (M=3, R=12) 外，FDS 能有效偵測出詐欺交易的最高比例是 TPR=75 %。FPR 隨著閥值增加而遞減，其中表十二 (M=4, R=10) 的 FPR 在閥值 0.9 時達到最低的 36.943 %。此外 OA 隨著閥值增加而增加，表十二 (M=4, R=10) 的的 OA 值在閥值 0.9 時達到最高的 63.354 %。因此對編號 C944695695 持卡人而言，(M=4, R=10) 且閥值=0.9 是最佳的 FDS。

重複先前步驟對編號 C73919470 持卡人進行分析得到表十三∼表十七，該持卡人有一筆詐欺交易，FDS 在不同閥值下都能有效偵測出該筆詐欺 (TPR=100 %)。FPR 隨著閥值增加而遞減，除了表十三 (閥值=0.4 時) 與表十七 (隨閥值先增後降)，其中表十五 (M=3, R=12) 的 FPR 在閥值 0.9 時達到最低的 24.204 %。此外 OA 隨著閥值增加而增加，表十五 (M=3, R=12) 的的 OA 值在閥值 0.9 時達到最高的 75.949 %。因此對編號 C73919470 持卡人而

言，(M=3, R=12) 且閥值=0.9 是最佳的 FDS。

持卡人 C944695695 在不同參數下的 TPR、FPR 和 OA 表八~表十二：

表八、C944695695 不同閥值的 TPR、FPR 和 OA ($M = 3,\ R = 10$)

閥值	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
TPR	75	75	75	75	75	75	75	75	50
FPR	58.599	58.599	58.599	57.962	57.962	54.140	54.140	51.592	23.567
OA	42.236	42.236	42.236	42.857	42.857	46.584	46.584	47.826	75.776

表九、C944695695 不同閥值的 TPR、FPR 和 OA ($M = 3,\ R = 8$)

閥值	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
TPR	75	75	75	75	75	75	75	75	50
FPR	55.975	55.346	55.346	55.346	54.088	54.088	45.283	45.283	41.509
OA	44.785	45.399	45.399	45.399	46.626	46.626	55.215	55.215	58.282

表十、C944695695 不同閥值的 TPR、FPR 和 OA ($M = 3,\ R = 12$)

閥值	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
TPR	75	75	75	75	75	50	25	25	0
FPR	60	59.355	59.355	58.065	52.258	52.258	52.258	52.258	29.677
OA	40.881	41.509	41.509	42.767	48.428	47.799	47.170	47.170	68.553

表十一、C944695695 不同閥值的 TPR、FPR 和 OA ($M = 3,\ R = 10$，間隔 2)

閥值	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
TPR	75	75	75	75	75	75	75	50	50
FPR	69.872	69.872	69.872	71.154	71.154	71.154	71.154	53.846	34.615
OA	31.25	31.25	31.25	30	30	30	30	46.25	65

表十二、C944695695 不同閥值的 TPR、FPR 和 OA ($M = 4,\ R = 10$)

閥值	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
TPR	75	75	75	75	75	75	75	75	75
FPR	60.510	59.873	59.873	59.873	57.325	57.325	57.325	59.236	36.943
OA	40.373	40.994	40.994	40.994	43.478	43.478	43.478	41.615	63.354

持卡人 C73919470 在不同參數下的 TPR、FPR 和 OA 表十三~表十七：

表十三、C73919470 不同閥值的 TPR、FPR 和 OA ($M = 3,\ R = 10$)

閥值	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
TPR	100	100	100	100	100	100	100	100	100
FPR	32.704	32.704	32.704	89.308	31.447	31.447	31.447	93.082	30.189
OA	67.5	67.5	67.5	11.25	68.75	68.75	68.75	7.5	70

表十四、C73919470 不同閥值的 TPR、FPR 和 OA ($M = 3,\ R = 8$)

閥值	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
TPR	100	100	100	100	100	100	100	100	100
FPR	0	95.031	95.031	95.031	93.789	90.062	45.342	45.342	44.099
OA	0.617	5.556	5.556	5.556	6.790	10.494	54.938	54.938	56.173

表十五、C73919470 不同閥值的 TPR、FPR 和 OA ($M = 3,\ R = 12$)

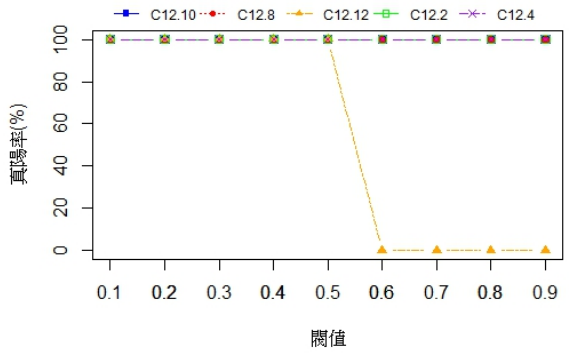
閥值	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
TPR	100	100	100	100	100	100	100	100	100
FPR	91.083	91.083	61.783	59.236	57.962	77.707	57.325	34.395	24.204
OA	9.494	9.494	38.608	41.139	42.405	22.785	43.038	65.823	75.949

表十六、C73919470 不同閥值的 TPR、FPR 和 OA ($M = 3,\ R = 10$，間隔 2)

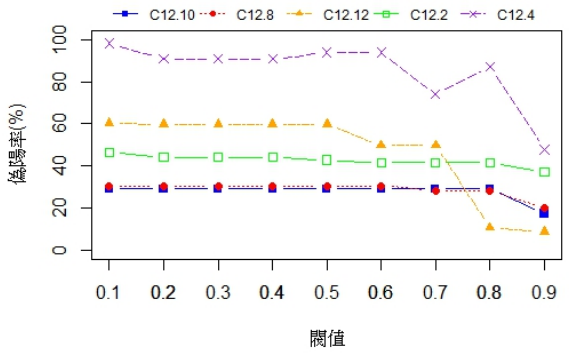
閥值	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
TPR	100	100	100	100	100	100	100	100	100
FPR	50.943	49.686	49.686	49.686	45.912	45.912	45.912	45.912	40.881
OA	49.375	50.625	50.625	50.625	54.375	54.375	54.375	54.375	59.375

表十七、C73919470 不同閥值的 TPR、FPR 和 OA ($M = 4,\ R = 10$)

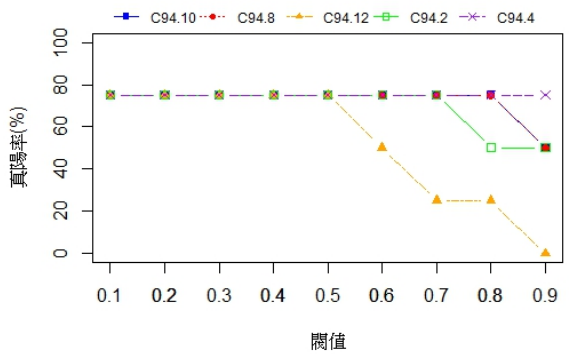
閥值	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
TPR	100	100	100	100	100	100	100	100	100
FPR	67.925	67.925	67.925	93.711	93.711	93.711	93.711	93.711	67.925
OA	32.5	32.5	32.5	6.875	6.875	6.875	6.875	6.875	32.5



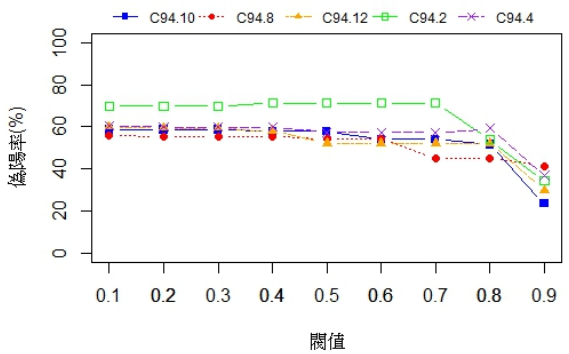
圖十四、C125481968 的 TPR 折線圖



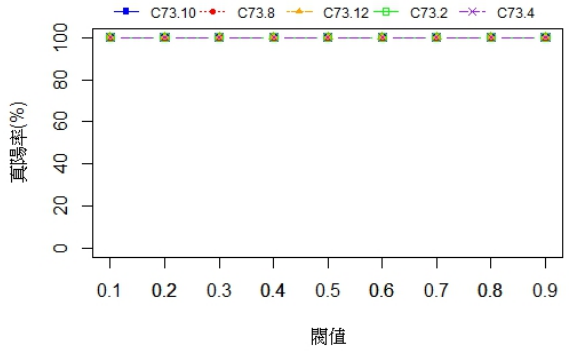
圖十五、C125481968 的 FPR 折線圖



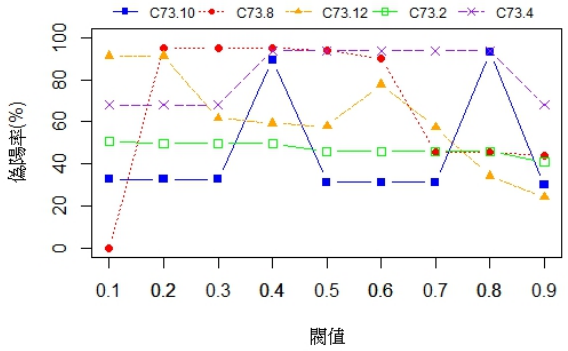
圖十六、C944695695 的 TPR 折線圖



圖十七、C944695695 的 FPR 折線圖



圖十八、C73919470 的 TPR 折線圖



圖十九、C73919470 的 FPR 折線圖

* C12.10 (M=3, R=10)， C12.8 (M=3, R=8)， C12.12 (M=3, R=12)， C12.2 (M=3, R=10，間隔 2)， C12.4 (M=4, R=10)

* C94.10 (M=3, R=10)， C94.8 (M=3, R=8)， C94.12 (M=3, R=12)， C94.2 (M=3, R=10，間隔 2)， C94.4 (M=4, R=10)

* C73.10 (M=3, R=10)， C73.8 (M=3, R=8)， C73.12 (M=3, R=12)， C73.2 (M=3, R=10，間隔 2)， C73.4 (M=4, R=10)

第三節　詐欺偵測系統：機構

單一持卡人為中心的交易，由於交易數量小，因此使用重疊的窗口來偵測欺詐交易，同時加入新交易到觀察序列中，能夠持續不斷的學習持卡人的使用習慣。然而以機構為中心的交易，由於交易數量龐大且有衆多持卡人，並且可能產生當真實行為遠離建模行為時的概念漂移 (concept drift) 現象 (例如偶發事件改變群衆的消費行為) (Robinson and Aria, 2018)。此外連續欺詐是另一個問題，一新交易序列可能是異常的，而其中的任何單個交易都是有效的，例如，商店的一項產品連續銷售超過 50 個，則可能發生欺詐。這些因素都讓詐欺規則變得複雜，增加系統執行的複雜度。因此為了檢視與分析基於商店行為的欺詐行為，我們使用非重疊窗口偵測欺詐交易。

不同於第四章第一節提取 BankSim 資料其中某一持卡人交易資訊，探討該名持卡人所有交易筆數套入 HMM 模型，不同閥值下的真陽率和偽陽率。此時將 BankSim 資料視為一個個體，針對的方向由挖掘持卡人與平時有所差異

的刷卡行為，轉為詐欺者盜刷信用卡的行為，認為盜刷行為會在某些時段集中且密集發生，猶如當詐欺者盜取持卡人卡片，會在某個特定時段不段的產生消費筆數，購買高單價商品。

基於上述，將正常交易與詐欺交易當作兩種不同的狀態，兩者之間的轉換視為一個馬可夫鏈，其中兩者發生的機率不同，而同一型態交易(正常或詐欺)呈現多筆連續交易情況，因此可視為一種馬可夫調控卜瓦松過程 (Markov modulated Poisson process)。短時間內對龐大的資料數目進行篩選，使用非重疊窗口做詐欺檢測，意味著不會將系統認定為正常的交易，投入系統學習成新的 MMPP，再對下一筆新的序列進行篩選。

下表十八 ∼ 表二十一，詐欺交易發生的數目固定為 2，訓練觀察值長分別為 25、50、75 和 100，每隔 25 筆做一個表。例如將 BankSim 資料的前 25 筆訓練一個 MMPP 模型，將 BankSim 資料 26 筆之後的資料，每隔 25 筆做一次詐欺檢測，當檢測的 25 筆和模型所預測的 25 筆都出現詐欺交易，便視為模型有檢驗出詐欺交易。意味著對每 25 筆信用卡交易做 1 次的快速篩檢，檢驗那 25 筆是否存在著詐欺交易。

表十八、MMPP 每隔 25 筆交易做一次偵測

正常交易發生數目	48	98	148	198	298	498
TPR	30.969	16.90	11.810	8.121	5.490	3.088
FPR	31.297	17.135	11.218	8.516	6.055	3.496
OA	63.153	73.163	77.461	79.224	80.935	82.765

表十九、MMPP 每隔 50 筆交易做一次偵測

正常交易發生數目	48	98	148	198	298	498
TPR	51.477	30.453	21.182	15.022	10.528	6.820
FPR	51.941	30.278	20.377	15.966	11.222	6.375
OA	48.974	59.211	63.980	65.562	67.833	70.390

表二十、MMPP 每隔 75 筆交易做一次偵測

正常交易發生數目	48	98	148	198	298	498
TPR	48.869	30.240	22.099	16.409	11.790	7.630
FPR	49.830	31.440	21.802	17.111	11.863	6.796
OA	33.938	38.602	41.649	42.491	44.080	45.727

表二十一、MMPP 每隔 100 筆交易做一次偵測

正常交易發生數目	48	98	148	198	298	498
TPR	76.729	49.762	36.187	27.662	21.186	11.416
FPR	75.304	50.888	36.718	29.274	19.713	13.018
OA	48.620	49.411	50.824	50.925	53.112	52.238

表二十二、不同間隔筆數的正常比例

間隔筆數	25	50	75	100
正常交易比例	85.293	73.234	62.924	54.021

從表二十二，可以看出間隔筆數的長短，會影響正常交易所佔的比例，意味著當間隔筆數越長時，詐欺交易所佔的比例會上升，間隔筆數為 100 時，正常和詐欺交易的比例甚至趨於平衡。

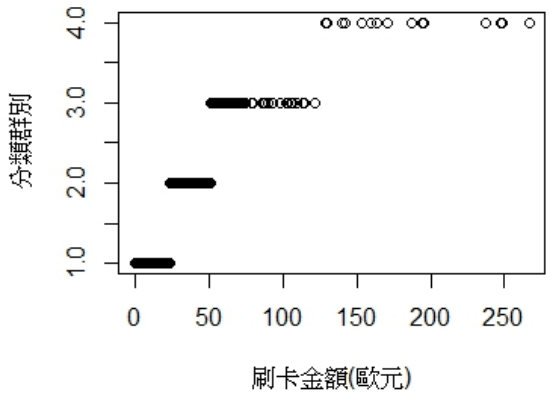
表十八~表二十一，MMPP 都能夠偵測出詐欺，但是較低的正常交易數目，能夠偵測出較多的詐欺數目，因為當正常交易數目下降，TPR 卻逐漸上升。而 FPR在正常交易數目多寡的走向和 TPR 一致，這也代表著能夠找出最多詐欺交易的模型，並不一定是個最佳的模型，基於在誤判正常交易為詐欺交易的比例相較於其他模型還高。OA 則和 TPR、FPR 相反，會隨著件隔筆數增加而上升。表十八~表二十一中，表十八的間隔比數 498 具有最佳的 OA 為 70.39 %，然而在抓取詐欺跟誤判正常交易的表現較為不佳。若要從中選擇一個平衡點，表十八每隔 25筆做一次偵測，在正常交易數目為 98 時，能夠找到詐欺交易、FPR 也不會過高，而 OA 也具有一定的水準 73.163 %。

第四節　詐欺偵測系統：機構與持卡人兩者間的關係

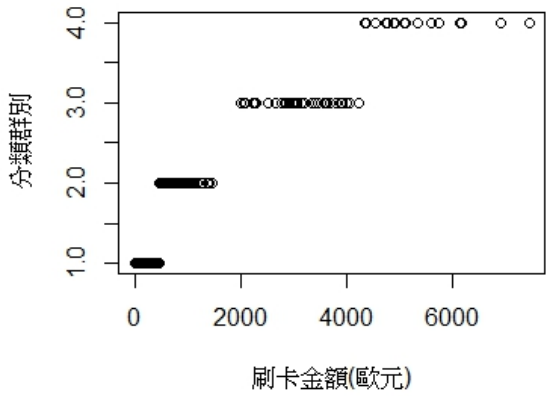
由表二十三，可以看出正常交易和詐欺交易大部分的消費類型相同，在 food (食物) 和 transportation (交通) 這兩種消費類型，沒有出現在詐欺交易的消費類型當中；相對的 leisure (休閒) 和 travel (旅遊) 這兩種消費類型，沒有出現在正常交易的消費類型當中。正常交易和詐欺交易的刷卡金額也存在差異，如圖十四和圖十五。詐欺交易的刷卡金額偏極大和極小，小筆金額出現在盜刷者想嘗試卡片是否可使用。如果當下持卡人和銀行端，沒有即時發現該筆交易為詐欺交易，便可能遭受到後續的損失。

表二十三、 BankSim 資料前 1000 筆正常和詐欺交易，各自的消費類型

正常交易	詐欺交易
barsandrestaurants	barsandrestaurants
fashion	fashion
food	health
health	home
home	hotelservices
hotelservices	hyper
hyper	**leisure**
otherservices	otherservices
sportsandtoys	sportsandtoys
tech	tech
transportation	**travel**
wellnessandbeauty	wellnessandbeauty



圖二十、前 1000 筆正常交易的 k-means 分群圖



圖二十一、前 1000 詐欺常交易的 k-means 分群圖

* 1 = l (低刷卡金額), 2 = lm (中低刷卡金額), 3 = lh (中高刷卡金額), 4 = h (高刷卡金額)

第五節　偵測方法間的比較

表二十四、三位持卡人在各偵測方法的表現

偵測方法 / 編號	HMM		ANN	
	TPR	FPR	TPR	FPR
C125481968	100	17.391	0	0
C944695695	50	23.567	0	0
C73919470	100	30.189	0	0
偵測方法 / 編號	KNN		SVM	
	TPR	FPR	TPR	FPR
C125481968	0	0	NA	NA
C944695695	0	0	NA	NA
C73919470	0	0	NA	NA

*表二十四數值單位為百分比(%)

表二十四為持卡編號 C125481968、C944695695 和 C73919470 在各種偵測方法間的比較。HMM 相較於 ANN 和 KNN 準確率偏低，SVM 則基於訓練的序列當中沒有詐欺交易存在，因而無法運作。 HMM 可以從資料中抓取詐欺交易，加上 BankSim 資料本身為一筆偏斜資料，詐欺交易筆數只占資料中的 1.21 %，能夠從持卡人的刷卡交易中當中偵測出詐欺交易，銀行端才能採取後續防治措施，減少損失。

第五章　結論與未來工作

互聯網的便利性和電子商務的創新，除了刷卡交易每年增長，詐欺交易發生的頻率及金額也隨之上升，發展一個良好且即時偵測的信用卡詐欺偵測系統，成為銀行端所關注的目標。本論文提出了 HMM 在信用卡欺詐檢測中的應用。文中使用 BankSim 模擬的資料，該筆資料接近真實的資料。分析 HMM 和其他偵測方法在數名持卡人的真陽率和假陽率表現。HMM 維持不錯的準確率和偵測詐欺交易，與其他方法比較，ANN 和 KNN 無法偵測出詐欺交易，將測試的資料都分為正常交易，SVM 沒有詐欺交易在訓練交易當中，模型無法運作。在針對單一持卡人進行詐欺偵測，HMM 確實能夠找出詐欺交易，有不錯的準確率。找到真正的信用卡交易資料帶入模型，依舊是努力的目標之一。HMM 確實能夠偵測到詐欺資料，但是相較於其他偵測方法，HMM 的準確率尚有提升的空間，或許可以嘗試從觀察值序列找出一個較佳的訓練模式，例如狀態之間的合併。亦或者發展一套方法，設定 HMM 的模型初始值。

HMM 在針對單一持卡人的交易中是否存在詐欺交易上，相較於其他模型，擁有不錯的 TPR，然而在 FPR 的表現上還存在著可以改善的部分。對於發卡銀行來說，無法偵測出詐欺交易會導致財務損失；相對的，將正常交易誤分到詐欺交易，當下停止授權該筆交易，則會犧牲客戶後續的生命週期價值。因此在提升 TPR 和降低 FPR 之間，存在一個如何使整體損失降低的公式。第四章的表三~表十七，呈現出三位持卡人在不同閥值和不同參數選擇的結果。在閥值的表現上，這些表在閥值為 0.9 時，對比於其他閥值擁有更佳的效能。然而在參數選擇上，這三位持卡人沒有達到一致參數的選擇，意味著這三位持卡人的最佳效能出現在不同的參數選擇之下。若後續將 BankSim 資料中所有持卡人的交易都試過一遍，則能夠找出大部分持卡人在哪一個參數選擇下擁有最佳效能，進而得到更完善的參數選取結果。除此之外，觀察符號所使用 K-means 聚類演算法將交易金額範圍分類，嘗試使用其他聚類演算法，或者是將交易金額範圍再多幾個分類，都會對 HMM 偵測詐欺的造成影響。

參考文獻

[1] 財金資訊股份有限公司業務說明信用卡業務網頁，服務流程。檢至 https://www.fisc.com.tw/tc/business/Detail.aspx?caid=08b6275d-19f1-495c-8fd6-749e558e4383(Dec. 12, 2018)

[2] 財團法人聯合信用卡處理中心中心年報 (2016)。105 年度中文年報。檢至 https://www.nccc.com.tw/wps/wcm/connect/9d778d09-69fd-4279-9835-d8b2ec4d72b9/105%E5%B9%B4%E5%A0%B1%E4%B8%AD%E6%96%87%E7%89%88.pdf?MOD=AJPERES&CACHEID=ROOTWORKSPACE-9d778d09-69fd-4279-9835-d8b2ec4d72b9-lu8-n6c (Dec. 12, 2018)

[3] Aleskerov, E., Freisleben, B., and Rao, B., (1997) CARDWATCH: A Neural Network Based Database Mining System for Credit Card Fraud Detection, Proc. IEEE/IAFE: Computational Intelligence for Financial Eng., pp. 220-226.

[4] Chen, R.C., Chiu, M.L., Huang, Y.L., Chen, L.T. (2004) Detecting credit card fraud by using questionnaire-responded transaction model based on support vector machines. Proceedings of the Fifth International Conference on Intelligent Data Engineering and Automated Learning, vol. 3177, pp. 800–806.

[5] Dymarski, P. (2011). Hidden Markov Models, Theory and Applications. Rijeka, Croatia: InTech, 8-12.

[6] Lopez-Rojas, E. A. and Axelsson, S. (2014). Banksim: A bank paymentssimulator for fraud detection research. The 26th European Modeling and Simulation Symposium, Bourdeaux, France, 144–152.

[7] Ephraim, Y. and Merhav, N. (2002). Hidden Markov processes. IEEE Trans. Inform. Theory, 48, 1518–1569.

[8] Ghahramani Z. (2001). An introduction to hidden Markov models and Bayesian networks. International Journal of Pattern Recognition and Artificial Intelligence, 15(1): 9–42.

[9] Ghosh, S and Reilly, D.L. (1994). Credit Card Fraud Detection with a Neural-Network, Proc. 27th Hawaii Int'l Conf. System Sciences: Information Systems, vol. 3, pp. 621-630.

[10] Luis Jose S. Santos, and Shirlee R. Ocampo (2018). Bayesian Method with Clustering Algorithm for Credit Card Transaction Fraud Detection. Romanian Statistical Review.

[11] Pan, J., Rao, V., Agarwal, P., and Gelfand, A. (2016). Markov-modulated marked poisson processes for check-in data. In International Conference on Machine Learning, 2244–2253.

[12] Panigrahi, S., Kundu, A., Sural, S., Majumdar, A. K. (2009). Credit card fraud detection: A fusion approach using Dempster–Shafer theory and Bayesian learning. Information Fusion, 10(4), 354–363.

[13] Prakash A, Chandrasekar C. (2012). An ensemble approach for credit card fraud detection. International Journal of Computer Applications; 59(19): 1–6.

[14] Rabiner, L.R. (1989). A tutorial on hidden markov models and selected applications in speech recognition. Proceedings of the IEEE, 77(2), 257-286.

[15] Srivastava, A., Kundu, A., Sural, S., & Majumdar, A.K. (2008). Credit card fraud detection using hidden Markov model. Dependable and Secure Computing, IEEE Transactions on Dependable and Secure Computing, 5 (1), 37–48 .

[16] Robinson, W.N., Aria, A. (2018). Sequential fraud detection for prepaid cards using hidden Markov model divergence, Elsevier, Expert Systems with Applications, Volume 91, 235–251.

[17] The Nilson Report (2016). Top 10 Issuers of Payment Cards Worldwide 2016, The nilson report, no. 1121, 2017.

[18] The Nilson Report (2016). Card Fraud Losses Reach $22.80 Billion, The nilson report, no. 1118, 2017.