**三、研究計畫內容（以中文或英文撰寫）：**

1. **研究計畫之背景。請詳述本研究計畫所要探討或解決的問題、研究原創性、重要性、預期影響性及國內外有關本計畫之研究情況、重要參考文獻之評述等。如為連續性計畫應說明上年度研究進度。**

* **研究目的與重要性**

隨著網際網路快速的發展，越來越多的網路服務唾手可得，這也讓現代人越來越依賴網路服務平台。如週五下班前人們可能會先透過外送服務平台訂購一些派對食物，接著由影音串流平台挑選一部合適的電影來度過美好的周末。雖然網路服務平台提供了令人目不暇的商品與服務，但過多的資訊也讓使用者難以應付。舉例來說，在北美區域，Netflix提供線上使用者超過4000部電影與1000部電視劇[[1]](#footnote-1)，對使用者而言，要從中找出一部喜歡的影集並不是一件容易的事情。許多研究發現這種『資訊過載』(information overload)現象會削弱使用者對網路服務平台的使用者體驗，進而影響他們再訪平台的意願**[19]**，而這種現象也會損害平台獲益，因為平台獲益通常與使用者人數呈正比。為了留住使用者，進而增進網路服務平台的收益，推薦適切的個人化商品就變得十分重要**[4, 10]**。實務上，推薦系統(recommendation systems)已是許多知名網路服務平台的核心功能，如電子商務巨擘Amazon與影音串流龍頭Netflix都仰賴推薦系統來提升使用者滿意度，進而提高平台獲利。以Netflix為例，首席產品官Hunt表示，80%以上的Netflix影音觀看都是透過推薦系統發生的，若能持續提昇Netflix推薦系統效能，就可增加用戶在平台上的黏著度，那Netflix每年可減少因客戶流失造成的10億美元損失**[8]**。由上可知，推薦系統除了能讓使用者快速且正確的獲取他們想要的資訊，對平台業者也能帶來可觀的效益，為促成使用者與平台商雙贏局面，開發有效的推薦系統便是一項重要的研究議題。

一般而言，推薦系統會分析使用者於服務平台上的行為(如點選某影音或對某商品評分)來推論使用者興趣(user preferences)，並藉此濾除無關使用者興趣的物品。經典的「協同式過濾法」(collaborative filtering)**[17]**便是透過上述概念來進行商品推薦，以「以使用者為基礎的協同式過濾法」(user-based collaborative filtering)為例，該方法會收集目標使用者(target user)喜歡(或採買過)的物品，藉此找出一群喜好物品與目標使用者高度重疊的參考使用者(reference users)，由於這些人的喜好與目標使用者相似，該方法會推薦這些參考使用者喜歡的物品給目標使用者。如今在Amazon網站上選購了一件商品後，網站會自動推薦「購買此商品的顧客，同時也購買了什麼商品」，這便是一個協同式過濾的經典案例。

如上所述，傳統的推薦演算法需仰賴(分析)使用者行為來進行商品推薦，也因此，它們在面對新進使用者(new user)時較難推薦符合使用者興趣的商品，這是因為新進使用者只有少許的行為資料，推薦演算法難以捕捉使用者興趣來進行合適的商品推薦，而這樣的困境被稱作「新進使用者冷啟動問題[[2]](#footnote-2)」(new user cold start)。一些文獻指出，使用者體驗是決定新進使用者是否會成為忠實客戶的重要關鍵**[6]**，以Netflix為例，新客戶往往在瀏覽少於20部推薦影片的情況之下決定是否加入平台。為留住新客戶，進而為網路服務平台業者帶入更多的利潤，設計好的推薦演算法來減緩新進使用者冷啟動現象是非常重要的。

過去已有不少推薦演算法被設計來解決新進使用者冷啟動問題，其中一類的方法為「基於內容推薦法」(content-based method)**[24]**。有別於收集使用者行為資料，這類方法收集使用者個人資料，如使用者簡歷、年齡或性別等，並透過這些資料來將使用者表示成數據向量(vector representation)，藉由向量相似度計算(如cosine similarity)來找出與新進使用者資料相似的參考使用者，且歸納這些參考使用者所喜歡的商品來完成推薦。基於內容推薦法的運作前提是演算法要能獲取使用者個人資料，實務上，使用者不一定有意願提供這類資訊，網路隱私與安全性考量也影響了這類方法的可行性。另外一種解決新進使用者冷啟動問題的方法為「基於代表商品推薦法」(representative-based method)**[7, 13, 20]**，該方法的核心是要挖掘出推薦系統內“具代表性的商品(representative items)”且推論出這些代表性商品與一般商品的關聯性。不同於傳統方法須仰賴足夠的使用者行為來進行商品推薦，若新進使用者對部分代表性商品有互動行為(如評分或購買)，這類方法便可透過上述商品關聯性間接地推論出新進使用者對其他商品的喜好，並進行商品推薦。

近年來，隨著深度學習(deep learning)在文字、圖像、語音等領域的亮眼表現，越來越多的學界與業界專家也開始研究該如何將深度學習應用於推薦系統。在業界，Covington等人使用深度學習技術來建構YouTube影片推薦系統**[3]**。該方法先運用一組具有三層隱藏層(hidden layers)的Deep Neuron Network (DNN)模型分析使用者的影片瀏覽歷史、搜索歷史、人口統計學訊息等資訊，找出該使用者可能喜歡的候選影片集合，接著再將使用者與候選影片的關聯資訊，例如：使用者瀏覽該頻道的次數、使用者最近一次瀏覽該頻道距現在的時間等，以另一個三層隱藏層的DNN模型來學習使用者對這些候選影片的喜好排序，進而完成影片推薦。任職於Google的Cheng等人也使用深度學習來建構Google Play的APP推薦**[2]**。該方法將Google Play中使用者的數據型資料(例如：使用者年齡、使用者已安裝的APP數量等)與離散型資料(例如：使用者使用的裝置種類、使用者已經安裝的APP等)，作為DNN模型的輸入，藉此產生使用者的特徵向量(embedding)，接著考量使用者特徵向量與APP的相關資訊來決定是否推薦某APP給某個使用者。在學界，有鑑於越來越來的推薦系統論文使用深度學習技術，推薦系統領域知名的ACM RecSys研討會便於2016年起固定徵求以深度學習為基礎的推薦系統論文，藉此促進深度學習技術在推薦系統上的發展與交流。見到如此多業界與學界專家趨之若鶩的投入這項新興研究議題，就能得知開發以深度學習為基礎的推薦系統是一項無法避免的趨勢。因此，本計畫希望能立足在這些研究之上，建構一套有效的深度學習模型來解決推薦系統中新進使用者冷啟動問題。透過本研究，我們將有機會推薦符合新進使用者需求的商品，提高他們於網路服務平台上的黏著度，創造出使用者與平台業者雙贏局面。此外，藉由推薦合適的商品，新進使用者可逐漸脫離冷啟動狀態，進而獲得更精確的商品推薦，如此形成使用者與平台之間正向的互動循環。

在閱讀大量推薦系統與深度學習論文後，我們認為「去雜訊編碼解碼器」 Denoising Auto-Encoder (DAE)**[22]**這種深度學習架構非常適合模擬推薦系統冷啟動問題。DAE模型常用於處理高維的圖像和語音資料，其的目的是要訓練一組資料編碼與解碼器。以圖像資料為例，DAE的訓練分成兩階段：第一階段為**雜訊產生階段**(**corruption stage**)，該階段會把圖片中的像素以隨機或常態機率分布模式將某些像素值設定為0，這些被修改過的像素即視為原始圖片上的雜訊(noise)。第二階段為**編碼解碼階段**(**encode-decode stage**)，該階段會將這含有雜訊的圖片以像素向量的樣式輸入encoder-decoder深度學習網路。在encoder端，輸入的像素向量會經過一層以上的隱藏層來降低圖片向量維度；而decoder端則是將圖片的降維向量透過一層以上的隱藏層還原回原始未受雜訊影響的圖片。簡單來說，DAE模型的訓練目標是要產生一組好的encoder和decoder，好將受雜訊干擾的圖片還原成起初未經雜訊干擾的圖片。訓練好的DAE模型除了可用來去除圖像雜訊外，其中介產物 – 降維向量還可用來表徵原始圖片。

如上所述，DAE常應用於處理圖像與語音資料，且能有效移除資料雜訊和將高維資料壓縮降維。為要將DAE套用於推薦系統情境，進而解決新進使用者冷啟動問題，我們需通盤了解DAE模型以進行模型改良。針對DAE這兩大優點，我們初步規劃兩種以DAE為基礎的新進使用者冷啟動推薦方法，第一個方法是基於DAE之向量雜訊移除，第二種做法是基於DAE之向量維度壓縮，並在本計畫前兩年研究這兩大做法，而計畫第三年將嘗試整合這兩種方法，且反思如何將所設計的方法運用於新進產品冷啟動推薦。本多年期計畫目標如下:

* + 第一年: 設計以DAE向量雜訊移除為基礎之新進使用者冷啟動推薦方法

推薦系統常用高維向量來描述一位使用者對商品的互動行為，以電影推薦為例，演算法會將每一個向量維度(dimension)視為一部電影，而使用者對電影的評分即為向量的維度元素值。DAE之向量雜訊移除與新進使用者冷啟動推薦的關聯性在於，我們可將高維使用者向量視為DAE要還原的原始資料，而DAE訓練時的雜訊產生就如同從向量中抹去部分使用者對物品的互動資訊，藉此把使用者模擬(倒退)成剛進平台的新進使用者。而後續的encoder-decoder訓練就是讓模型把資料稀疏的新進使用者向量成長(快轉)為後來資料豐富使用者向量。待DAE模型訓練完畢後，我們就可以用該模型來為新進使用者推算未來喜歡的商品。也就是說，我們可將新進使用者的稀疏向量輸入訓練好的encoder-decoder網路，再檢視decoder產出的豐富向量來為新進使用者建立推薦清單。為更貼近新進使用者冷啟動問題，我們改稱DAE訓練時的雜訊產生階段為「**使用者返老還童階段**」(**user rejuvenation stage**)。

因DAE主要應用於圖像與語音資料，其訓練時雜訊產生的方式並不一定適用於新進使用者冷啟動情境。若我們在返老還童階段隨機抹去使用者對商品的互動資料，所產生的資料稀疏向量未必能確切反映新進使用者冷啟動狀態。因此，在本研究的第一年，除通盤了解DAE，我們將著重於雜訊產生方法。我們初步的想法是要從大量的使用者與物品互動資料中找出具代表性的商品(representative items)，在將使用者還原成新進使用者狀態時，我們並不隨機抹去商品互動資料，而是盡可能抹去非代表性商品互動，好讓返老還童後的稀疏使用者向量盡量保留這些具代表性商品的互動資訊。我們認為使用者會開始使用一項服務平台(既成為新進使用者)，應該是受到代表性商品的吸引，例如電商平台上的熱門手機特賣可能會吸引新客戶註冊搶購。若能使稀疏向量盡量保留代表性商品，則越有可能把使用者向量還原成新進使用者狀態，也有利於後續encoder-decoder訓練。我們將考量多種商品面向(如熱門程度、覆蓋率與時間性)，並從大量的使用者商品互動資料中挖掘出可能的代表性商品，並研究不同面向對返老還童的影響，與討論它們能否有效解決新進使用者冷啟動推薦。

* + 第二年: 設計以DAE向量維度壓縮為基礎之新進使用者冷啟動推薦方法

由於網路服務平台提供的商品繁多，使用者往往只會和少部分商品有互動行為，而新進使用者的商品互動行為更是寥寥可數。為減少這稀疏現象對推薦系統的影響，許多推薦演算法會將使用者(或商品)高維向量進行降維，讓使用者的商品互動行為以一低維向量表示，進而提高推薦效率。舉例來說，若我們從使用者的採買明細(用一高維向量描述)歸納出該使用者對3C與運動(低維商品類型)商品很有興趣，則我們可推薦這兩類商品給該使用者。知名的矩陣拆解推薦法(matrix factorization) **[12]**就是透過向量降維來提升推薦系統效能，其推薦結果優異，也讓它從眾多競爭者中脫穎而出，獲得2009 Netflix推薦競賽的一百萬美金競賽獎金。由此可見，向量降維對推薦系統有絕對的影響性。

Auto-encoder深度學習的強項之一就是能將高維資料進行降維，求出資料的低維代表向量，而DAE更是能有效壓縮受雜訊汙染的高維資料。在本計畫第二年，我們將挑戰以壓縮降維的方式來解決新進使用者冷啟動推薦。我們會使用DAE降維後的向量來表示使用者，並透過向量相似度計算來進行協同式過濾推薦。簡單來說，在DAE訓練階段，我們會將所有使用者(包含新進使用者與一般使用者)的高維向量丟進DAE模型進行網路參數訓練，待訓練完畢後，每個使用者於encoder階段的輸出會被保留下來，這些低維向量就是每個使用者的向量表示。不同於第一年的做法，我們並不使用decoder來進行商品推薦，相反的，我們會採用協同式過濾法來進行新進使用者冷啟動推薦。對一位新進使用者，我們運用向量相似度來挑選與該使用者低維向量極為相似的參考使用者，接著推薦這些參考使用者喜歡的商品來解決冷啟動問題。除了協同式過濾法，我們也會研究其它推薦方法，如我們也可以運用DAE模型來將商品表示成低維向量，並設計其它網路架構來分析使用者與商品在低維空間上的非線性關聯性以產生商品推薦。而分析商品低維向量也可協助我們思考新進商品冷啟動推薦的可行性。此外，其它研究細節，如DAE訓練時的雜訊產生方式，也將在這一年內探討。

* + 第三年: 整合DAE維度壓縮與雜訊移除並擴展研究新進商品冷啟動推薦

前項研究分別針對DAE兩大特點來進行新進使用者商品推薦，在本計畫最後一年，我們將累積前項研究經驗，整合前述兩大方法以期提升新進使用者冷啟動推薦效能。在第二年，我們將所有使用者(包含新進使用者與一般使用者)的高維向量丟進DAE模型進行網路參數訓練，藉此得到降維後的使用者向量，但我們也可把一般使用者與新進使者分開處理，先透過第一年所研究的返老還童程序來將一般使用者推回新進使用者狀態，再用這些稀疏向量與其原始向量進行DAE訓練，待模型訓練完畢後，我們再把新進使用者的稀疏向量輸入模型以取得這些新進使用者的降維向量，並透過decoder來產生一組商品推薦清單。此外，我們還會將新進使用者的降維向量與一般使用者的降維向量進行比較以產生協同式過濾的推薦清單，最後再將這推薦清單與decoder產生的推薦清單整合來推薦商品給新進使用者。我們整合這兩套方法的考量是這兩方法各自有不同的推薦思維，透過decoder產生的清單是仰賴新進使用者的稀疏向量，它應會著重使用者自身喜好。而協同式過濾因考量其它(一般)使用者，它較能產生多元的商品推薦，我們期待整合這兩大特色以推薦合適的商品給新進使用者。

上述方法雖然專注於新進使用者冷啟動問題，但透過不同的向量表示方式，這些方法將有機會用於新進商品冷啟動(new item cold start)推薦。在計畫尾聲，我們將微調所設計的方法來嘗試解決新進商品冷啟動問題。如我們可將商品以一高維向量表示，每一個維度代表一個使用者對該商品的互動行為(如評分或採買)，如此，我們也能為商品向量進行返老還童與降維壓縮，並借助DAE模型來推薦新商品給合適的使用者。

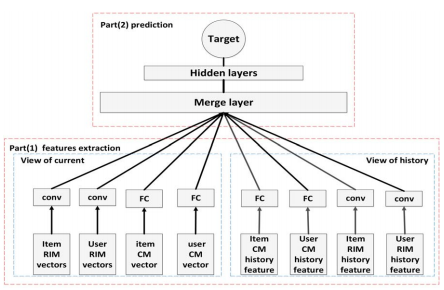
* + - **研究背景與國內外重要參考文獻**

推薦系統是一項歷史悠久的研究議題，也因此累積了為數可觀的研究文獻，但如前節所述，近期的推薦系統研究多著重於運用深度學習，這主要是因為以深度學習為基礎的推薦系統有以下兩大優點**[26]**: (i)可處理複雜的非線性關係。以往知名的推薦方法(如前節所提到的matrix factorization)多半以線性組合(如向量內積)的方式來模擬使用者與商品關係，而深度學習整合多種非線性啟動函式(activation function，如sigmoid與relu)，因此較能效捕捉現實生活中使用者與商品複雜的互動關係。(ii)可有效擷取資料特徵。許多推薦方法會將高維的使用者(商品)向量進行降維，藉此找出有用的使用者(商品)特徵。特徵工程(feature engineering)是深度學習的強項之一，無論是監督式或非監督式學習，深度學習都可透網路階層來自動擷取不同精細程度的資料特徵，這特性也有助於推薦系統整合多種使用者(商品)資料與特徵，使推薦結果更加精準。

在本章節，我們介紹近期重要的推薦系統，我們尤其著重於運用深度學習的推薦系統與冷啟動推薦。由於我們的研究重點之一為挑選重要物品來進行DAE返老還童，我們另有一節探討具代表性商品相關研究。

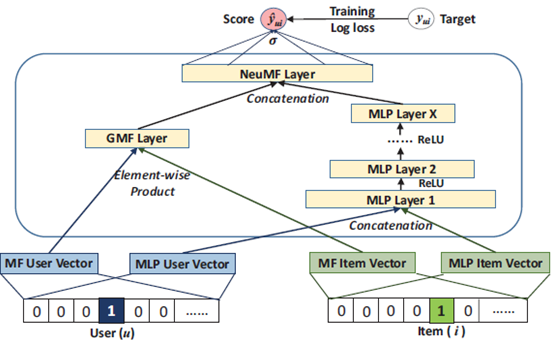
* + 以深度學習為基礎之推薦系統文獻探討

以深度學習為基礎的推薦方法多使用使用者行為來進行模型訓練，以期模型能捕捉(模擬)使用者對商品的行為，進而完成商品推薦。模型訓練主要包含兩大階段，以電影評分與推薦為例，第一階段須決定以何種向量表示法來陳述使用者對電影(商品)的評分，並將該向量作為深度學習模型的輸入。第二階段則是設計適合的類神經網路架構，使模型預測的電影評分值與真實評分誤差越小越好。在文獻**[5]**，作者提出了兩種方法來產生user和item的embedding vector，一個稱作Constraint Model (CM)，以item為例，其做法是利用找出對商品*A*和商品*B*皆評過分且評分相同的使用者數，藉此描述商品*A*與*B*的關聯性，並將所有商品關聯性以一關係矩陣*R*表示，而商品的embedding vector即為*R*矩陣的分解結果。另一個產生user和item embedding的做法為Rating Independent Model (RIM)，以item為例，其做法則是找出對商品*A*和商品*B*皆評過分但是評分不同的使用者數量來描述*A*、*B*的關聯，再運用如上的矩陣分解法來求出商品embedding。除了透過前述方法產生出item RIM vector、user RIM vector、item CM vector和user CM vector，該方法還會產生4個稱作「View of history」向量，該方法會將所有對某item評過分的user向量進行平均，用來表示該item的history向量，而user的history vector則是平均該user所有評分過的item vectors。如下圖所示，上述的8個向量會分別進行fully connected DNN和Convolution Neural Network (CNN)處理，再將所擷取的特徵向量合併，且輸入最終的網路層來預測某個使用者對某個商品的評分。訓練完模型後，該模型就可以預測使用者對任一商品的評分，高評分的商品就可推進給使用者。



圖一、文獻**[5]**的神經網路架構圖

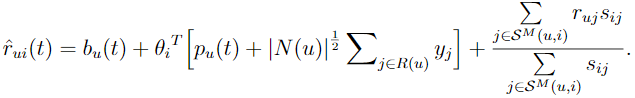
在文獻**[9]**，推薦系統要模擬的使用者行為不是商品評分，而是使用者對某商品是否有興趣。如下圖所示，該方法在第一階段會以one-hot encoding的方式來表示一個使用者與商品，接著利用4組單層的embedding layer來產生user與item的embeddings，這些embeddings再分別經過Generalized Matrix Factorization (GMF) layer和Multi-Layer Perceptron (MLP) layer，來學習user和item之間線性與非線性關係的embeddings。其中GMF的運算是將user和item向量中對應的維度元素值相乘，其目的希望以matrix factorization的角度來捕捉向量間線性的運作關係，在MLP， user和item的向量在合併後會經過一連串非線性的運算，以此捕捉它們非線性的關係。模型最後再同時處理GMF與MLP的embedding來預測使用者的使否對物品有興趣，輸出1代表有興趣，0則表示無興趣。訓練完模型後，若想要知道user *A*是否對item *B*感興趣，只要將其各自轉換成one hot encoding後輸入模型，便可以透過模型輸出的數值得知使用者對商品的喜好程度。



圖二、文獻**[9]**的神經網路架構圖

除上述方法，另一派深度學習推薦系統是基於Auto-Encoder (AE)架構，這類方法又可分為兩大類，(i)使用AE對使用者或商品及其附帶資訊(side information，如個人資料或商品描述)進行特徵向量降維**[15, 22, 24]**，(ii)使用AE來預測使用者與物品互動行為**[25]**。在第一類的應用中，文獻**[24]**提出了一個架構來解決完全冷起啟動商品(complete cold start item)與非完全冷起啟動商品(incomplete cold start item)推薦，該架構會預測商品評分，藉此推薦高評分的新進商品給使用者。

這個推薦架構組合了Stacked Denoising Auto-Encoder (SDAE)模型和一知名推薦法timeSVD++**[11]**。對於完全冷啟動商品(即沒有任何使用者互動資訊的商品)，SDAE首先以冷啟動商品的商品描述作為模型輸入，並取用SDAE中間隱含層的輸出作為 商品的降維表示向量，該向量會帶入下列的timeSVD++公式來推測使用者對該全新商品的喜好分數。

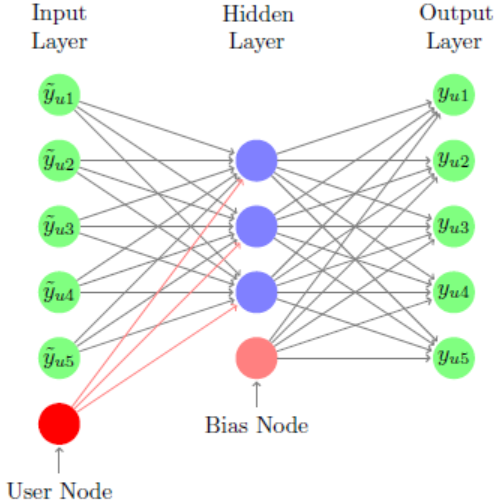


其中表示使用者*u*對商品*i*在時間點*t*的預測評分，表示使用者*u*在時間點*t*對商品的評分偏差(bias)，就是經過SDAE處理後商品*i*的降維向量，該向量會與使用者降維向量及所有使用者*u*評過分的商品平均降維向量進行內積運算，末項則是考量*M*個與商品*i*最相似的商品，依照使用者*u*給這些商品的評分及這些商品與冷啟動商品的相似度來推估商品的評分。針對非完全冷啟動商品(即有一點點使用者評分的商品)，模型預測評分的公式如下：



其中μ為所有items的平均評分，表示item *i*在時間點*t*的評分偏差，因為該商品有些許評分資訊，為該商品透過矩陣拆解後的低維向量。

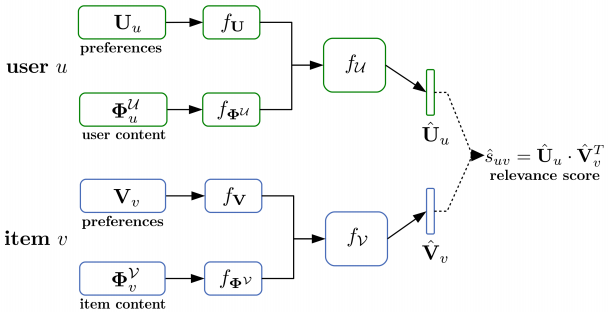
在第2類的應用中，文獻**[25]**提出了Collaborative Denoising Auto-Encoder (CDAE)模型，該方法將使用者與商品的互動行為以一向量表示，向量中的每一個維度對應到一個商品，而維度值表示使用者對商品的喜好，1代表使用者喜歡此商品；0代表使用者尚未看過此商品。如下圖所示，CDAE的輸入為使用者對各個商品的喜好值，除了向量，模型還需要使用者ID作為輸入，接著模型會隨機將某些輸入值設為0，藉此產生denosining auto-encoder訓練時的資料雜訊，再透過一層維度遠比輸入向量小的隱藏層來將含有雜訊的輸入還原回原始資料。透過大量的使用者物品互動資料，我們可訓練一可靠的喜好行為預測模型，待模型訓練完畢後，針對一個使用者，我們將其輸入CDAE，並從輸出的中挑選維度值越接近1且使用者先前未曾互動過的商品給該使用者。



圖三、文獻**[25]**的Denoising Auto-Encoder網路架構圖

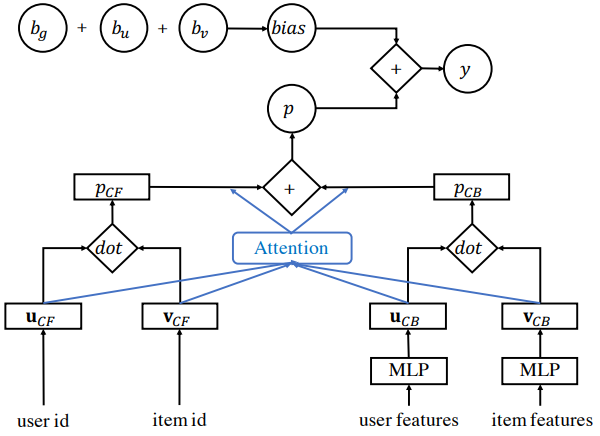
* + 運用深度學習之冷啟動推薦文獻探討

雖然以深度學習為基礎的推薦方法比其它方法有更強的推薦效能，但在面對新進使用者或新進商品時，仍然會面臨冷啟動的問題，這是因為這些新進使用者或商品只擁有非常少的互動資訊(如評分)，稀疏的資訊影響了深度學習的效果。因此，許多以深度學習為基礎的推薦方法也嘗試使用使用者或商品的附帶資訊來解決冷啟動推薦問題。如文獻**[23]**整合了preferences (即rating行為)和content等資訊當作深度學習模型的輸入，並使用了input dropout的機制來強化深度學習模型。下圖為該方法的網路架構，在訓練模型時，若遇到某個沒有評分資訊的cold start user (item)，該方法會將對應的user (item) preferences輸入設為0向量，這樣的設計稱之為「dropout」，其想法源自於denoising auto-encoder的noise產生程序，希望透過模型訓練的過程來還原出該cold start user (item)的latent preference representation。在學得使用者與商品的latent preference representation後，就可透過這些latent vectors的內積來估算使用者對商品的喜好評分。



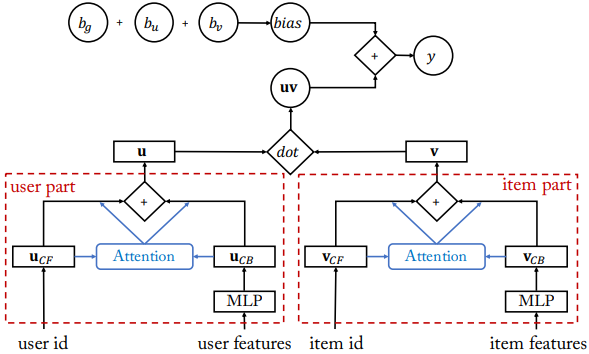
圖四、文獻**[23]**的網路架構圖

文獻**[21]**則是提出以注意力模型(attention model)為基礎**[1]**的冷啟動推薦系統ACCM – Attentional Content & Collaborate Model。如上所述，因為新進使用者與新進商品的行為資訊不足，該方法也考量了使用者與商品的內容資訊(content)，並透過attention model動態地給予內容資訊和歷史行為資訊(historical feedback)合適的權重以進行商品推薦。簡單來說，對於新進使用者或商品，由於他們的歷史行為資訊嚴重不足，該模型會傾向使用相對可靠的內容資訊來產生商品推薦。相反的，對於非新進使用者或商品，由於他們的歷史行為資訊較多，該模型會傾向於使用歷史互動資訊來進行商品推薦。下圖為該方法提出的兩個模型架構，分別被稱為result level attention和vector level attention。



圖五、文獻**[21]**之result level attention網路架構

在result level attention的模型架構中，模型的輸入為user id的one-hot encoding、item id的one-hot encoding、user content features (例如：年齡、職業)和item content features (例如：物品描述)。user id和item id的one-hot encoding會分別通過一層embedding layer得到向量，這些embedding layer中的權重參數會根據訓練時預測的商品評分值與真實評分值誤差來更新。user content features和item content features則是事先透過其他方法訓練出來的向量，而非與result level attention模型一起訓練，它們會分別通過一層MLP後產生出向量。接著計算向量內積來進行類似傳統協同式過濾以預測使用者對商品的評分該模型同時也會對向量內積計算以取得評分。有了評分與後， Attention model就會動態學習與各自的權重，再依權重對兩個評分加權得到評分。除了*p*外，該模型在輸出預測評分時還會考慮user bias 、item bias 和global bias 。



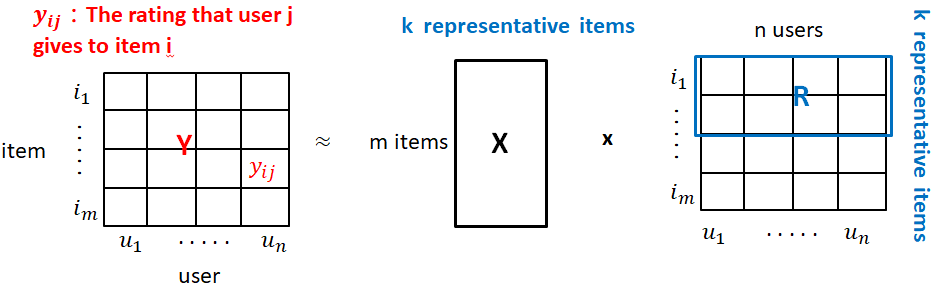
圖六、文獻**[21]**之vector level attention網路架構

Vector level attention的模型輸入與result level attention完全相同。唯一的差異是在於attention model會用在使用者向量*u*與商品向量*v*的產生階段，而不是對預測評分做加權。在得到使用者向量*u*與商品向量*v*後，該方法會將這兩個向量做內積來預測商品評分。如同result level attention，該模型同樣也會考慮user bias 、item bias 和global bias 。

* + 代表性商品探勘之文獻探討

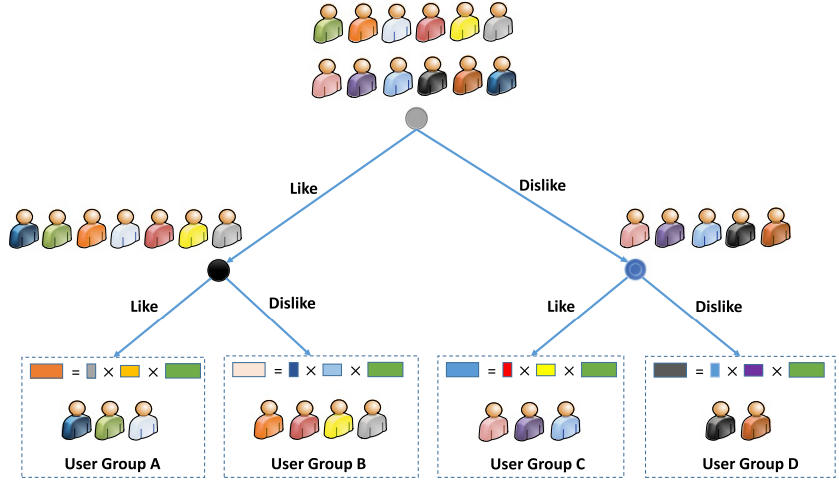
如前節所述，在缺少使用者商品互動資訊的情境下，多數冷啟動推薦方法都仰賴使用者與商品的side information，且多半是使用文字資料來推測使用者是否喜歡某商品。另一種解決冷啟動問題的方法是基於商品或使用者代表性(representativeness)，以冷啟動使用者商品推薦為例，不同於前述方法需額外蒐集有關使用者或商品的描述資訊，這類方法挖掘推薦系統中具代表性的商品，以及這些商品和其他一般商品的關係，當冷啟動使用者對一代表性商品有興趣時，這類方法會透過代表性商品與其它商品的關聯性來進行商品推薦。

文獻**[13]**提出一套名為RBMF (Representative-Based Matrix Factorization)的方法。如下圖所示，該方法將使用者對商品的評分以一矩陣*Y*∈表示，接著從這矩陣內找出*k*個具代表性的商品，透過這*k*個商品，該方法要將*Y*拆成兩個子矩陣*X*∈與*R*∈，其中*R*是*Y*矩陣內擷取出這*k*個代表性商品之使用者評分所組成，換言之，*R*是*Y*的子集合。確定*R*後，RBMF所要學習的模型參數是*X*矩陣，該矩陣描述這*k*個代表性商品與其它商品的關聯性，而模型的學習目的是要找出一組好的*X*，讓*X*與*R*能還原原始資料*Y*，即(*Y*≈*XR*)。該篇作者是透過maximal volume algorithm來選定*k*個representative items，此演算法能從原始矩陣中選出最能還原原始矩陣的*k*個columns來產生子矩陣*R，*作者認為這*k*個columns所對應的items就是representative items。



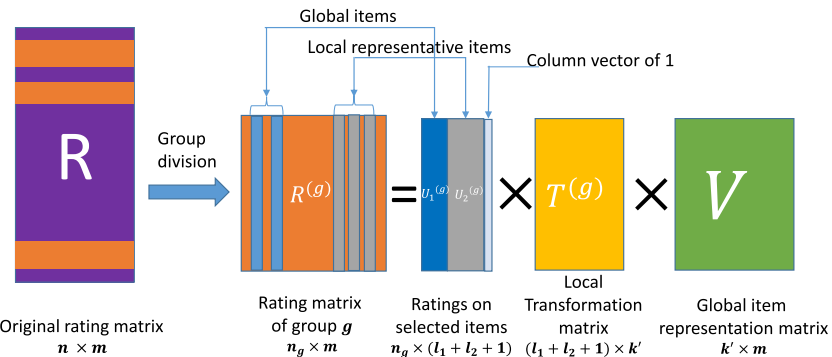
圖七、文獻**[13]**的矩陣拆解示意圖

文獻**[20]**改良**[13]**所提出的RBMF方法，相較於RBMF讓所有使用者共用一組具代表性商品，該論文作者認為不同屬性的使用者族群應該有不同的代表性商品。如下圖所示，該論文提出的LRBMF方法會先以決策樹的方式，依「全域具代表性商品」(global representative items)來將使用者分組。分組後，再根據各組的使用者組成來挑選「區域具代表性商品」(local representative items)。



圖八、文獻**[20]**之LRBMF使用者分群

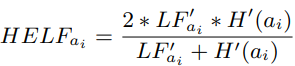
選取「全域具代表性商品」時的依據是希望各組所形成的評分矩陣與預測的評分矩陣誤差越小越好。而在選取「區域具代表性商品」時，該方法也是透過maximal volume algorithm來選出前*k*個「區域具代表性商品」。在決定好這些商品後，使用者向量的維度可由全域與區域具代表性商品共同組成，而如下圖所示，每個使用者群組*g*可由這些使用者對物品的評分來構成該組的評分矩陣，每列(row)代表該組的使用者，每行(column)代表一個商品。該方法會從所有商品中，依照第一階段選出的個全域具代表性商品，和第二階段利用maximal volume algorithm所選出的個區域具代表性商品，來形成矩陣*U*，而圖中另兩個矩陣與*V*便是模型訓練時要用來逼近(還原)的模型參數。待訓練完畢，商品推薦的生成會類似於RBMF，只要能取得某使用者對於「全域具代表性商品」與「區域具代表性商品」的評分值，便可以組成該使用者的向量*u*，再將向量*u*與模型參數、V做內積，就可以推測此使用者對所有商品的評分，進而推薦合適的商品給該使用者。



圖九、文獻**[20]**之LRBMF矩陣拆解

文獻**[7]**則是找出推薦系統中具代表性的使用者，該方法將所有使用者進行k-means分群，找出各群集最具代表性的使用者，再用這些使用者對商品的喜好關係來對群內其他使用者進行商品推薦。該方法將每群的重心使用者視為該群的代表性使用者，換句話說，該使用者需要與群集內其它使用者的相似度總和最大。而相似度的計算方式則是該論文的核心，該論文改良知名的similarity metrics (如cosine similarity、Pearson correlation)來避免這些similarity metric會偏愛highly active users的挑選偏差。如在Pearson correlation計算中乘上表示user *i*和user *j*共同評分過的item數量，：表示user *i*評分過的item數量。若兩個users共同評過分的items數量越多，則他們之間的相似度就會越高，但若user *i*或user *j*有一個人是highly active user時，則分母會越大，造成兩users間的similarity變小。在進行商品推薦時，該方法首先會找出目標使用者所屬的群集，然後推薦該群representative user喜歡的物品給目標使用者。

文獻**[16]**提出了Harmonic mean of Entropy and Logarithm of Frequency (HELF)方法來挑選具代表性商品。該篇論文的作者認為代表性商品必須被大多數使用者評分過，換句話說，代表性商品必須要夠熱門(popularity)。除熱門外，代表性商品的評分分布還必須夠分散，意即使用者對商品的評分若越不一致，則該商品所包含的資訊量越多，也越具代表性。HELF以商品的評分數與entropy作為popularity和資訊量的考量依據，並用下列調和平均數(harmonic mean)來綜合這兩指標，高HELF值的商品即為代表性商品。



其中為商品的評分次數取對數後再除以所有使用者人數來進行次數正規化(normalization)，商品評分分布之entropy。

* + - **章節結語**

由近期文獻可知，現行的冷啟動推薦方法鮮少採用DAE模型，文獻**[25]**與**[22]**雖採用DAE架構，但這些方法並不是在處理冷啟動推薦，且在對輸入向量做雜訊干擾時僅使用傳統的Gaussian noise、masking noise或是salt-and-pepper noise。這些雜訊產生方式雖然能強化DAE模型對原始資料的還原穩定度，但應用在推薦系統上卻沒那麼直觀與可解釋性。在本計畫，我們將嘗試運用DAE來解決冷啟動推薦問題，我們希望將DAE、代表性商品探勘和冷啟動推薦完美整合，並將DAE的雜訊產生視為一返使用者老還童過程，透過所設計的代表性商品探勘法來保留使用者對重要商品的互動資訊，且將輸入DAE的使用者商品評分向量回推(模擬)成冷啟動狀態，進而完成DAE模型訓練與後續的冷啟動商品推薦。

1. **研究方法、進行步驟及執行進度。請分年列述：1.本計畫採用之研究方法與原因及其創新性。2.預計可能遭遇之困難及解決途徑。3.重要儀器之配合使用情形。4.如為須赴國外或大陸地區研究，請詳述其必要性以及預期效益等。**

我們將逐年完成本計畫三大目標：(1)設計以DAE向量雜訊移除為基礎之新進使用者冷啟動推薦方法，(2)設計以DAE向量維度壓縮為基礎之新進使用者冷啟動推薦方法，(3) 整合DAE維度壓縮及雜訊移除並擴展至新進商品冷啟動問題。我們將於下列小節說明這三大目標的初步研究方法、可能遭遇問題及解決途徑。

* **第一年 – 設計以DAE向量雜訊移除為基礎之新進使用者冷啟動推薦方法**
  + - 1. **研究方法與原因**

如文獻探討所述，在缺乏使用者商品互動資訊的情況下，冷啟動推薦方法多使用使用者或商品的附帶資訊(side information)來產生商品推薦清單，但很多時候使用者並不一定有意願提供這些資訊，而在瀏覽近期推薦系統與深度學習的文獻時，我們發現DAE模型十分適合於冷啟動推薦，為說明為何採用DAE來解決新進使用者冷啟動問題，在本節，我們首先介紹DAE架構，說明它的學習目標與應用情境，再解釋DAE與新進使用者冷啟動推薦的關聯性。最後，我們將說明第一年研究期間可能遭遇的問題與解決途徑。

1. **DAE基本架構**

在深入了解DAE之前，我們必須對其原型Auto-Encoder (AE)有一定的認識。AE是由encoder和decoder兩部分所組成的，其目的是希望能找出一組好的encoder及decoder，讓輸入的高維向量資料透過encoder進行有效的壓縮降維，且降維後的資料能再透過decoder還原回原始資料，而這壓縮過的降維向量可視為原始輸入向量的簡約代表。Encoder通常以全連接(fully-connected)類神經網路為基礎，將輸入向量*x*映射成低維度向量*y*，我們習慣以函式 表示encoder，其公式如下所示。

= *y* =

其中的參數為，*W*是一個*KM*的權重矩陣(*KM*)，而*b*是一個維度為*K*的偏差向量(bias vector)，是sigmoid函式。透過函式，我們可將高維(*M*維)的資料*x*轉成低維(*K*維)的*y*。Decoder的架構通常與encoder對稱，而最簡約的decoder可以是一層全連接類神經網路，我們可用函式來表示，該函式會將低維向量*y*映射回與輸入向量相同維度的向量 。其公式如下所示。

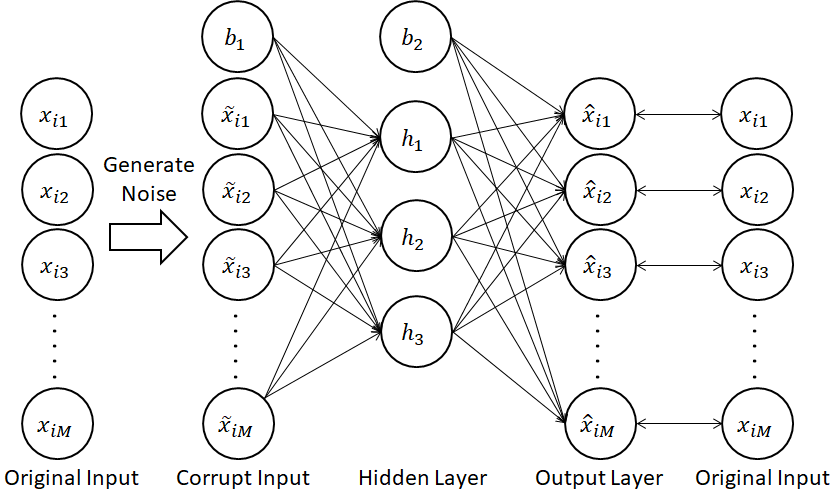
= =

其中的參數為，是一個*MK*的權重矩陣，而是一個維度為*M*的偏差向量(bias vector)，是sigmoid函式。

上述的模型(函式)參數是透過大量訓練資料求得的，假定我們有*N*筆訓練向量 *X* = {*x*1, *x*2, …, *xN*}，AE模型的訓練目標是要優化下列損失函式，即是要最小化輸入向量*xi*與輸出向量*i*之間的誤差總和。

其中是兩向量的平方誤差、代表第*i*筆訓練向量、為第*i*筆向量的模型輸出向量。

DAE是AE的一種變形，該模型的訓練目標仍是希望最小化輸入向量與輸出向量間的誤差，不過在訓練時，DAE會對輸入向量進行雜訊干擾，會將向量內部分的元素值修改，才將帶有雜訊的向量輸入encoder-decoder中進行參數尋找與輸入向量還原。透過雜訊產生的機制，DAE可以讓模型擷取出更有代表性的資料特徵，使模型不易因雜訊干擾就改變預測的結果。DAE的架構如下所示。



圖十、DAE深度學習架構

常見的雜訊產生方法有三種：第一種方法是Gaussian noise，Gaussian noise適用於當輸入向量的元素值是實數的情況，該方法會依照常態分布來產生雜訊值，並將雜訊附入原始向量中的素值值。第二種方法是masking noise，masking noise會隨機將輸入向量內部份素值值設定為0。第三種方法是salt-and-pepper noise，Salt-and-pepper noise會隨機將輸入向量內的素值值設定成該維度的最大值或最小值，如應用於彩色圖片，該方法會將圖片內部分像素(每一個像素會對應到一向量維度)設定成最小像素值0或最大像素值255。

總結DAE的訓練過程：

1. 首先，我們會將訓練資料集*X* = {*x*1, *x*2, …, *xN*}內的向量透過上述三種雜訊產生方法或自行設計的方法，隨機將訓練向量*x*映射成受雜訊干擾向量 。此過程可以表示為：

其中的為產生雜訊的機率分布。

1. 接著，將受雜訊干擾的向量 輸入encoder，使其映射到低維度向量*y*。此過程可以表示為：

= *y* =

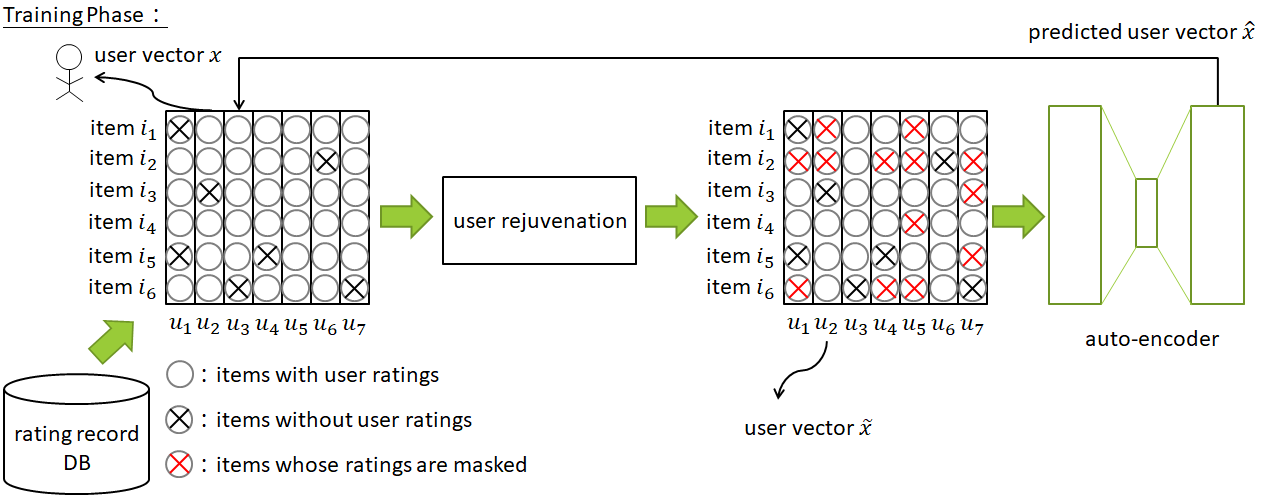
1. 最後，decoder將隱藏層的向量*y*映射回原輸入向量的空間中，產生輸出向量。此過程可以表示為：

= =

而DAE的模型參數、優化可以使用一般的「反向傳遞演算法」(backpropagation)，通過計算原始向量*x*與輸出向量間的差距(loss)來逐步由模型輸出層往輸入層傳遞，以此來更新參數。

1. **採用DAE架構原因與初步方法**

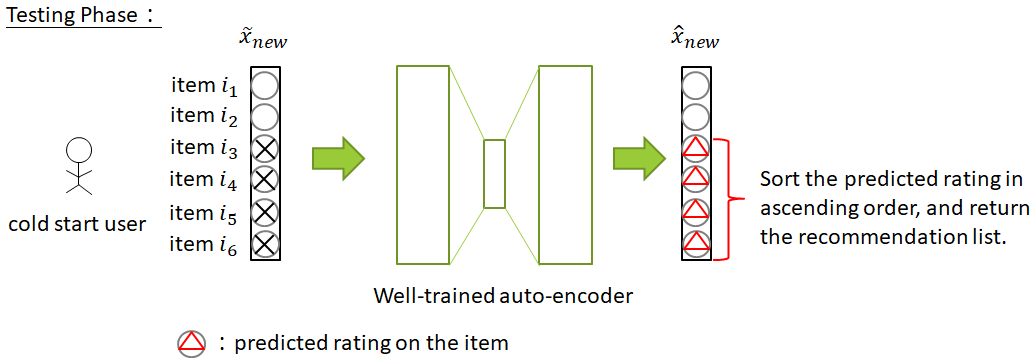
如上所述，DAE非常適合處理雜訊資料，並能有效將雜訊移除且還原原始資料。在處理新進使用者冷啟動推薦時，若我們將使用者對商品的喜好(如評分)視為DAE要恢復的資料，而DAE訓練期的雜訊產生程序視為使用者返老還童階段(user rejuvenation stage)，其雜訊產生既抹去使用者對商品的評分，藉此將使用者推回新進使用者狀態。則DAE的訓練目的就可轉化為將「把評分稀疏的新進使用者狀態成長(快轉)為將來評分豐富狀態」。待模型訓練完畢後，我們就可以用該模型來為新進使用者推測未來喜歡的商品。



圖十一、DAE冷啟動推薦之訓練架構

上圖為我們初步的系統架構圖，在訓練階段，我們會收集大量一般(非新進)使用者商品評分資訊，且把使用者以高維的商品評分向量*x*表示。向量內每一個維度元素對應到使用者對一個商品的評分。接著，使用者返老還童階段會考量商品代表性，把部分使用者評分從向量*x*中抹去，我們用符號 來表示這新進使用者狀態向量，而向量 在通過DAE的encoder-decoder後，模型會將其推算成向量 ，而DAE的訓練目標是希望這輸出向量 與使用者原始的向量*x*越接近越好。透過大量的*x、*，我們期望DAE模型能捕捉使用者如何由新進評分狀態轉化為後續評分狀態。

待DAE模型訓練完畢後，我們就可使用模型的encoder-decoder來進行新進使用者商品推薦，令為一新進使用者的商品評分向量，透過encoder-decoder可將其轉化成向量，我們可從中挑選高評分的商品來推薦給該使用者。



圖十二、DAE冷啟動推薦之新進使用者推薦流程

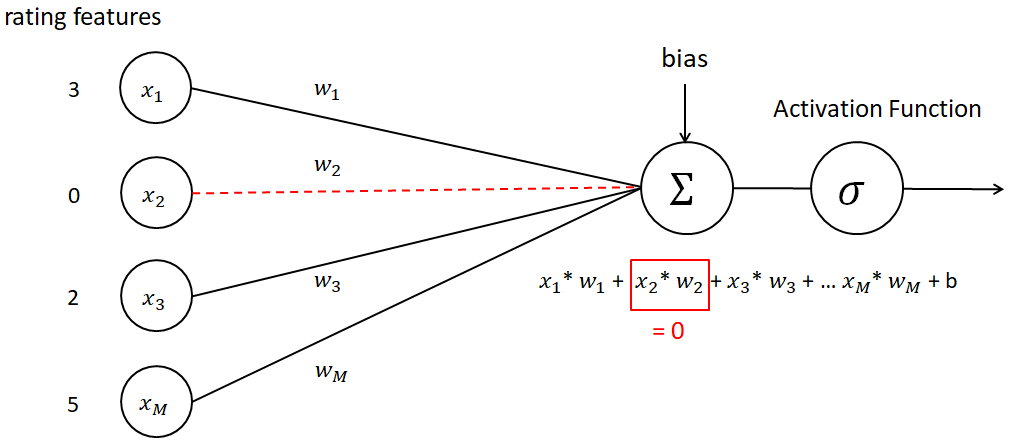
* + - 1. **可能遭遇之困難及解決途徑**

我們已考量到下列可能的挑戰並提出初步的解決方法。

1. **如何抹去使用者商品評分**

許多推薦演算法會將使用者未曾互動的商品以評分0來表示，但這種表示法常會與1至5分的商品評分產生混淆，讓推薦演算法誤以為使用者不喜歡該商品而影響推薦效能。為避免上述干擾，有些推薦演算法(如matrix factorization)會忽略這些沒有互動的商品，且只針對使用者有評分紀錄的物品來進行模型訓練。在返老還童階段，我們須抹去部分商品評分來將使用者推回新進狀態，這些要被抹去的商品評分該如何在使用者高維向量 內呈現則是一件需要審慎思考的問題。

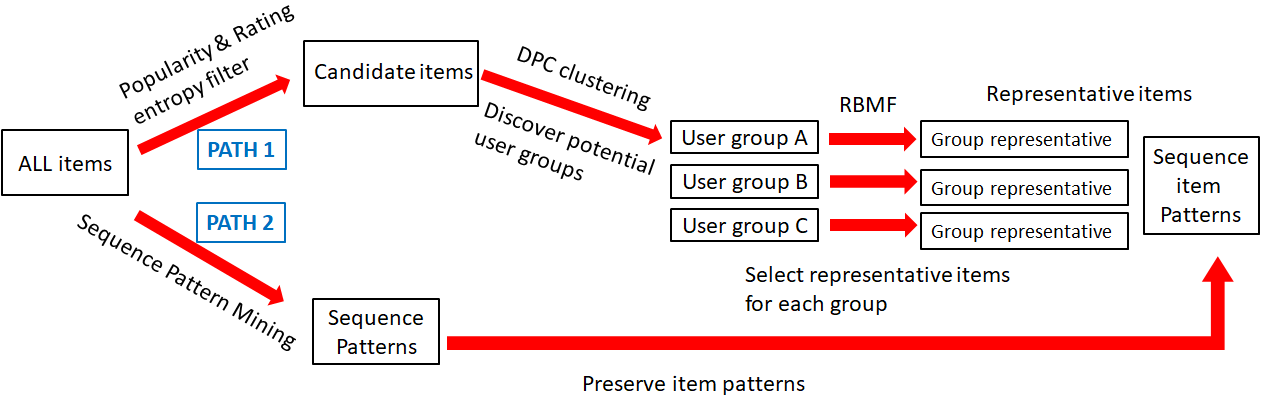
我們初步規劃在內以0來表示使用者未曾評過分的商品，換言之，被挑選要抹去的商品，以及原本沒有評分過的商品，在 都是以0分表示。我們採用這種表示法的原因在於，對神經網路之perceptron而言，一個0的輸入意謂該輸入對應到的特徵(feature)未啟動(inactive)。如下圖所述，在計算該perceptron的輸出時，這些未啟動特徵對perceptron的輸出貢獻值為零，也等於在訓練(測試) perceptron時，我們不考慮(忽視)這些特徵，這與我們想要抹去使用者互動商品的想法一致，也因此，這種以0的表示法應可以避免上述干擾。



圖十三、Perceptron與0值輸入之關係

1. **如何挑選合適的代表性商品**

圖十四為我們初步規劃的代表性商品挑選方法。在探討文獻時，我們了解不同的使用者族群應有不同的代表性商品，如喜歡3C類型的使用者應對美妝類的代表性商品沒有興趣。因此，在挑選代表性商品前，我們會先對使用者進行分群(clustering)，找出興趣相似的使用者族群，再分析不同族群使用者對商品的評分行為來挑選代表性商品。



圖十四、代表性商品挑選流程

我們雖然可計算使用者高維向量的相似度來進行使用者分群，但如先前所述，這些向量通常非常稀疏，會影響相似度計算與分群的品質，也因此，我們採用文獻回顧裡提到的HELF指標來將高維向量降維。HELF指標主要考慮兩個因素：其一是商品的熱門度(popularity)，也就是某個商品曾被多少使用者評分過；其二是使用者對某個商品評分的資訊量，該資訊量透過熵(entropy)來表示，意即使用者們對某商品的評分分散程度，舉個例子來說，若所有使用者對某商品一致給予5分的評分，則該商品有最小的熵值0，意即資訊量為零；反之，若所有使用者評分均勻分布在1~5分中，則該商品可獲得最大的熵值(資訊量)。HELF為上述兩因子的調和平均數，該值越大，代表一商品不但熱門，且大家對該商品的喜好意見分歧，換言之，該商品很有鑑別力。在進行使用者分群前，我們會先濾掉HELF值較小的商品來為使用者高維向量進行初步降維。

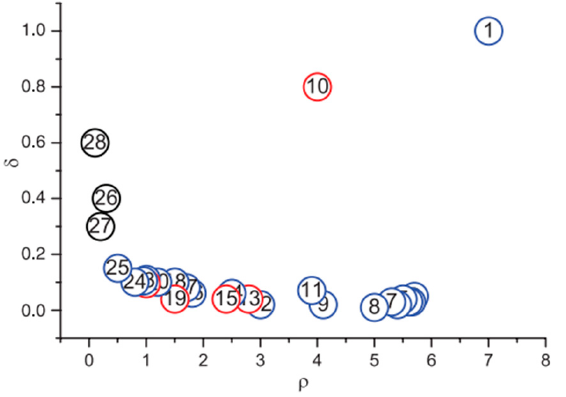
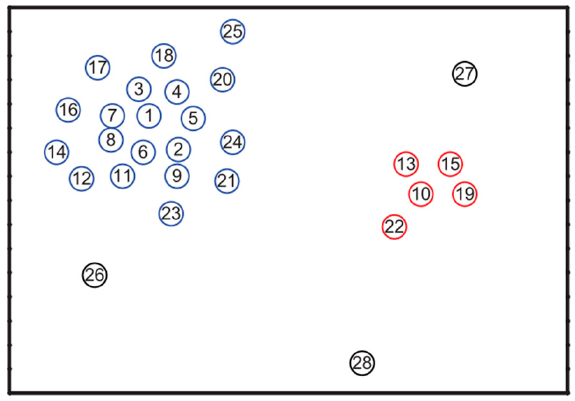
在將使用者以一維度較低的向量表示後，我們會運用Density Peaks Clustering (DPC)演算法**[18]**來將使用者分群。DPC是一種基於區域性密度的分群演算法，透過一個事先設定的距離範圍 (cutoff distance)，我們可算出資料點*i* (在此即為一個使用者向量)在範圍內所包含的其他資料點數量，該數值就是點*i*的區域密度(local density) 。的計算公式如下：

其中資料點*i*與資料點*j*的距離，該值可採用歐式距離(Euclidean distance)計算求得。得到區域密度後，DPC便會依據算出群體中心相對距離(relative distance) ，計算的方法如下公式所示：當計算出所有資料點的區域密度後，我們將區域密度由大到小排序，假設大小關係為 > > > ... 。對於密度最大的點*i*，它的群體中心相對距離會等於與點*i*相距最遠的那個點*n*的距離。而其他非最大密度的點(如點*k*)之群體中心相對距離等於密度大於該點的點集合中(即*i*、*j*)，與該點距離最小的點距離。

若為所有資料點中，區域密度最大的點

若為所有資料點中，區域密度最大的點

DPC分群演算法會依據區域密度和群體中心相對距離來產生如下的資料中心決策圖(x軸：、y軸：)。



圖十五、DPC分群與資料中心決策圖(截圖於**[18]**)

上圖左為28個資料點在資料空間的分布情況，資料點上的數字是依照區域密度由大到小進行編號，即編號1的區域密度最大、編號28的區域密度最小。右圖則是將這28個資料點依其 ( )所繪製的資料中心決策圖，*x*軸為區域密度、*y*軸為群體中心相對距離。從右圖可明顯看出當與皆大時，表示這些點(如點1、10)的區域密度很大，且密度比這些點大的點集合，離這些點的距離很遠，這些點因此很適合當作群聚中心。我們將會考量資料於資料中心決策圖的分布來挑選合適的使用者作為群聚中心，再把其他使用者依照距離大小指派給不同群體，完成使用者分群工作。

在將使用者分群完畢後，我們就要挑選每群的代表性商品。直覺上，由於我們先前已經透過HELF濾掉一些鑑別力差的商品，我們似乎可以直接挑選每群使用者偏愛的商品作為該群的代表商品，但HELF只考慮商品自身的熱門度與資訊量，它並沒有考量商品間的關聯性，也因此，我們可能會挑選過多同質性商品而影響了代表性商品的多樣性。

我們初步規劃以文獻回顧裡提到的RBMF挑選每群的代表性商品。RBMF是一種矩陣拆解演算法，我們一開始先將所有使用者的評分以一個評分矩陣∈ 來表示，其中*N*代表系統中使用者的數量，*M*代表經過HELF指標篩選過的商品數量。假設DPC將使用者分為A、B、C三群，我們先依照使用者分群結果將分割成三個評分子矩陣∈ 、∈ 、∈ ，其中、、分別代表不同群的使用者數量。接著我們再運用RBMF之Maximal Volume Algorithm分別從這三個評分矩陣中挑出*K*個列，使拆解後的子矩陣能最逼近原始矩陣，而這*K*個列所代表的商品即為該群的代表性商品。

透過上述流程，我們相信所挑選的代表性商品不但具有熱門性與資訊量，還能兼顧商品評分關聯來找出多樣性的代表商品。

1. **能否運用使用者與商品的互動時間戳記來進行返老還童**

使用者返老還童的目的是希望能將使用者評分向量推回成新進使用者狀態，藉此讓DAE捕捉使用者後續的評分演變來進行新進使用者冷啟動推薦。我們假設使用者會進入一服務平台是受到代表性商品吸引，且竭力保留代表性商品來將使用者回推成新進狀態，但若我們能拿到使用者與商品互動的時間戳記，則能更精準的回推使用者狀態。

運用時間戳記來回推使用者狀態可視為一種時序性樣式探勘問題(sequential temporal pattern mining)**[14]**。舉例來說，若我們將使用者的商品評分以一序列來表示，如序列{*i*5:4, *i*21:3, *i*79:5, *i*42:2, *i*37:4}代表一使用者最先評分的商品為*i*5，且評分為4分，以此類推。則在收集大量使用者評分序列後，我們可運用或改良現有sequential temporal pattern mining演算法來找出商品評分的順序性，如在時序探勘後發現有不少使用者會從*i*21開始評分，接著評*i*42與*i*37。我們可依照挖掘出的時序樣式把後續發生的商品評分從使用者評分向量移除，藉此回推使用者初始狀態。值得一提的是我們的問題情境會比傳統的sequential temporal pattern mining更複雜，這是因為我們除了要考量商品的評分順序，還需要兼顧評分的一致性，如一開始評*i*21高分跟一開始評*i*21低分應屬於不同的時序樣式，這些挑戰都有待我們在執行計畫時克服。

1. **是否要為不同使用者群體訓練各自的DAE模型**

在為使用者分群且選出各自的代表性商品後，我們需思考是否要為每個群體訓練各自的DAE模型。又或者不分群體，直接集結所有使用者資料來訓練單一DAE模型。為每個群體建立一個DAE模型的好處在於這些模型有機會捕捉到不同族群的評分演變過程，但在訓練完模型後，我們會遇到一個困難的問題，就是我們難以知道要進行商品推薦的新進使用者是屬於哪一個族群。這是因為新進使用者只有少許的商品評分，因此很難判斷使用者所屬的群體。也因此，我們初步規劃將所有使用者資料集結起來訓練一個整體的DAE模型。進行推薦時，我們就不須煩惱上述問題，直接將新進使用者的評分向量輸入訓練好的encoder-decoder來進行商品推薦。在執行計畫期間，我們也會思考上述問題，並期待能採用族群導向的DAE模型來提高推薦效能。

1. **Encoder-decoder能否套用複雜的網路架構**

我們目前規劃以MLP做為encoder-decoder網路架構。但我們發現，許多處理圖像或語音的encoder-decoder會用到複雜的網路架構，如圖像類的encoder-decoder常使用CNN來進行資料編碼與解碼，且有不錯的效能表現。雖然如此，我們的問題情境卻不適用CNN或RNN來進行編碼與解碼，這是因為我們的輸入向量是使用者對商品的評分，所有維度都是彼此獨立，不似圖像的向量維度是由像素所構成的，相鄰的維度即為相鄰的像素，CNN等考量資料鄰近關係的網路架構因此能利用這相鄰性來抽取不同精細程度的資料特徵。

但如先前時間戳記討論所言，若我們能歸納出商品評分的時序關係，則能將要輸入模型的使用者向量進行維度整理，如我們若發現商品*j*總是在商品*i*評過分後才被使用者評分，則我們可將商品*i*的維度擺在商品*j*之前。雖然商品時序性極有可能是一個partial order，但如果我們能有辦法利用時序性來為商品產生一個topological sort，並依照排序來重整向量維度，則將有機會套用複雜的網路架構。

* **第二年 – 設計以DAE向量維度壓縮為基礎之新進使用者冷啟動推薦方法**

**研究方法與原因**

DAE另一個強項在於能有效取出資料特徵，將高維資料以低維向量表示。在第二年的研究，我們將發揮DAE這項優點，整合DAE與協同過濾法(collaborative filtering)於新進使用者冷啟動推薦，並延伸這個想法來取出商品代表向量，且結合這兩種低維向量表示法來進行新進使用者冷啟動推薦。

協同過濾法是一種常見的推薦方法，其推薦法可細分為二種：基於使用者的協同過濾法(user-based collaborative filtering)與基於商品的協同過濾法(item-based collaborative filtering)。基於使用者的協同過濾法會將使用者以一特徵向量表示，透過向量相似度來找出與目標使用者特徵相似的參考使用者，再分析這些參考使用者對商品的喜好來進行商品推薦。基於商品的協同過濾法也是將商品以特徵向量表示，再透過相似度計算找出與目標商品類似的參考商品，並推薦目標商品給喜歡這些參考商品的使用者。

協同過濾法雖然簡單，但它卻易受資料稀疏性影響，以user-based collaborative filtering為例，該方法會透過使用者對商品的評分(或其他互動行為)來將使用者表示成高維向量，每一向量維度代表一商品，向量維度值為使用者對商品的評分。由於電商服務平台提供的商品繁多，有使用者評分的商品寥寥可數，這種現象會造成上述高維向量非常稀疏，影響協同過濾法尋找可靠的參考使用者，進而拖累推薦效能。為減少這種稀疏現象對推薦系統的影響，許多推薦系統會將使用者(商品)高維向量進行降維，讓使用者以一低維向量來表示，藉此減緩稀疏現象的影響，且提高推薦效率。AE的主要任務就是要將高維資料降維，萃取出低維度的資料特徵，而DAE更是能有效壓縮受雜訊汙染的高維資料，產生更穩健(robust)的特徵代表。在本計畫第二年的執行期間，我們將透過DAE來將使用者表示成低維向量，並利用這些使用者向量來進行基於使用者協同過濾。

圖十六為我們初步規劃的系統架構圖。我們的方法可分成兩大步驟，步驟一為「利用DAE進行使用者向量降維」，步驟二為「基於降維向量之使用者協同過濾」。在步驟一，我們會將使用者對商品的評分以一評分矩陣∈ 表示，矩陣中的每一行紀錄一使用者*i*對所有商品的評分，我們將使用者以該行向量表示，而所有使用者向量會形成DAE的訓練資料來進行模型參數訓練。每一筆高維向量會先經過雜訊產生階段將原始向量轉化成受雜訊汙染的向量，接著將輸入encoder-decoder且產生輸出向量訓練的目標是要最小化的誤差(訓練目標函式與過程可參考第一年計畫內容)，讓DAE能有效還原未受雜訊汙染的使用者高維向量。

在完成DAE訓練後，我們可取出每個使用者於encoder的產出，即模型內隱藏層的數值向量，向量的維度為，且，該向量即為原始向量的低維向量表示，也可視為每個使用者的低維特徵代表(embedding)，我們會利用這使用者embedding來進行基於使用者的協同過濾推薦。



圖十六、基於DAE降維向量之協同過濾推薦

步驟二將進行基於使用者降維向量之協同過濾推薦。此步驟包含「預測」與「推薦」兩階段，在預測階段，我們會從步驟一取得每個使用者的embedding，並將這些embedding組成降維特徵矩陣，每一行為一個使用者的embedding。對一位目標使用者，我們利用這些embedding來計算其他使用者與目標使用者的相似度，而相似度的計算方法是採用向量空間常見的餘弦相似度(cosine similarity)，其公式如下：

其中為向量內第*d*個元素值。該值越大，代表使用者*p*與使用者*q*越相似。我們會挑選相似度前高的使用者為參考使用者，利用這些參考使用者對商品的評分來推算目標使用者*p*對未評分商品的評分值，令參考使用者集合，而對商品有評分的使用者集合，代表使用者對商品的評分。依照下列公式，我們可計估計目標使用者對商品的評分：

最後，在推薦階段，我們用商品預測評分來進行商品排序，並將排名前幾名的商品推薦給目標使用者。

**可能遭遇之困難及解決途徑**

我們已考量到下列可能的挑戰並提出初步的解決方法。

1. **是否要正規化使用者商品評分**

如上所說，我們的方法會從評分矩陣*R*內取出使用者高維向量*x*，並用這些向量進行DAE訓練與找出降維向量*y*。但若將*x*原封不動的當成訓練資料，則訓練出來的模型可能會受用者評分傾向影響。評分傾向意指不同使用者對商品會有不同的評分基準，如有些人的基本評分為4分，除非非常討厭一個商品，否則大部分的商品會在4分上下震盪。但也有些人的基本評分是3分，所給的商品評分會在3分左右，此時，這位使用者所給的3分跟前述使用者的4分有相同的意義，但對類神經網路而言，不同的輸入值代表不同的資訊，也因此，我們應對使用者評分進行正規化，消除不同使用者的評分傾向，讓模型能正確捕捉大量使用者評分關係，且取出具代表性的降維向量。

許多推薦方法都意識到使用者評分傾向對推薦系統的影響，並設計模型參數來消除不同使用者的評分傾向。如文獻回顧裡提到的timeSVD++就使用偏差變數(bias variable)來解決上述問題。我們初步提出一個簡單的方式來進行評分正規化，對每個使用者向量*x*，我們會針對其內非零的商品評分進行平均(或取眾數)，並視平均值為該使用者的基礎評分，接著再將所有非零的評分減去這個平均評分，如此，*x*內的正值元素表示使用者對該商品較為有興趣，負值元素表示使用者不喜歡該商品，而元素值越接近零的商品表示使用者對這些商品無感。

上述方法雖是一個常見的正規化方法，但我們須小心在正規化後，*x*內的零值元素會有兩個含意，一個是該商品的原始評分為使用者平均評分，即使用者對該商品無明顯喜好。另一個是使用者未曾評過該商品。雖然這兩種情況差異很大，但我們初步認為這差異對模型訓練沒有太大的影響。如我們在第一年可能遭遇之困難及解決途徑(1)所言，對perceptron而言，一個0的輸入意謂該輸入對應到的特徵(feature)未啟動(inactive)。在計算該perceptron的輸出時，這些未啟動特徵對perceptron的輸出貢獻值為零。在正規化後，我們期望透過輸入元素的正負值來讓DAE捕捉使用者對商品的喜惡，因此，沒有強烈喜好的商品可以淡化無視，而將其與未評分商品設定為零，模型就不會考慮這些商品對模型參數的影響。

1. **DAE訓練時是否要排除新進使用者**

在第一年的研究中，我們在準備DAE訓練資料時會排除新進使用者，這主要是因為第一年研究的目的是要訓練一個能從新近使用者狀態還原回豐富評分狀態的DAE，未達此目的，訓練資料必須為評分豐富的使用者向量，而不能是評分稀疏的新進使用者向量。

但在第二年的研究，我們的目標是要透過DAE找出具代表性的使用者低維向量，無論新進使用者或評分豐富使用者，模型都要能有效找出他們的代表向量。若只用豐富向量進行模型訓練，所產生的encoder-decoder可能無法有效捕捉新進使用者的代表向量，因此，在這項研究主題，訓練資料並不會排除新進使用者。

1. **能否同時將商品降維且用於推薦**

我們初步只將使用者降維向量來進行商品推薦，但許多以維度降維為基礎的推薦法(如matrix factorization)除了將使用者向量降維外，還會同時求出維度相等的商品降維向量，再利用向量相似度來推薦使用者合適商品。使用降維向量的好處在於降維向量常被視為資料的潛在因子(latent factor)，如某維度可能代表一特殊商品類別，而依照潛在因子產生的商品推薦通常較貼近使用者需求。

若要從評分資訊擷取出商品降維向量，我們可將前述初步方法應用在評分矩陣*R*的列上，但因為商品與使用者高維向量的維度數不同(商品高維向量的維度為*N*而使用者高維向量維度為*M*)，我們需額外訓練另一個DAE模型，且將*R*內每一列所代表的商品高維向量當作訓練資料。雖然我們可透過這樣的程序求出商品低維向量，甚至在架構DAE模型時要求商品低維維度數與使用者低維維度數相同，但我們不能直接計算兩者的相似度以進行商品推薦，這是因為這兩組低維向量沒有維度應對關係，如商品的低維向量之第一維度與使用者低維向量之第一維度彼此毫無關係，因此無法運用傳統的向量相似度(如cosine similarity)來計算彼此的關聯性。為善用這兩組低維向量，我們初步規劃在原本的DAE上架上一組DNN來進行商品推薦，初步架構圖如下所示。



圖十七、整合DNN與DAE之推薦架構

整合後的系統包含三大步驟，步驟一為「利用DAE建立特徵代表」，步驟二為「利用深度神經網路預測商品評分」，步驟三為「建立推薦列表進行商品推薦」。步驟一主要目的是建立每位使用者及每個商品的低維向量(embedding)，此步驟即使用DAE來將使用者與商品的高維向量進行向量降維。當給定評分矩陣∈ ，矩陣中的每一行*xn*為一個使用者的高維評分評分，每一列*im*為商品的高維受評分向量。我們分別將這兩組高維向量餵給不同的DAE模型來進行參數訓練，在訓練完畢後，這兩個DAE模型會各自在隱藏層的地方產生低維向量*yn*與*zm*，其向量維度為，且或，而這兩組低維向量即為使用者與商品的embedding。

如先前所述，由於這兩組低維向量的維度並沒有對應性，我們無法直接運用向量相似度計算商品與使用者的關聯性，為挖掘出這兩者的關係，步驟二會先將使用者與商品的embeddings連接(concatenate)起來，並將其做為一組深度神經網路的輸入，每一層隱藏層的運算可以透過下式表示：

其中為第*l*層的輸出，該輸出同時也是第*l*+1層的輸入，而到上面提到的接合embedding，*Wl*為該層的網路參數，*bl*為對應的偏差向量，sigmoid啟動函式。而網路輸出層的計算如下：

我們令網路的輸出為預測的使用者對商品評分，模型訓練時的損失函數可以表示為：

換言之，在訓練時，我們希望網路預測的商品評分訓練資料內的實際評分差距越小越好。如同一般的深層網路訓練，我們將使用backpropagation的方式，不斷的套用stochastic gradient descent來學習參數每一層的參數值。

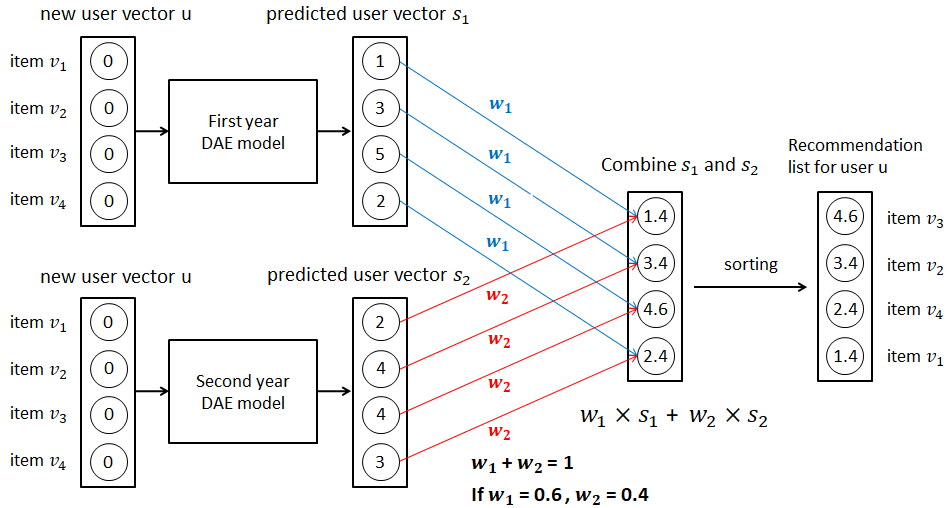
步驟三為推薦階段，當我們訓練好DAE與上述深層網路後，我們就可以利用這些模型來預測一位新進使用者是否會喜歡某一商品，對一位新進使用者與一商品，我們先用DAE模型取出它們各自的embedding，接著接合這兩個embeddings且將其輸入到步驟二的深層網路，該網路會預測使用者對該商品的評分值，我們可將所有商品依照預測評分值大小排列，並產生Top-N推薦清單給該使用者。

* **第三年 – 整合DAE維度壓縮與雜訊移除並擴展研究新進商品冷啟動推薦**

**研究方法與原因**

在本計畫最後一年，我們將整合前述兩大方法以期能提升新進使用者冷啟動推薦效能。我們想整合這兩套推薦方法的原因在於，這兩套方法是透過不同角度來解決新進使用者冷啟動問題。第一年的方法是從新進使用者自身的稀疏向量著手，推論後續可能的喜好商品。而第二年著重在運用協同式過濾，考量其他行為相似的使用者來進行商品推薦。我們期望能順利整合這兩套方法，進而產生多元的商品推薦。

我們初步的想法是透過權重組合的方式來將不同的商品推薦進行整合。如下圖所示，對一位新進使用者，我們先運用前兩年所設計的方法各自產生商品推薦評分，我們把所有可推薦的商品及其評分用一個向量*si*表示，向量的維度為所有可推薦的商品，其元素值為某方法對一商品的評分預測。舉例來說，若*s*1表示第一年的方法對新進使用者的商品推薦評分向量，則*s*1內第一個元素代表該方法對第一項可推薦商品的預測評分。我們將會給予這兩個方法一個權重值*w1*與*w2* (*w1, w1 ≥ 0*且*w1 + w2* = 1)，而最後整合的商品推薦評分就是將這兩個評分預測向量進行權重組合。在產生最後的評分向量後，我們可依照其評分值來將商品進行排序，並推薦該新進使用者高評分的商品。



圖十八、初步推薦整合架構

新進商品冷啟動(new item cold start)推薦也是一項實用性極高的推薦議題。由於新商品鮮少有使用者互動資訊(如評分或購買)，推薦系統難以分析可能的潛在用戶來進行有效的商品推薦，而這種現象常會產生惡性循環，讓新商品越來越冷門，最終被其他推薦商品淹沒。新進商品冷啟動推薦除了可以避免上述問題，讓電商服務平台提供更多元的商品外，它還可以預測商品的受歡迎程度，協助平台業者調整行銷策略。前述方法(包含第一年與第二年的內容)雖然專注於新進使用者冷啟動問題，但透過不同的向量表示方式，這些方法將有機會用於新進商品冷啟動推薦。在本計畫尾聲，我們將微調所設計的方法來嘗試解決新進商品冷啟動問題。如我們可將商品以一高維向量表示，每一個維度代表一個使用者對該商品的互動行為(如評分或採買)，如此，我們也能為商品向量進行返老還童與降維壓縮，並借助DAE模型來推薦新商品給合適的使用者。

**可能遭遇之困難及解決途徑**

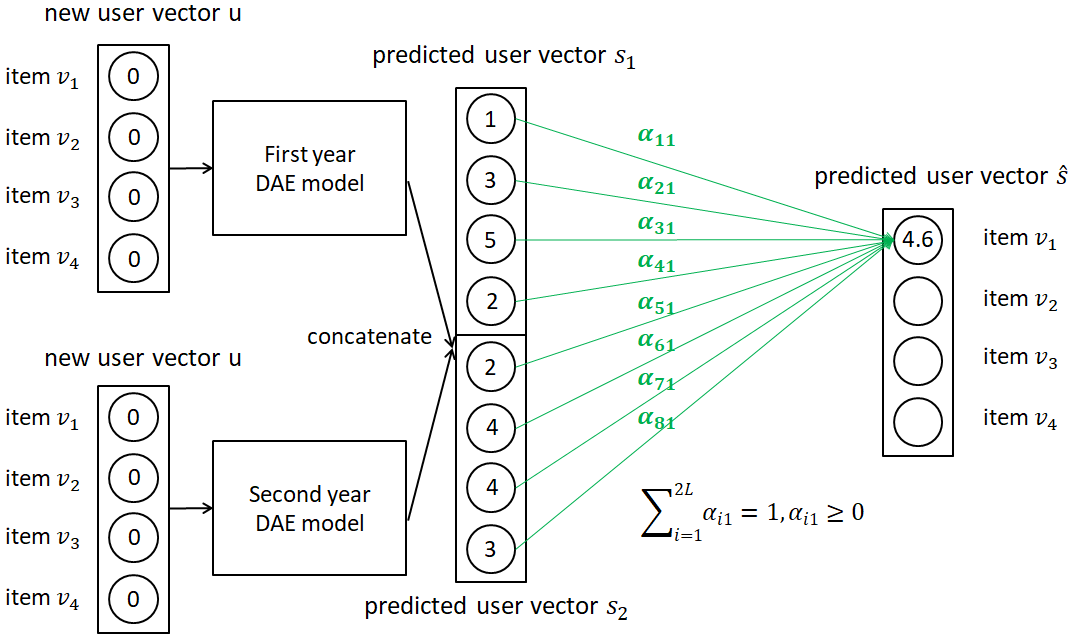
我們已考量到下列可能的挑戰並提出初步的解決方法。

1. **如何決定評分向量組合權重**

如上所述，我們會對第一年與第二年方法產生的商品評分向量給予不同的權重，再用線性組合的方式來產生最後的商品推薦清單。這些權重可以是人訂的，其好處是可依照不同策略調整推薦特質。如當我們想要增強推薦的多樣性時，我們可給予*w*2較高的權重。

我們也正在構思如自動產生合適的權重，我們初步的想法是運用「注意力模型」(attention model)**[1]**，藉由大量的使用者真實評分與各方法的預測評分來學出合適的權重，並產生最終的商品評分預測。注意力模型也是一種深度學習架構，它常用於解決序列對應(sequence to sequence)的問題，如語言翻譯。以英翻中為例，該模型的核心是，在決定一個翻譯輸出時，模型要能自動去決定哪些英文字詞是重要的，而英文字詞重要性(即每一個英文字詞的權重)會隨著不同翻譯輸出而自動調整，不是一個定值。例如，在翻譯I love running along the river bank時，attention model會逐一掃過英文字詞來產出中文翻譯，每個翻譯產出都會參考原文內所有字詞，且調整這些字詞的權重來尋找合適的翻譯，如在翻譯bank時，attention model會給予river與bank較高的權重，讓bank有較高的機會翻成河岸而非常用的銀行。

我們初步的架構如下圖所示，針對一位新進使用者，我們先用第一年與第二年的方法來產生商品評分預測向量*s*1與*s*2，接著將這兩個向量合併，若可推薦的商品數量為*L*，該合併向量的維度就會為2*L*。attention model會輸出一個*L*維的向量，其中第*j*個元素為商品*j*經過attention model考量*s*1與*s*2後的整合預測評分。如圖所示，與上述人為權重的方式不同，在產生該預測分數時，我們除了*s*1與*s*2各自對商品*j*的評分外，我們還會一併考量其他商品的評分，並用一組權重: 來進行最後的評分運算，這些權重將透過大量的*s*1與*s*2及使用者對商品的真實評分來進行attention model訓練，待模型訓練完畢後，我們就可以整合第一年與第二年的方法來進行新進使用者商品推薦。值得一提的是使用attention model可將商品間的關係納入評分預測的考量，如同上述river bank的翻譯情境，在計算商品*j*的整合評分時，若商品*i*與商品*j*的評分很有連動性，除了各方法對*j*的預測分數外，attention model會同時透過高權重將商品*i*的預測分數納入最後的評分計算。



圖十九、Attention model為基礎之推薦整合架構

1. **能否整合第一年計畫所提到的使用者分群機制**

在第一年的研究中，我們將對使用者分群且選出各群的代表性商品，並在返老還童階段盡量保留各群使用者對其代表性商品的評分。但在訓練DAE時，我們卻選擇總和所有使用者來訓練一個整合的DAE模型，而非為每個使用者族群建立獨特的DAE。若我們能在上述attention model取得成功，則我們可將相關的經驗與設計方法延續到這個議題。簡單來說，我們可以依使用者分群結果來為每個群體建立一個DAE模型，這樣的好處是這些模型有機會捕捉到不同族群的評分演變過程。在推薦階段，面對一位新進使用者，我們不判別該使用者所屬的群體，而是將新進使用者的稀疏評分向量輸入每個群體訓練好的encoder-decoder，讓這些encoder-decoder各自產出每個群體可推薦商品的評分(向量)，我們再運用attention model來調配這些評分向量，完成新進使用者商品推薦。

* 章節結語

在本多年期研究計畫，我們將研究新進使用者冷啟動推薦這個困難的研究議題。學術方面，我們會深入探究近年來熱門的深度學習技術，設計合適的返老還童機制來回推使用者成為新進狀態，並結合denoising auto-encoder這個深度學習架構來進行商品推薦。此外，我們還會研究encoder產出的代表性向量，且整合協同式過濾法來進行多元商品推測。最後attention model將協助我們整合這兩大方法的推薦清單，藉此整合不同推薦面向來提升新進使用者冷啟動推薦效能。實務方面，除了使用知名測資來評估我們方法的效果，我們也會尋找可能的產學合作機會，實際應用我們的方法到多種不同網路服務平台，以擴大本計畫研究影響力。

1. 預期完成之工作項目及成果。請分年列述：1.預期完成之工作項目。2.對於參與之工作人員，預期可獲之訓練。3.預期完成之研究成果（如實務應用績效、期刊論文、研討會論文、專書、技術報告、專利或技術移轉等質與量之預期成果）。4.學術研究、國家發展及其他應用方面預期之貢獻。
2. **預期完成之工作項目：**

我們第一年預計完成的工作為：主題趨勢預測及文字特徵擷取相關文獻探討、以排序學習為基礎之文字特徵擷取演算法設計、台灣景氣資料蒐集、效能評估模式建立、研究成果整理、論文撰寫並投稿。

第二年預計完成：文字特徵擷取文獻探討、混合式文字特徵擷取設計、多型態主題趨勢資料收集、系統效能評估、研究成果整理、論文撰寫並投稿。

第三年預計完成: 推薦演算法文獻探討、潛在興趣模型建立、以推薦演算法為基礎之趨勢預測模型開發、系統評估、主題趨勢預測網站建立、研究成果整理、論文撰寫並投稿。

1. **對於學術研究、國家發展及其他應用方面預期之貢獻：**

對政府及企業決策者而言，精確的主題趨勢預測是重要且必須的，在快速變化的社經形勢下，決策者可透過主題趨勢預測結果制定合適的政策以因應環境變化，本計畫將研究這重要且實際的研究議題。為提升主題趨勢預測效能，我們會著手改善文字特徵選取及趨勢預測模型，我們將使用排序學習從小量高排名文件集中取得大眾關注的主題關鍵字，並將其搜尋頻率視為觀察指標以進行趨勢預測，我們更會結合既有特徵選取技術以發展混合式文字特徵擷取法，並將研究成果應用於不同型態的趨勢主題。此外，我們還會大膽嘗試以推薦系統為架構的趨勢預測模型設計。我們期望本計畫的研究成果可有效提升政府及企業的決策能力，幫助政府及企業在快速變化的社經形勢中制定即時且正確的政策，以降低資訊延遲對政府及企業所造成的損失

1. **對於參與之工作人員，預期可獲之訓練：**

本計畫的研究重點是排序學習、文字特徵選取及推薦演算法。研究人員(含博/碩士生)可藉本計畫學得上述相關技術，我們會研讀近期國際知名研討會與期刊論文，並藉此訓練研究人員閱讀與評鑑論文的能力。此外，由於本計畫為一新興研究議題，許多實作環節(如資料收編與實驗設計)是必須自行設計與組織的，除論文研究訓練，我們還期望藉本計畫來訓練研究人員獨立思考與團隊合作能力。

學理方面，我們將專注於排序學習且研討各種文字特徵選取法，此外，我們會分析矩陣拆解式推薦演算法，並用其發展主題趨勢預測模型。參與本計畫的人員因此可學習高等線性代數、特徵選取、文字探勘及機器學習等知識。

實務方面，研究人員將建構並管理搜尋引擎關鍵字搜尋頻率以收編研究所需資料，由於趨勢相關資訊是不斷產生且型態多變，這無疑是一種大數據，我們因此期許研究人員能學習大數據相關的環境建置技術，如建立叢集運算(clustering)來加速模型建置與分散式處理。同時，研究人員還需撰寫最佳化學習演算法並開發可供線上查詢的主題趨勢預測系統，我們期望藉由這些實作機會來增進研究人員機器學習、文本分析及大數據處理的實務經驗。

1. **預期完成之研究成果及績效（如期刊論文、研討會論文、專書、技術報告、專利或技術移轉等質與量之預期績效）：**

學術方面，我們目前已進行了主題趨勢預測及相關文字特徵擷取的文獻探討，亦整理了前人研究以界定本研究之獨特性。我們也開始進行相關模型設計與程式撰寫，預計明年(2017)上半年就能將初步研究方法與演算法設計完畢，並著手收編實驗資料以進行效能評估。我們暫定於計畫第一年中後期將初步研究成果參與 SIGKDD、ICIS與 PACIS 等重要國際會議徵稿。在實驗進行的同時(計畫第一年中後期)，我們會開始規劃混合式文字特徵擷取研究，並期望能在計畫第二年中期將初步研究成果參與重要國際會議投稿。我們也期望能在研究過程中挖掘出新的研究議題(如利用排序學習挖掘同義文字)，並於計畫中後期實做系統且規劃這些新議題的研究方向。我們預期會產出4~7篇與本計畫相關的國際會議論文與4~6篇期刊論文。除論文發表外，我們也會在不侵犯智財權的情況下將收編的資料集公開給任何有興趣的學者使用，藉此吸引其他學者共同研究與討論這議題，進而增加本研究的能見度。

實作方面，我們會將所設計的主題趨勢預測系統應用在不同領域，並建立一個主題趨勢預測網站且定期公告國內外熱門趨勢主題狀態。此外，我們也會嘗試與行銷產業接觸以尋求產學合作機會，藉本系統協助行銷單位了解消費者搜尋行為與關注焦點(如與產品銷售相關的重要關鍵字)以改善行銷成果。我們也會積極申請相關專利發明，並藉此尋求技轉的可能性。

**參考文獻**

[1] D. Bahdanau, K. Cho, Y. Bengio, Neural machine translation by jointly learning to align and translate, arXiv preprint arXiv:1409.0473, (2014).

[2] H.-T. Cheng, L. Koc, J. Harmsen, T. Shaked, T. Chandra, H. Aradhye, G. Anderson, G. Corrado, W. Chai, M. Ispir, R. Anil, Z. Haque, L. Hong, V. Jain, X. Liu, H. Shah, Wide & Deep Learning for Recommender Systems, in: Proceedings of the 1st Workshop on Deep Learning for Recommender Systems, (ACM, Boston, MA, USA, 2016), pp. 7-10.

[3] P. Covington, J. Adams, E. Sargin, Deep Neural Networks for YouTube Recommendations, in: Proceedings of the 10th ACM Conference on Recommender Systems, (ACM, Boston, Massachusetts, USA, 2016), pp. 191-198.

[4] R. Dewan, B. Jing, A. Seidmann, Adoption of Internet-Based Product Customization and Pricing Strategies, Journal of Management Information Systems, 17(2) (2000) 9-28.

[5] M. Fu, H. Qu, Z. Yi, L. Lu, Y. Liu, A Novel Deep Learning-Based Collaborative Filtering Model for Recommendation System, IEEE Transactions on Cybernetics, 49(3) (2019) 1084-1096.

[6] J.J. Garrett, Customer Loyalty and the Elements of User Experience, Design Management Review, 17(1) (2006) 35-39.

[7] O. Georgiou, N. Tsapatsoulis, The Importance of Similarity Metrics for Representative Users Identification in Recommender Systems, in: IFIP International Conference on Artificial Intelligence Applications and Innovations, (Springer Berlin Heidelberg, Berlin, Heidelberg, 2010), pp. 12-21.

[8] C.A. Gomez-Uribe, N. Hunt, The Netflix Recommender System: Algorithms, Business Value, and Innovation, ACM Trans. Manage. Inf. Syst., 6(4) (2015) 1-19.

[9] X. He, L. Liao, H. Zhang, L. Nie, X. Hu, T.-S. Chua, Neural Collaborative Filtering, in: Proceedings of the 26th International Conference on World Wide Web, (International World Wide Web Conferences Steering Committee, Perth, Australia, 2017), pp. 173-182.

[10] Y. Jiang, J. Shang, Y. Liu, Maximizing customer satisfaction through an online recommendation system: A novel associative classification model, Decision Support Systems, 48(3) (2010) 470-479.

[11] Y. Koren, Collaborative filtering with temporal dynamics, in: Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining, (ACM, Paris, France, 2009), pp. 447-456.

[12] Y. Koren, R. Bell, C. Volinsky, Matrix factorization techniques for recommender systems, Computer, 42(8) (2009) 30-37.

[13] N.N. Liu, X. Meng, C. Liu, Q. Yang, Wisdom of the better few: cold start recommendation via representative based rating elicitation, in: Proceedings of the fifth ACM conference on Recommender systems, (ACM, Chicago, Illinois, USA, 2011), pp. 37-44.

[14] N.R. Mabroukeh, C.I. Ezeife, A taxonomy of sequential pattern mining algorithms, ACM Comput. Surv., 43(1) (2010) 1-41.

[15] A. Majumdar, A. Jain, Cold-start, warm-start and everything in between: An autoencoder based approach to recommendation, in: 2017 International Joint Conference on Neural Networks (IJCNN), (2017), pp. 3656-3663.

[16] A.M. Rashid, G. Karypis, J. Riedl, Learning preferences of new users in recommender systems: an information theoretic approach, SIGKDD Explor. Newsl., 10(2) (2008) 90-100.

[17] P. Resnick, N. Iacovou, M. Suchak, P. Bergstrom, J. Riedl, GroupLens: an open architecture for collaborative filtering of netnews, in: Proceedings of the 1994 ACM conference on Computer supported cooperative work, (ACM, Chapel Hill, North Carolina, USA, 1994), pp. 175-186.

[18] A. Rodriguez, A. Laio, Clustering by fast search and find of density peaks, Science, 344(6191) (2014) 1492-1496.

[19] J.B. Schafer, J. Konstan, J. Riedl, Recommender systems in e-commerce, in: Proceedings of the 1st ACM conference on Electronic commerce, (ACM, 1999), pp. 158-166.

[20] L. Shi, W.X. Zhao, Y.-D. Shen, Local Representative-Based Matrix Factorization for Cold-Start Recommendation, ACM Trans. Inf. Syst., 36(2) (2017) 1-28.

[21] S. Shi, M. Zhang, Y. Liu, S. Ma, Attention-based Adaptive Model to Unify Warm and Cold Starts Recommendation, in: Proceedings of the 27th ACM International Conference on Information and Knowledge Management, (ACM, Torino, Italy, 2018), pp. 127-136.

[22] P. Vincent, H. Larochelle, I. Lajoie, Y. Bengio, P.-A. Manzagol, Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion, Journal of machine learning research, 11(Dec) (2010) 3371-3408.

[23] M. Volkovs, G. Yu, T. Poutanen, Dropoutnet: Addressing cold start in recommender systems, in: Advances in Neural Information Processing Systems, (2017), pp. 4957-4966.

[24] J. Wei, J. He, K. Chen, Y. Zhou, Z. Tang, Collaborative filtering and deep learning based recommendation system for cold start items, Expert Systems with Applications, 69(2017) 29-39.

[25] Y. Wu, C. DuBois, A.X. Zheng, M. Ester, Collaborative Denoising Auto-Encoders for Top-N Recommender Systems, in: Proceedings of the Ninth ACM International Conference on Web Search and Data Mining, (ACM, San Francisco, California, USA, 2016), pp. 153-162.

[26] S. Zhang, L. Yao, A. Sun, Y. Tay, Deep Learning Based Recommender System: A Survey and New Perspectives, ACM Comput. Surv., 52(1) (2019) 1-38.

1. <https://www.finder.com/global-netflix-library-totals> [↑](#footnote-ref-1)
2. 文獻上另有新進商品冷啟動問題(new item cold start)，該問題在研究如何將新上架商品推薦給使用者。在本計畫，我們先專注新進使用者冷啟動問題，後續章會提到如何改良我們所設計的方法來減緩新商品冷啟動問題。 [↑](#footnote-ref-2)