(一)摘要

近年來，隨著 Facebook、Instagram 以及 Twitter 等社群網站與 Yelp、以及 IMDB 等評論網站的興起，人們在上面打卡、按讚、評論甚至發表貼文的這些互動紀錄資料可以協助建構推薦系統，在幫助使用者找到他們可能感興趣資訊上發揮著重要作用。已有研究指出，如果惡意攻擊者能有權限接觸到推薦系統中的推薦結果與系統中的部分使用者資訊，他們也就能夠獲取使用者與項目的互動歷史紀錄（User-Item Interaction），因此隱私問題也就伴隨出現了。

隱私問題有重新識別攻擊（Re-Identification Attack）與個資推斷攻擊（Private-Attribute Inference Attack）兩大種類。前者已有不少研究解決此類問題，而後者通過被外洩的使用者互動歷史來獲取使用者的私人個資，如年齡、性別、職業和位置這些使用者不希望被公開的資料，針對這種類型的攻擊目前仍未存在有效措施來保護使用者。綜合上述所提，本專題希望以建立一個強健的生成對抗網絡系統架構並訓練出表現良好的攻擊者與推薦者，**設計出一套能有效保護使用者個資的推薦系統**，在不採取過往直接更動推薦系統中的使用者資料的強況下，不僅能有效推薦給使用者有用的資訊，更重要的是能夠防範攻擊者個資推斷攻擊。期望透過這樣的推薦系統能讓使用者對所使用的推薦系統平台感到信任，不再擔憂個資外洩的問題 。

本專題的特色為，相較以往多數的研究旨在防禦重新識別攻擊，此次研究的要點將朝向較少有研究成果的個資推斷攻擊發展出具有保護機制的推薦系統。此外，本專題也能一次考量多個隱私個資屬性，不僅在隱私保護上將有較佳的結果對於實際佈署時也較有效率。考量到相關的研究發展迅速，此次研究的一大目標則是結合十分新穎的「神經圖形協同過濾」與「用於順序推薦之分層門控網絡」兩大新技術，來分別改善原先舊有研究架構中的推薦者與攻擊者系統，期望能在推薦效果、隱私保護成效甚至是運算效能都能有優於以往的發展。

本次專題首先將會注重在使用 X. Wang [11]所提出的概念進行個性化排名推薦系統的建立；第二階段則是會參考 C. Ma [12]研究中提及的技術來建立攻擊系統，此階段的主要目的是不斷透過模型的自適與調整，試圖推斷出使用者的隱私個資。最後階段則是評估模型，此階段將會將所建立的模型與多種模型比較，並且分別評估其在隱私保護與推薦精準度的表現，並且觀察模型間的差異。

在計畫的後期，希望能有機會結合企業提供的即時社群資訊進這套系統，以便改善本研究採用的並不是即時資料的缺點，以便能提供更好的推薦與保護結果。

(二)研究動機與研究問題

研究動機

近年來，隨著 Facebook、Instagram 以及 Twitter 等社群網站與 Yelp 以及 IMDB 等評論網站的興起，人們不僅能利用這些社群網站認識新的朋友，同時也能在上面打卡、按讚、評論甚至發表貼文。我們每天每日都能在這些平台上看到來自不同朋友、不同來源的資訊，透過這些互動紀錄資料更可以協助建構推薦系統，在幫助使用者找到他們可能感興趣資訊上發揮著重要作用。

建立推薦系統時通常需要擁有使用者的個人屬性資料如年齡與性別以及使用者過往的偏好歷史紀錄如對電影或商家的評分資料，推薦系統再基於上述的資料進行演算向使用者推薦相關項目。儘管推薦系統很有效，但同時其也可能成為使用者隱私被侵犯的來源。已有研究指出，如果惡意攻擊者能有權限接觸到推薦系統中的推薦結果與系統中的部分使用者資訊，他們也就能夠獲取使用者與項目的互動歷史紀錄（User-Item Interaction），會有這樣的隱憂存在的一個主要原因是，推薦系統的推薦結果部分來自其他使用者的選擇結果（即使用者與項目的互動歷史），因此隱私問題也就伴隨出現了。

隱私問題的其中一個種類是重新識別攻擊（Re-Identification Attack），攻擊者透過尋找使用者是否存在於現有資料庫中去推斷使用者對項目的評分。目前已有研究解決此類問題，例如修改推薦系統的推薦結果讓單一評分結果或是整個使用者資料缺失，使得攻擊者不能推斷使用者的實際評分和偏好；另一種類型則是個資推斷攻擊（Private-Attribute Inference Attack），通過被外洩的使用者互動歷史來獲取使用者的私人個資，如年齡、性別、職業和位置這些使用者不希望被公開的資料，針對這種類型的攻擊目前仍未存在有效措施來保護使用者，主要手段為對使用者資料進行更動去混淆攻擊者，但這樣的方法往往會降低推薦系統的品質。此外，僅更動推薦系統中已存在的使用者資料不一定能防止攻擊者在未來獲取使用者新推薦結果時（例如購買新產品）推斷出使用者的個資。

綜合上述所提，本專題希望**設計出一套能有效保護使用者個資的推薦系統**，在不採取過往直接更動推薦系統中的使用者資料的強況下，不僅能有效推薦給使用者有用的資訊，更重要的是能夠防範攻擊者個資推斷攻擊。期望透過這樣的推薦系統能讓使用者對所使用的推薦系統平台感到信任，不再擔憂個資外洩的問題 。

研究問題

縱貫整個問題的架構，為了有效地建立一個能兼具保障良好推薦效果與保障使用者隱私的推薦系統，本專題將著重以下兩個目標：

（1）開發一個個性化的隱私保障推薦系統來防範隱私個資推斷攻擊

（2）確保使用者在收到個性化推薦後的隱私內容得到有效的保護不被輕易獲取

在建立對抗個資推斷攻擊的推薦系統時可視為一個對抗式學習的議題[1]，參考過往的研究，可將本專題的具體系統架構以建立一個強健的生成對抗網絡（Generative Adversarial Network，GAN）為主要方向，架構示意圖如圖 1，並區分為兩個發展主軸：

（1）個資推斷攻擊系統，目標是根據現有的推薦者反覆運算地調整其模型，使其不斷嘗試準確推斷使用者的個資

（2）推薦系統，利用使用者與項目的潛在特徵表示（Latent Representation）進行個性化推薦，同時利用個資推斷攻擊系統來調整與規範整個推薦的過程，以便躲過攻擊系統的攻擊

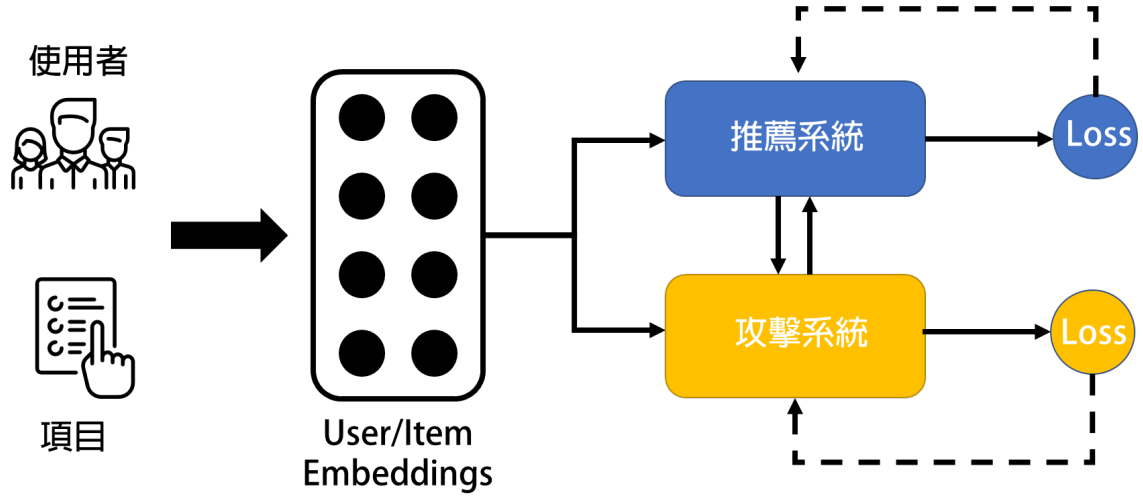


圖 1：GAN 架構圖

為了方便說明接下來的研究問題，將在下面定義此次研究問題需要的名詞與符號。

定義一：$I = \{i_1, \dots, i_M\}$ 代表 Item 項目的集合，如產品或是電影

定義二：$\mathcal{U} = \{u_1, \dots, u_N\}$ 代表使用者的集合

定義三：I_h 代表使用者 h 所評分的一組項目集合，$\mathcal{R}_h$ 則是推薦給使用者 h 的一組推薦項目結果

定義四：$\mathcal{P} = \{p_1, \dots, p_T\}$ 代表一組共有 T 個個資屬性的集合（例如，年齡、性別）

定義五：**R** 代表使用者-項目的評分矩陣（User-Item Rating Matrix）

本專題的目標是向人們推薦他們所感興趣的資訊。但是，同時也希望保護使用者們的隱私，防止惡意的攻擊者試圖根據使用者的項目資訊清單去推斷他們的隱私個資。每個使用者 h 的項目清單 S_h 是過去曾經評分過的項目和新推薦給使用者的項目之聯集，即 $\mathcal{S}_h = \{I_h \cup \mathcal{R}_h\}$ 。在此專題所定義的框架下，值得一提的是攻擊系統將會針對目標使用者之互動紀錄嘗試去推斷使用者的個資。

綜合上述所提的，可將問題具體定義為使整個系統能有辦法自我學習一個 f 函數，讓推薦系統可以向使用者 u_h 推薦會讓其有興趣的資訊 $\mathcal{R}_h$，並在滿足這樣的條件同時達到以下兩個要求：

（1）攻擊系統不能從使用者的項目清單 $\mathcal{S}_h = \{I_h \cup \mathcal{R}_h\}$ 中推斷出目標使用者的個資屬性 $\mathcal{P}$

（2）為使用者所推薦項目結果 $\mathcal{R}_h$ 必須是使用者所感興趣的問題

因而可以被定為找出以下目標函數： $\mathcal{R}_h = f(I_h, R, \mathcal{P})$。

(三)文獻回顧與探討

推薦系統中現有的隱私保護研究中多側重於保護使用者免受重新識別攻擊，在這種攻擊中，攻擊者將試圖推斷目標使用者的實際評分，並尋找目標是否存在於資料庫中。這些研究可被區分為基於差分隱私（Differential Privacy Based）以及基於擾動（Perturbation Based）兩種保護方式。

差分隱私為密碼學中的一種手段，原先用於保護統計資料庫中機密個人資料的隱私性，此種方法主要目標為能在資料查詢時最大化準確性並同時最大限度減少辨識其記錄的可能。由於推薦系統中與統計資料庫都擁有部分相似的特性（例如同樣存放著許多個人資料），因此也有不少研究將差分隱私的方法運用在推薦系統的隱私保護上。J. Hua 的研究中[2]，以差分隱私的概念提出一種矩陣分解方式，該方式將雜訊加到項目的向量中，並且採用第三方機制以減少每次迭代中添加的雜訊，此項研究可防止不受信任的惡意使用者獲取任何使用者的評分或個人資料；同樣基於差分隱私的概念，Z. Jorgensen 等人[3]考慮到先前差分隱私的保護機制會導致推薦系統品質上的劣化，提出將社群網路的群體結構進行分組，對使用者依社會關係進行分群，並進而在不同群間生成使用者偏好平均值，這樣的方法顯著地減少為了滿足差分隱私所需產生的雜訊；R. Bassily 等人[4]則為了滿足差分隱私的需要而修改使用者評分數而後才將其結果與推薦者分享。

以目前基於擾動的研究，多數是採用對使用者的過往互動資料中添加假項和評分來混淆攻擊者。H. Polat 等人[5]為了解決隨著運用協同過濾(Collaborative filtering)於推薦系統上的技術正變得越來越流行同時產生的隱私問

題，提出了一種隨機擾動（RP）技術，讓給定的使用者評分項目共享偽裝的 z-score，藉此保護使用者的隱私也同時仍能提供準確推薦的建議；同樣採用基於擾動的保護方式，D. Rebollo [6]的研究提出了一種基於資訊理論的隱私度量方法，從而將隱私風險降到最低；J. Parra [7]研究中，不僅透過添加或刪除使用者紀錄中項目和評分的手段來降低隱私外洩風險，同時考量到採取這樣措施後對推薦系統造成不可避免的效能降低，因此在為使用者和總體項目分佈之間的 Kullback-Leibler 差異（KL 散度在信息系統中稱為相對熵（relative entropy））進行量化措施後，特別針對隱私、偽造率和抑制率之間的權衡進行研究，找出折衷的結果。

個資推斷攻擊的重點是從使用者的公開資訊中推斷出其隱私個資，除了可利用目標使用者的好友資訊和社群成員資訊（如共同打卡紀錄）個資推斷。目標的個資，也可以從使用者個人的行為資訊（例如對電影的評分）進行攻擊。在現階段的研究中，較少有研究特別關注於保護使用者免受個資推斷攻擊。J Jia [8]等人提出了一種防禦系統 AttriGuard，計算上易於處理，在推薦系統上的效能損失小，在給定隱私個資相對於使用者實際個資的某個機率分佈進行抽樣後，找到了影響最小的雜訊並將其加到使用者-項目互動資料中，如此攻擊者就會將抽樣到的個資預測為特定使用者的隱私個資；U. Weinsberg [9] 的研究則提出，將一定數量的虛擬項添加到每個使用者的紀錄資料中才發佈那些使用者-項目評分數據，而添加虛擬項與其實際內容呈負相關，然而此研究只能逐項針對單一個資屬性佈署無法一次考量到多個屬性；G. Beigi [10] 所發展的基於對抗式學習之模型 RAP ，則結合了 BPR 推薦系統與以 RNN 為基礎的攻擊者系統達成兼具推薦效能跟保護使用者免受個資推斷攻擊的成效。

作者＼元素	針對個資推斷攻擊防護	可針對多項個資屬性佈署	採用效能更佳並考慮序列關係的推薦系統與攻擊系統
R. Bassily [4]		✓	
U. Weinsberg [9]	✓		
G. Beigi [10]	✓	✓	
本計畫	✓	✓	✓

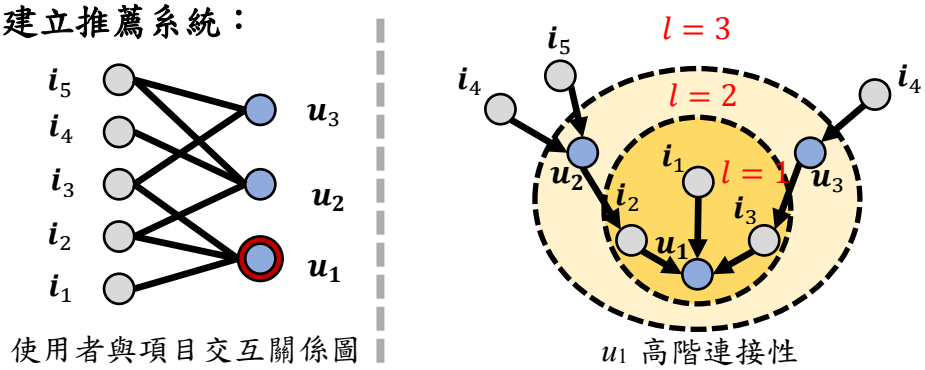
表 1：各項研究與本次專題之比較

雖然 G. Beigi 所提出的 RAP 模型在隱私防護與推薦成效上相較過往研究都有不錯的表現，然而因為其所採用的推薦與攻擊系統模型皆較為傳統與基本，並且沒考慮到項目推薦的序列性問題，因此本專題希望能結合目前十分新穎的技術如 X. Wang [11]所提出「神經圖形協同過濾」(Neural Graph Collaborative Filtering) 這個全新推薦系統框架以及 C. Ma [12]所提出的「用於順序推薦之分層門控網絡」(Hierarchical Gating Networks for Sequential Recommendation，HGN) 來分別改善原先 RAP 架構中的推薦者與攻擊者系統，期望能在推薦效果、隱私保護成效甚至是運算效能都能有優於以往的發展，本專題與其他研究的詳細比較請參閱表 1。

(四)研究方法及步驟

本次專題將會使用 Python 作為分析資料與實作演算法的主要工具，並且主要分成三個階段。第一階段將會注重在使用 X. Wang 所提出的概念進行個性化排名推薦系統的建立，目的是從使用者資料中提取他們的實際偏好並向他們推薦有關的項目；第二階段則是會參考 C. Ma 研究中提及的技術來建立攻擊系統，此階段的主要目的是不斷透過模型的自適與調整，試圖推斷出使用者的隱私個資。此外，推薦系統也會利用攻擊系統確保先前評分的項目和新推薦的項目聯集不會洩漏使用者的隱私個資，並進一步欺騙攻擊系統來調整推薦系統本身；接著最後階段則是評估模型，此階段將會將所建立的模型與多種模型比較，並且觀察模型間的差異。

階段一 建立推薦系統：



使用者與項目交互關係圖　　u_1 高階連接性

圖 2 高階連接性與使用者與項目交互圖

多數現有構建 Embedding 方法僅使用描述性特徵（例如，ID 和其他屬性）而沒有考慮使用者與項目互動過程，因此使用者與項目互動中潛在的協同訊號（Collaborative Signal）並沒有在 Embedding 過程中被進行編碼，如此所得的 Embedding 可能不足以達成有效的協同過濾效果。因此 X. Wang 研究中就提出將使用者與項目互動關係整合到 Embedding 中，以神經圖形協同過濾的架構，採用使用者與項目互動的「高階連接性」(High-Order Connectivity) 來建

模，用更為顯著方式將協同訊號有效地結合到 Embedding 過程中。於圖 2 說明了高階連接性的概念，以 u_1 為要向其推薦的目標對象，其高階連接性代表的意義為任何節點到達 u_1 的路徑長度 l 大於 1 者。例如，路徑 $u_1 \leftarrow i_2 \leftarrow u_2$ 表示 u_1 和 u_2 之間的行為具有相似性，因為兩個使用者都和 i_2 共同進行了交互作用。而另外較長的路徑 $u_1 \leftarrow i_2 \leftarrow u_2 \leftarrow i_4$ 則暗示 u_1 可能對 i_4 有興趣，因為與 u_1 相似的使用者 u_2 已經與 i_4 擁有交互作用。此外，從 $l=3$ 的整體角度來看，i_4 比 i_5 更受 u_1 的關注，因為有兩條路徑連接 i_4 與 u_1，而只有一條路徑連接 i_5 與 u_1。

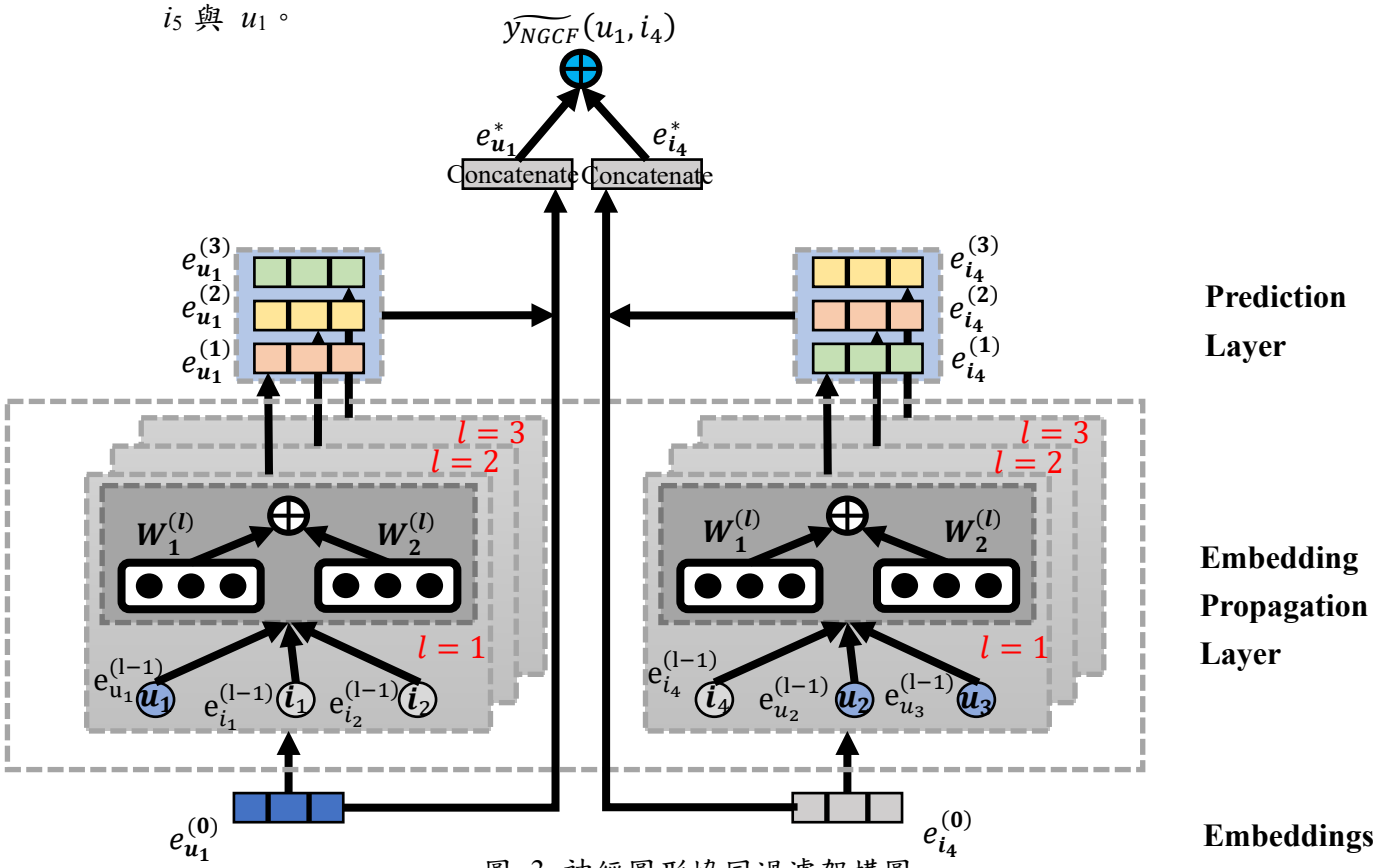


圖 3 神經圖形協同過濾架構圖

在討論具體模型建立時，首先先將使用者與項目的集合進行 Embedding 分別得到 $\mathbf{e_i}$ 跟 $\mathbf{e_u}$ ，對於每對連接起來的使用者-項目對（Pair）(u, i)定義從 i 到 u 的訊息為 $\mathbf{m_{u \leftarrow i}} = f(\mathbf{e_i}, \mathbf{e_u}, p_{ui})$，$\mathbf{m_{u \leftarrow i}}$ 為訊息 Embedding，$f(\cdot)$是以 $\mathbf{e_i}$ 跟 $\mathbf{e_u}$ 為輸入的訊息編碼函數並使用係數 p_{ui} 控制每次傳播的衰減因子。在本架構中，可將 $f(\cdot)$ 定為：$\mathbf{m_{u \leftarrow i}} = \frac{1}{\sqrt{|\mathcal{N}_u||\mathcal{N}_i|}}\left(\mathbf{W_1 e_i} + \mathbf{W_2}(\mathbf{e_i} \odot \mathbf{e_u})\right)$，$\mathbf{W_1}$ 與 $\mathbf{W_2}$ 為可訓練的權重矩陣而⊙代表元素的乘積。由圖形卷積網路的研究中，可將 p_{ui} 設為 Graph Laplacian Norm $1/\sqrt{|\mathcal{N}_u||\mathcal{N}_i|}$，$\mathcal{N}_u$ 與 $\mathcal{N}_i$ 分別是使用者 u 與項目 i 的第一層鄰居。通過更多層的聚合與傳播，在第 l 階段的使用者表示式可以遞迴地 1 成： $\mathrm{e}_u^{(l)} = \mathrm{LeakyReLU}\left(\mathrm{m}_{u \leftarrow u}^{(l)} + \sum_{i \in \mathcal{N}_u} \mathrm{m}_{u \leftarrow i}^{(l)}\right)$，被傳遞的訊息可表

示成以下：$m_{u\leftarrow i}^{(l)} = p_{ui}\left(W_1^{(l)} e_i^{(l-1)} + W_2^{(l)}\left(e_i^{(l-1)} \odot e_u^{(l-1)}\right)\right)$、$m_{u\leftarrow u}^{(l)} = W_1^{(l)} e_u^{(l-1)}$。在傳播了 L 層之後，使用者 u 與項目 i 有了多種表示式 $\left\{\mathbf{e_u^{(1)}}, \cdots, \mathbf{e_u^{(L)}}\right\}$、$\left\{\mathbf{e_i^{(1)}}, \cdots, \mathbf{e_i^{(L)}}\right\}$，在不同層中所輸出的表示式代表經過不同傳遞管道的訊息，因此它們在反映使用者偏好方面具有不同的作用，而我們將它們串聯起來構成使用者的最終 Embedding。以||代表串聯的動作，最終的 Embedding 可表示為 $\mathbf{e_u^*} = \mathbf{e_u^{(0)}}||\cdots||\mathbf{e_u^{(L)}}$，$\mathbf{e_i^*} = \mathbf{e_i^{(0)}}||\cdots||\mathbf{e_i^{(L)}}$。最後一步，以內積求得使用者對目標項目的偏好：$\widehat{y_{NGCF}}(u,i) = e_u^{*\top} e_i^*$。

階段二 建立攻擊系統：

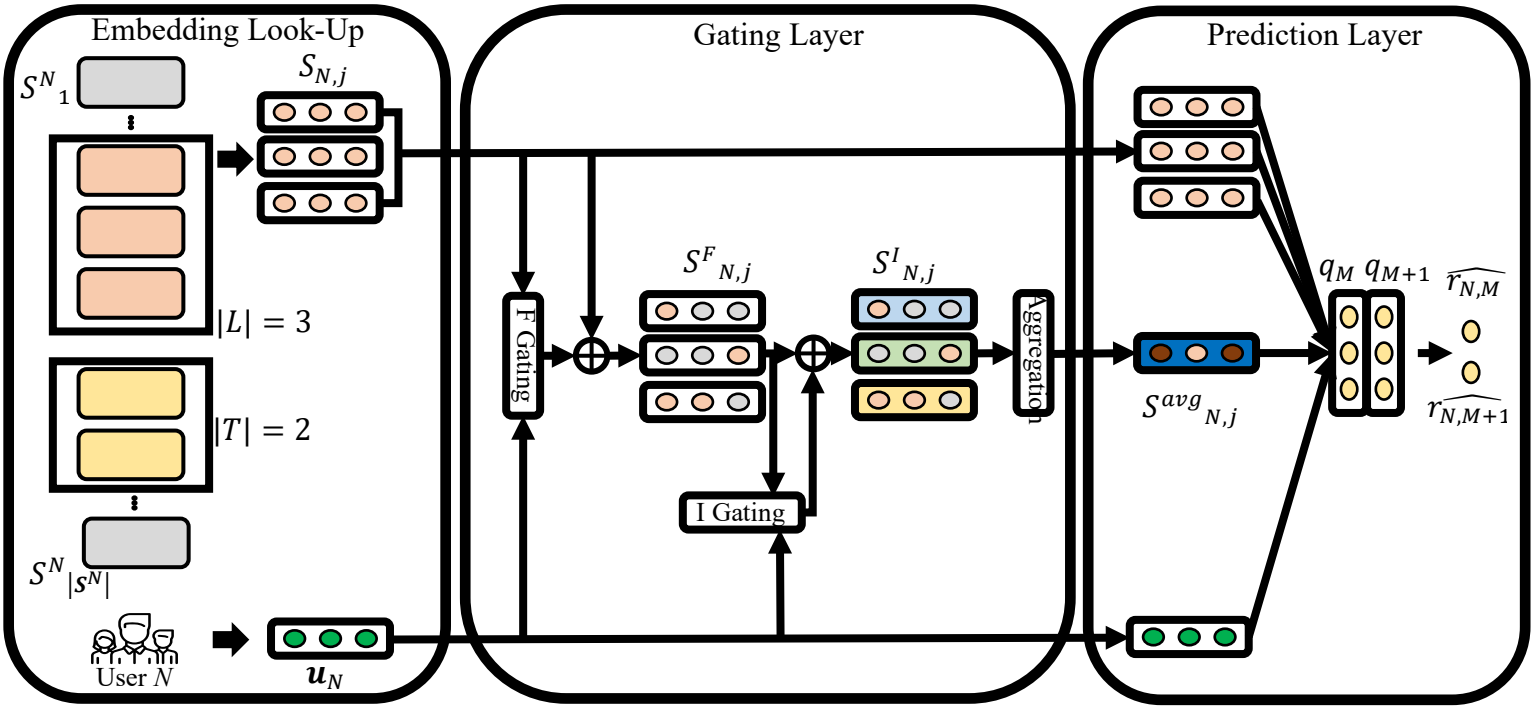


圖 4 HGN 架構圖

過往有不少研究採取 CNN 或 RNN 的架構去學習使用者行動上的順序，那樣的架構卻未考慮到不同項目特徵的特定部分。忽略代表性特徵可能無法真實地捕捉到使用者的短期興趣。其次，這些基於 CNN 或 RNN 的方法也不會根據使用者的偏好來區分項目的重要性，將這些有意義的項目與其他項目同等對待可能會造成對使用者意圖的理解不完整。此外，過往的研究也相對不注重項目之間的關係，而實際上使用者很可能會與密切相關的項目一個一個的產生互動。有鑑於此，C. Ma 所提出的研究特別著墨在對於短期興趣建立新的模型架構，採取這樣的技術來將序列性問題考慮進攻擊系統的模型，可使攻擊系統更加完善，而因本專題所採用 GAN 的架構，使攻擊系統強化的同時也對推薦

系統上是有所助益的。

參考圖 4 HGN 的架構圖，S^N 為使用者 N 的項目序列，可將問題理解於給定 $|L|$ 個連續項目，在使用者-項目互動序列 S^N 中，使用者隨後將會與哪項目進行互動。而由於此為攻擊者系統，所輸出的 q_M 為使用者個資的 Embedding 欄項目。

階段三 模型評估：

在模型評估上，將採用公開數據 MovieLens [13]進行實驗，該數據集包括 943 位使用者對 1682 部電影的 100000 個評分。每個使用者至少對 20 部電影評分，評分分數在 1-5 之間。在收集的數據集中，每個使用者都與三個隱私個資屬性，即性別（男性/女性），年齡和職業。

在評估指標上，主要採取針對隱私保護的效果以及推薦系統是否有成功地推薦有品質的內容給使用者。

（1）隱私保護：由於數據分配不平衡，因此我們採用了 Micro-AUC 以提供更準確的評估。較低的 AUC 在掩蓋隱私個資屬性方面會有較好表現。

（2）推薦品質：採用 $P@K$ 與 $R@K$ [14]進行評估。

$$P@K = \frac{|\{ test\ items \} \cap \{top - K\ returned\ items\}|}{K}$$

$$R@K = \frac{|\{ test\ items \} \cap \{top - K\ returned\ items\}|}{|\{ test\ items \}|}$$

(五)預期結果

如研究問題中提及，本專題將會建構一套推薦系統架構，預期能達到兼具保護使用者隱私個資的同時能有良好的推薦品質。此外，本專題也將結合目前推薦系統的最新技術去改善先前其它以個資推斷攻擊保護為主的研究，希望透過建構與訓練出更強大的推薦系統與攻擊系統的同時，能在隱私保障、推薦精準度甚至是運算效能上都能有顯著的提升

此外，由於我們的資料是來自過去文獻、網站上彙整而來，並不是即時的資訊，因此如果未來有企業願意提供即時的社群資訊並結合這套系統，便能提供更好的推薦與保護結果。

(六)參考文獻

[1] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. 2014. Generative adversarial nets. In Advances in neural information processing systems. 2672–

2680.
[2] Jingyu Hua, Chang Xia, and Sheng Zhong. 2015. Differentially Private Matrix Factorization.. In IJCAI. 1763–1770.
[3] Zach Jorgensen and Ting Yu. 2014. A Privacy-Preserving Framework for Personalized, Social Recommendations. EDBT 582.
[4] Raef Bassily and Adam Smith. 2015. Local, private, efficient protocols for succinct histograms. In Proceedings of the forty-seventh annual ACM symposium on Theory of computing. ACM, 127–135.
[5] Huseyin Polat and Wenliang Du. 2003. Privacy-preserving collaborative filtering using randomized perturbation techniques. In International Conference on Data Mining. IEEE.
[6] David Rebollo-Monedero, Javier Parra-Arnau, and Jordi Forné. 2011. An information-theoretic privacy criterion for query forgery in information retrieval. In International Conference on Security Technology. Springer, 146–154.
[7] Javier Parra-Arnau, David Rebollo-Monedero, and Jordi Forné. 2014. Optimal forgery and suppression of ratings for privacy enhancement in recommendation systems. Entropy 16, 3 (2014), 1586–1631.
[8] J Jia and Gong NZhenqiang. 2018. AttriGuard: A Practical Defense Against Attribute Inference Attacks via Adversarial Machine Learning. In 27th {USENIX} Security Symposium ({USENIX} Security 18). USENIX Association.
[9] Udi Weinsberg, Smriti Bhagat, Stratis Ioannidis, and Nina Taft. 2012. BlurMe: Inferring and obfuscating user gender based on ratings. In Proceedings of the sixth ACM conference on Recommender systems. ACM, 195–202.
[10] Beigi, Ghazaleh & Mosallanezhad, Ahmadreza & Guo, Ruocheng & Alvari, Hamidreza & Nou, Alexander & Liu, Huan. (2019). Privacy-Aware Recommendation with Private-Attribute Protection using Adversarial Learning.
[11] Wang, Xiang, et al. "Neural graph collaborative filtering." Proceedings of the 42nd international ACM SIGIR conference on Research and development in Information Retrieval. 2019.
[12] Ma, Chen, Peng Kang, and Xue Liu. "Hierarchical gating networks for sequential recommendation." Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. 2019.
[13] F Maxwell Harper and Joseph A Konstan. 2016. The movielens datasets: History and context. Acm transactions on interactive intelligent systems (tiis) 5, 4 (2016).
[14] Cai-Nicolas Ziegler, Sean M McNee, Joseph A Konstan, and Georg Lausen. 2005. Improving recommendation lists through topic diversification. In Proceedings of the 14th international conference on World Wide Web. ACM, 22–32.

(七) 需要指導教授指導內容

資料探勘及社群網路分析：需要教授指導資料探勘和社群網路分析的概念及實際操作。

演算法以及資料結構：需要教授指導 Python 的使用，並協助指導演算法與資料儲存的設計，讓實際操作時即使遇到龐大的數據及運算也能夠順利完成分析。

預測模型的概念及選用：除了基本的機率預測模型外，尚需要請教授介紹、比較其他預測模型選用的優缺點， 並將其寫成程式語言供電腦運作。