

國立政治大學資訊科學系
Department of Computer Science
National Chengchi University

碩士論文
Master's Thesis

基於使用者表示法轉換之跨領域偏好排序於
推薦系統

User Embedding Transformation on
Cross-domain Preference Ranking for
Recommender Systems

研究生：陳先灝
指導教授：蔡銘峰

中華民國一百一十年八月
August 2021

基於使用者表示法轉換之跨領域偏好排序於推薦系統

User Embedding Transformation on Cross-domain
Preference Ranking for Recommender Systems

研究生：陳先灝

Student : Hsien Hao Chen

指導教授：蔡銘峰

Advisor : Ming-Feng Tsai

國立政治大學

資訊科學系

碩士論文

A Thesis

submitted to Department of Computer Science

National Chengchi University

in partial fulfillment of the Requirements

for the degree of

Master

in

Computer Science

中華民國一百一十年八月

August 2021

致謝

兩年前，爲了追求更好的職涯，以及因著某種迷惘，我離開當時的工作，並選擇回到學校，加入蔡老師的實驗室。因爲和之前領域不同，一開始自然不知所措，還好蔡老師很快地安排了一開始的玉山專案，讓我學習到何謂自然語言處理，以及執行專案的各種眉角。

之後，蔡老師也大方地採納我不成熟的研究提議，讓我和佳好、寶鈞共同執行數個推薦系統相關專案，跌跌撞撞地摸索這個我並不熟悉的領域。蔡老師和王老師也總是不耐煩地回答我一些相當粗淺的問題，並讓我肆意地探究了一些重要但當時無法有效取得成果的問題。那些在實驗室待到凌晨，拚命寫著數學證明和程式碼，最後沉沉睡去的日子，雖然痛苦掙扎，也終將成爲我未來人生的重要養分。

回頭看看這兩年的學習，最感謝的人莫過於蔡老師和王老師。兩位老師不論在學術上，以及爲人處世的態度上，都是我重要的榜樣，是推薦系統中重要的抽樣。而老師們對於報告的簡報以及表達能力的要求，我一開始是抗拒的，但後來卻理解到那些細緻的準備都是有意義的。能完整而簡單地講述某個研究，才是理解了該研究的初級指標，且優秀的溝通能力也是成爲好工程師的第一步。

我不會說什麼要謝的人太多，所以就感謝天吧這種話。

謝謝佳好，作爲最初的專案合作夥伴，你的能力一直讓我非常放心，好像什麼東西直接丟給你，都能有很穩定的產出，有時也會有一些真的很厲害的結果；謝謝寶鈞，對於那些我自己都不太清楚研究方向的議題，感謝你幫我做了各種實驗，並給予了很多重要的建議，讓它能在最後有個還不錯的歸處；謝謝均捷跟韋勝，對於我不成熟的程式技巧給了許多建議；謝謝喻能，每次三更半夜都願意跟我討論研究相關的問題，好的研究夥伴是前行的重要基石。特別感謝家人，讓我能自由地探索人生方向之外，也從來不要求我需要負什麼額外的責任，感謝你們一直願意包容我的任性。

陳先灝

國立政治大學資訊科學系

August 2021

中文摘要

隨著電子商務、影像串流服務等線上服務平台的發展，各大服務供應商對於「精準掌握用戶喜好」等相關技術的需求也逐季提升。其中，推薦系統作為這類方法的核心技術，如何在多變的現實問題中，提出符合特定需求的解決方式，也成為近年來相關研究的主要方向。在本研究中，我們特別關心的是推薦系統中的冷啟動 (Cold Start) 問題。冷啟動問題發生的主要原因，是因為特定情況造成的資料稀缺，比如推薦系統中的新用戶/物品等等。由於其困難性和實際應用中的無可避免，一直是推薦系統研究中，的一個具有挑戰性的問題。其中，緩解此問題的一種有效方法，是利用相關領域的知識來彌補目標領域的數據缺失問題，即所謂跨領域推薦 (Cross-Domain Recommendation)。跨領域推薦的主要目的在於，在多個不同的領域中實行推薦演算法，從中描繪出用戶的個人偏好 (Personal Preference)，再利用這些資訊來補充目標領域缺少的數據，從而在某種程度上解決冷啟動問題。

在本文中，我們提出了一個基於用戶轉換的跨領域偏好排序方法 (CPR)，它讓用戶從源域 (Source Domain) 和目標域 (Target Domain) 的物品中同時擷取資訊，並據此進行表示法學習，將其轉化為自身偏好的表示向量。通過這樣的轉換形式，CPR 將除了能有效地利用源域的資訊之外，也能直接地以此更新目標域中用戶和物品的相關表示，從而有效地改善目標域的推薦成果。在數據實驗中，為了能有效證明 CPR 方法的能力，我們將 CPR 方法實驗在六個不同的工業級資料上，並在差異化的條件設定 (目標域全體、冷啟動用戶、共同用戶) 中進行測試，也以先進的跨領域和單領域推薦演算法做為比較基準，進行比較。最後發現，CPR 不僅成功提高目標域整體的推薦效能，針對特定的冷啟動用戶也達到相當好的成果。

關鍵字：推薦系統、機器學習、跨領域推薦、冷啟動問題

Abstract

With the development of online service platforms such as e-commerce and video streaming services, the major service providers' demand for related technologies such as "accurately extracting user preferences" has also increased quarter by quarter. Among them, the recommendation system is the core technology of this kind of method. Therefore, how to propose solutions that meet specific needs in changing real problems has also become the main direction of related research in recent years. In this research, we are particularly concerned about the "cross-domain problem" in the recommendation system. The main reason for the cross-domain problem is the scarcity of data caused by different domains and specific circumstances, such as recommending source/target items in the system, and so on. Due to its difficulty and inevitable practical application, it has always been challenging in recommender systems research.

In this thesis, we propose a cross-domain preference ranking method (CPR) based on user conversion, which allows users to simultaneously extract information from items in the source domain and the target domain, and based on this, perform representation learning and transform it into a representation vector of their preferences. Through this conversion form, CPR effectively uses the information in the source domain and directly updates the relevant representations of users and items in the target domain, thereby effectively improving the recommendation results of the target domain. In the experiments, to effectively prove the ability of the CPR method, we tested with the CPR method on six different industrial-level data and conducted it in a differentiated condition setting (all target domains, cold-start users, shared users). The test also uses advanced cross-domain and single-domain recommendation algorithms as a benchmark for comparison. Finally, the experimental results show that CPR successfully improved the overall recommendation performance of the target domain and achieved quite good results for specific cold-start users.

Keywords: Recommendation System, Recommender System, Machine Learning, Cross Domain Recommendation, Cold-start

目錄

致謝	1
中文摘要	2
Abstract	3
第一章 緒論	1
1.1 前言	1
1.2 研究目的	3
第二章 相關文獻探討	5
2.1 表示法學習	5
2.2 推薦系統	6
2.2.1 冷啟動推薦	7
2.2.2 跨領域推薦	7
2.3 深度學習	8
2.3.1 圖神經網路	8
第三章 研究方法	10
3.1 問題定義	10
3.1.1 符號說明	10
3.1.2 推薦系統	11
3.1.3 跨領域推薦	11
3.2 貝葉斯個人化推薦排序	12
3.3 訊息傳遞模型與鄰居聚合	12
3.4 UET-CPR	14
3.4.1 基於使用者向量轉換的跨領域鄰居聚合	14
3.4.2 模型與實現方法說明	15
3.5 UET-CPR 的推薦方法	18
3.6 UET-CPR 的優化	19
3.6.1 優化貝葉斯個人化推薦排序	19
3.6.2 UET-CPR 模型設計的合理性	21
第四章 實驗結果與討論	22
4.1 實驗設定	22
4.1.1 模型比較對象	22
4.1.2 資料集	23
4.1.3 實驗設定與方法說明	24
4.2 研究問題說明	25
4.3 實驗結果	25
4.3.1 目標域上的使用者	25
4.3.2 目標域上的冷啟動使用者	27

4.3.3	目標域和源域上的共同使用者	28
4.4	關於 UET-CPR 模型與其他模型在不同資料集上表現差異的討論	28
4.4.1	不同資料集對結果的影響	29
4.4.2	TV-VOD 資料集	29
4.4.3	CSJ-HK 資料集	29
4.4.4	MT-B 資料集	30
4.4.5	小結	30
4.5	個案探討 - 以 TV-VOD 資料集為例	30
第五章	結論	32
	參考文獻	34



圖目錄

圖 1.1	k-hop 鄰居聚合	4
圖 3.1	無向的二分圖	11
圖 3.2	訊息傳遞模型	13
圖 3.3	UET-CPR 的模型	15



表目錄

表 3.1	本論文中使用的符號列表	10
表 4.1	資料集統計數據	24
表 4.2	模型在目標域的整體表現	26
表 4.3	模型在目標域上冷啓動使用者的表現	27
表 4.4	模型在目標域和源域上共同使用者的表現	28



第一章

緒論

1.1 前言

隨著網際網路的發達，諸如電子商務、影音串流等平台愈加活躍，除了讓人們的生活更加便利之外，在近年 Covid-19 的疫情肆意傳播之際，也為我們被受限制的日常生活，提供了部分替代方案，並帶來前所未有的商業機會。然而，由於相關資料的數量成指數型成長，且隨著使用習慣的改變與依賴程度的提升，使用者對服務要求的細緻化與個人化程度也迅速提高：

例如在 YouTube、Spotify 等影音串流平台中，使用者便不能再輕易地被制式化的推薦清單所滿足，而傾向追求更個人化、更符合其當下情緒的推薦結果；或是像近年炙手可熱的交友軟體中，傳統依照物理距離、個人檔案關鍵字等粗淺的配對方法，也正正面臨著時代的汰換，朝向基於使用者更隱性、更心理層面的特質進行推薦。

個人化推薦系統 (Personalized Preference Recommendation System) 便是因此被設計出來解決此類問題。雖然名為推薦，其本質更接近一種資訊過濾系統，目標在利用使用者過往的行為，例如點擊記錄、商品偏好等等，來推斷其內心隱性的偏好和特質，從而預測其未來的各項購買行為與使用情況。在將上述概念一般化之後，無論是電子商務、數位廣告投放、影音串流，甚至是生產力工具、語音助理、醫療行為等等，都可將其部份問題化約為一個推薦問題，並利用時下各項個人化推薦技術進行解決。

目前主要的個人化推薦技術，大約可分為如下三類：

- (1) 協同過濾 (Collaborative Filtering)
- (2) 內容推薦 (Content-based Recommendation)
- (3) 混合式推薦 (Hybrid Recommendation)

協同過濾的核心，在於「尋找使用習慣/過往行為類似的使用者」。當推薦系統能以使用者個人偏好的異同，將使用者分為不同的群集，便能假設同群集內的使用者會在某種程度共享類似的喜好，推薦同群集內的使用者所共同喜愛的物品，勢必會較推薦不同群集使用者喜歡的商品，來得更為精準有效。和協同過濾相反，內容推薦則以物品自身的特質出發，比如從物品顯示在網站的圖像、文字資訊、Hash Tag 等處抽取特徵，並試圖推薦內容特徵類似的物品。例如購物網站常推送給使用者相同作者的書籍，或是具有相同標籤和關鍵字的商品等等，都可以視為內容推薦的一種應用。而混合式推薦則是將上述兩種方法共同使用，以期兼收兩家之長，達到更好的成果，是近年來許多研究關注的重要議題之一。

然而，雖說盡可能更多地利用諸如個人活動紀錄、和商品特徵等各項資訊，是目前個人化推薦系統的潮流，但在現實世界的問題之中，受限於硬體成本、運算時間，以及使用者的人數攀升和其喜好的時時變動，如何有效地從繁雜的數據之中，挖掘出有價值的資訊並利用，勢必是未來研究與工業應用的一大挑戰。此外，隨著人們對隱私權關注的提升，以及歐盟之一般資料保護規範 (GDPR) 的通過，如何在使用個人資料的情況之下，去識別化並防止未來遭有心人士利用，也是相當值得關注的一環。

長時間裏，矩陣分解 (Matrix Factorization) 一直是協同過濾方法中的標竿，利用向量空間中的內積模擬使用者隱性偏好 (Implicit Preference) 的作法雖然簡單，效果卻出奇的好。如 BPR [24] 便是此類方法集大成的重要著作，不但提出了以最大後驗估計，直接對排序進行優化的作法，更透過引進負採樣 (Negative Sampling) 的概念，除了能更快速地優化最終模型，還能提升模型的堅固性 (robustness)。

深度學習 (Deep Learning) 從 2015 年 [28] 在影像上達到 SoTA (State of The Art) 之後，迅速地被應用於解決不同領域的各種問題，並在近幾年得到許多理論研究支持，證明其作為函數近似器 (Function Approximator) 的優秀能力，在推薦系統上的應用也開始獲得關注。 [5] 以神經網路預測使用者和影片之間的關聯性，並用於 YouTube 平台的推薦系統，使得每位使用者均享有個人化並不時更新的影音清單。

圖神經網路 (Graph Neural Networks) 最近也是推薦系統方法中重要的一支。大部分的神經網路模型，關注的是資料之間的幾何或時序關係，但圖神經網路卻透過將問題化約為圖 (Graph)，對於節點之間的拓撲結構進行學習和操作，和推薦系統的問題定義有異曲同工之妙。如 NGCF [36], LightGCN [10] 便是此類方法的佼佼者。

1.2 研究目的

在現實世界的各項應用中，跨領域推薦是工業級推薦系統時常需要面對的問題之一。以影音串流平台為例，以電影串流為主要商業模式的公司，若想要跨足電視劇領域，從頭開始收集相關資料並建立推薦模型，會花費極大成本和時間，且前期推薦結果肯定無法讓使用者滿意。此時，若能夠有效活用電影串流資訊作為源域 (Source Domain)，用以改善目標域 (Target Domain) 即電視劇的推薦模型，則可以達到比僅操作電視劇相關數據更好的結果。此即為跨領域推薦在工業級應用之目的和重要性。

而在上述所提及的情境中，跨領域推薦的目的，主要在於利用使用者在源域的活動紀錄，獲取其隱性個人偏好 (implicit personal preference)；讓他們即使在目標域因資料稀少屬於冷啟動 (cold start) 的情況下，也能得到令人滿意的推薦成果。然而，受限於模型複雜度、運算效率、資料大小及研究偏好等因素，近年來跨領域推薦的研究主軸，主要針對源域與目標域的共同使用者 (shared user)，即利用源域與目標域之間資訊的互動，補充/改善在目標域上已經表現尚可的推薦模型。

本文並不算質疑這些研究的價值，然而，在實際上的應用中，利用跨領域推薦之演算法改善冷啟動使用者/物品不但更符合現實需求，也因冷啟動問題本身相對困難的因素，更有解決的必要和價值。

綜上所述，本研究的目標在於設計一跨領域推薦之演算法，以改善推薦模型針對在源域有一定活動紀錄，但在目標域屬於冷啟動之使用者的推薦表現。此外，傳統的跨領域推薦演算法多半屬於兩階段式，即分別訓練使用者/物品在源域/目標域的代表向量 (embedding)，再透過另一獨立的模型，訓練一個從源域映射 (mapping) 至目標域的函數。然而。由於此類方法無法同時考量源域與目標域的情況，並動態調整其代表向量，因此，雖然在特定的情況下有效，但對於附加資訊 (side information) 稀缺的冷啟動問題的改善有限。有鑑於此，端到端 (end to end) 的演算法才是相形之下更合理的選擇。

本文提出了一個基於鄰居聚合 (neighbor aggregation) 的個人化偏好排序 (personalize preference ranking) 的端到端演算法，透過同時聚合使用者在源域與目標域的鄰居資訊，使模型同時考量源域與目標域的情況進行推薦，並修改牽涉到之節點的代表向量，再利用協同過濾 (Collaborative Filtering) 的性質，將這些訊息幅射至冷啟動的使用者，以補充其在目標域相對稀缺的資訊。

此外，在比較基準的部分，為了能更有效地比較各項跨領域推薦演算法的表現，我們將分成三種不同情境進行比較，並基於三組工業級資料集來觀察模型在不同情境下的結果差異，此外，也引進目前主流的單域 (single domain) 推薦模

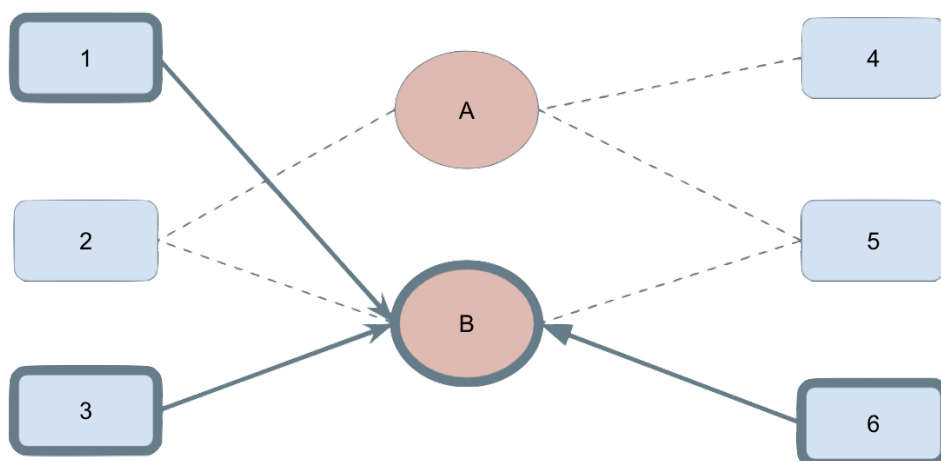


圖 1.1: k-hop 鄰居聚合

型，使本文提出之跨領域推薦演算法，能在方法設計、模型表現和訓練/預測效率的檢測上更具說服力，並讓後續研究者做為參考。

本研究的主要貢獻在於：

- 在推薦問題中，冷啟動問題由於其資訊的稀缺，一直是相對困難的部分，但在實際應用的情境卻又時常出現。而本文提出之跨領域推薦演算法，能夠透過使在其他領域之活動紀錄，補充其缺乏的附加資訊，能在一定程度上解決冷啟動問題的發生。
- 提出了一個端到端的向量表示學習法，能夠同時考量源域與目標域之情況，並動態調整雙方面之向量表示，不但在使用上更具彈性，且能夠有更好的推薦表現。
- 除了比較並分析過往之跨領域推薦演算法在不同資料集上的表現和差異，我們也引進了目前最先進的單領域推薦演算法作為基準，澈底比較各項跨領域推薦演算法在方法設計、模型表現和訓練/預測效率的優劣，可供後續研究者參考。
- 利用基於抽樣和異步梯度下降的實現，不但能有效平行化模型優化的過程，也能大幅提高模型的可擴展性，使其在工業級的資料上也能有效實現跨領域推薦。

第二章

相關文獻探討

2.1 表示法學習

機器學習任務的表現很大程度取決於其數據或特徵的表示方式，且該表示法最好基於其應用場域選擇。有鑑於此，目前許多機器學習演算法的實際應用，均著重於資料前處理的流程，或是數據轉換方式的設計，以產生「足以支持有效機器學習過程」的數據表示。雖然這類的特徵工程也是處理現實問題中重要的一環，但此種方法不但需要大量人力資源，也突出了傳統機器學習演算法的核心缺陷：無法從數據中提取和組織具判別性的資訊。特徵工程是一種試圖以人類的先驗知識以彌補這類弱點的方式，也在許多問題上展現出了有效的成果；然而，為了機器學習的適用性、一般性及其應用範圍，研究者更希望機器學習演算法的能降低其對人為特徵工程的依賴性，使得我們能更快速便捷地開發新的應用，並使我們的機器學習能夠更好地理解這個世界以及其背後的解釋性因素。

一般表示法學習的相關研究，其主要目的在於透過機器學習演算法，以得到具備有用資訊的數據表示，使得構建分類或回歸相關模型時，能藉由此數據表示，得到更好的下游任務表現。以機率模型為例，一個好的表示通常能從輸入的特徵中擷取其潛在解釋因素，並以之估計其後驗分佈；此外，在執行監督式學習時，好的特徵表示也能夠直接作為其模型的輸入使用。

以詞向量為例，word2vec [20] [19] 透過將預測單詞之間的共現 (Co-Occurrence) 作為任務，訓練全連接神經網路 (Fully-Connected Neural Network) 將單詞向量從獨熱編碼 (One-Hot Encoding) 壓縮成具備單詞特徵的低維向量，在數年間的各項自然語言處理任務中，取得相當好的成果。而在圖表示法學習 (Graph Embedding) 的研究中，也有許多研究者受其啟發：DeepWalk [21] 將圖 (Graph) 中的各項節點 (Node) 視為單詞處理，以隨機遊走 (Random Walk) 的方式，將路過的節點構建為串列，模擬出文字中句子的樣態，再以 word2vec 中 skip-gram 的方式學習各節點

的代表向量；而 Node2Vec [8] 則在隨機遊走的形式上加以改良，得以更精確地學習各節點的代表向量。LINE [30] 則另闢蹊徑，除了活用 word2vec 發揚光大的負採樣 (Negative Sampling) 之外，更引進二階相似度 (Second-Order Proximity) 的概念，考量節點作為自我 (Ego) 及鄰居 (Neighbor) 時的性質差異，在推薦系統等問題上，取得優異的成果。

2.2 推薦系統

推薦系統過往被視為資訊檢索 (Information Retrieval) 的一支，利用查詢 → 文本 (Query → Document) 的架構，延伸為例如使用者資訊 → 推薦物品的情況；且同樣因為在現實應用中，搜尋的池 (Pool) 較大，而同樣採用取回和重排序的架構 (Retrieve and Rerank)，先以較粗粒度 (Coarse-grained) 而有效率的方法，從所有文本/物品中建立一較小的候選池，再以粒度較細 (fine-grained) 但效率相對低的方法，對其進行重新排序。

近年來的推薦系統演算法多半奠基於協同過濾 (Collaborative Filtering) [26] 的概念，透過分析使用者/物品之間喜好/受喜好相似性，來對使用者進行推薦。而 BPR [24] 則是其中重要的里程碑，除了延續之前利用矩陣分解 [27] 的概念，也在使用者/物品的表示法學習中，引進了負採樣的概念，有效地利用了推薦系統中隱性回饋 (Implicit Feedback) 的性質，此外，更利用隨機梯度下降法 (Stochastic Gradient Decent) [2] 來提高表示法學習的效率，有效建立了後續大部分推薦系統研究的整體架構。

最近的推薦系統研究在方法上，主要透過引進深度學習和其相關領域的概念，對使用者/物品的表示法學習的整體流程進行改進，如 NCF [11] 就是透過利用全連接神經網路 (Fully-Connected Network) 對矩陣分解的架構進行模擬，並以相對更複雜的模型，取得較原生矩陣分解更好的成果，而 NGCF [36] 則是在其上將二分圖 (Bipartite Graph) 結構鑲嵌進其神經網路結構中，但本質仍是協同過濾和 BPR 的變體。GraphSAGE [9], PinSAGE [38] 則引進圖神經網路的概念，將二分圖中的每個節點和其鄰居共同進行表示法學習，並以鄰居抽樣器 (Neighbor Sampler) 來提高整個模型在工業資料集上的可擴展性 (Scalability)。

而有趣的是，近一兩年的研究則開始對這些將原始流程複雜化的手法，進行了深刻的反思：BPR 的作者 Stephen Rendle 在其論文 [25] 中質疑了 NCF [11] 以神經網路模擬矩陣分解的必要性，並以理論 [1] 和實驗證明了原生的矩陣分解結構，在細膩的超參數 (Hyper Parameters) 調整過後，較 NCF [11] 的表現更好；而 NGCF [36] 的原團隊更是提出 LightGCN [10]，說明了在 NGCF 的基礎上，刪去許多不必要的模型結構 (如代表向量之前的全連接層，以及神經網路多半具有的非線性層)，將能取得更好的表現，並大幅提高訓練效率。

而在研究的多樣性上，補充使用者/物品的元數據 (Metadata) 則是近年重要的一支，如 KGAT [35] 透過引進知識圖譜 (Knowledge Graph) 結構，讓物品在圖結構的高階資訊 (Higher-Order Information) 能夠被應用至表示法學習的過程之中，則是將各項多模態 (Multi-Modal) 資訊嵌入進使用者/物品的表示法學習過程之中，讓其產生的代表向量 (User/Item Embedding) 能夠包含這些外部資訊，以達到更好的推薦成果。

2.2.1 冷啟動推薦

冷啟動推薦 (Cold-Start Recommendation) 問題跟傳統推薦系統問題最大的差異在於，由於一般推薦演算法均使用使用者 \rightarrow 物品的交互關係為基礎進行推薦，因此這類資料的完整性和品質，直接會影響結果的好壞；而冷啟動推薦便是要試圖在此種資訊不足的情況下來討論。以電影推薦為例，對新的用戶進行推薦，或是將新的電影加入推薦清單，都屬於一種冷啟動問題。

資料補充是一種直接而有效的方法，比如電影串流公司 Netflix 針對新註冊的用戶，都會希望他們先選出一些自己喜愛的電影，再透過先行推薦一些熱門的影劇作品，來迅速建立起使用者的喜好輪廓。雖然上述屬於研究以外的做法，但亦有從這種方式借鑑的研究。如 DropoutNet [34] 便是將使用者和物品的內容特徵 (Content Feature) 以神經網路進行壓縮和轉換，再對某些使用者/物品使用 Dropout，將它們的資訊隱藏並進行推薦的下游任務，讓模型具備從有限資訊中，自行補足並取得好的推薦結果的能力，和 Netflix 使用方法的背後邏輯有異曲同工之妙。

基於內容的推薦也可以用以解決冷啟動問題。畢竟使用者的喜好和自身及物品的各項特徵有關，也可以利用雙方的內容資訊以達到對缺乏的交互關係的補充。然而，畢竟內容特徵在推薦系統中，屬於相對間接的資料，不如交互關係那麼直接，因此研究者多半會將內容資訊和有限的使用者 \rightarrow 物品交互關係混合運用，如 [33]。

2.2.2 跨領域推薦

跨領域推薦和一般推薦問題最大的差別在於，一般推薦問題中不太區分所謂「資訊來源」，通常將所有資訊直接混合進行模型訓練和推薦。而對跨領域推薦而言，通常會將資訊依其來源分為資訊相對豐富的源域，以及需要進行資訊補充，並主要觀察其上表現的目標域。因此，大部分的跨領域推薦相關研究，便著重如何有效擷取源域資訊，並將其目標域上進行利用。

CMF [29] 中提出的集體矩陣轉換 (Collective Matrix Factorization) 是跨領域推薦中重要的方法之一。它透過訓練一線性轉換來將源域和目標域中共同的使用者/物品表示法進行轉換，從而將資訊從源域轉換至目標域，以達成跨領域推薦的目的。而 EMCDF [18] 則將其 CMF 提出的線性轉換以多層感知器 (MLP) 進行替代，以達到更好的效果。

Bi-TGCF [17] 則以圖神經網路協同過濾 (NGCF) 為基礎，建立有效的圖特徵擷取模組，並結合交替式的學習策略，輪流訓練源域和目標域的代表法，從而讓模型掌握來自源域和目標域的不同資訊。

另一種進行跨領域推薦的做法，則是透過引進額外的資訊，如圖像、文字的元數據 (metadata)，從而獲取目標域中缺乏活動紀錄的使用者/物品的資訊。這類方法主要透過模型設計，取得基於使用者/物品元數據的特徵之後，以源域較豐富的活動紀錄為基礎進行表示法學習，並將其結果應用的目標域，從而改善其資訊稀少的問題。

2.3 深度學習

深度學習 (Deep Learning) 一詞出自 [16]，乃是該文三位作者整合了近年間神經網路的研究，並基於其隨著圖形處理器 (Graphics Processing Unit, GPU) 的發展，而能更有效率地建立起較以往更深層的神經網路 [22] 所命名，其理論基礎在於，基於通用近似定理 (Universal Approximation Theorem)，任意寬度的兩層以上神經網路，可以擬合任意連續函數 (離散函數則可以在該證明過程中被逼近) [13]。爾後，卷積神經網路 (Convolutional Neural Network, CNN) [28] [23] 在圖像領域大放異彩，輕易地擊敗了原先的 SoTA [3]，遞歸神經網路 (Recurrent Neural Network, RNN) [12]、注意力模型 (Attentional Model) [31] 則是近年自然語言處理的主流，Transformer [31]、BERT [6] 的出現更是一口氣影響了上述兩個領域的格局，將預訓練模型 (Pre-Trained Model) 的應用推到了一個新的高度。

2.3.1 圖神經網路

圖神經網路 (Graph Neural Networks) 的基礎，在於活用了訊息傳遞 (Message Passing) [7] 的架構，讓模型在學習各個節點的特徵時，能夠同時考慮到該節點鄰居的相關資訊，並將節點之間的遠近關係，也透過神經網路的分層 (Hierarchical) 結構有效地區分，並嵌入在各節點的特徵裏。而圖卷積網路 (Graph Convolutional Network, GCN) [14] 則是在訊息傳遞模型的基礎上，將圖上各節點的關係，視為一種能量傳遞的過程，並試圖將卷積神經網路 (CNN) 的結構，以所謂「圖傅立葉轉換」 (Graph Fourier Transform) 的形式，應用在這些節點的代表法學習上。

雖然 GCN 建構了圖神經網路的理論基礎，但其繁複的數學形式與在大資料集上可擴展性不高的模型，也讓後續的研究者在其上做了諸多改進，GraphSAGE [9] 在 GCN 原本的平均聚合架構之上，另外提出了連接聚合、LSTM [12] 聚合等等，並提出了鄰居抽樣 (Neighbor Sampling) 的概念，大幅提高了圖神經網路模型的效率和擴展性。另一方面，[37] 則是將圖神經網路的架構化約為鄰居聚合 (Neighbor Aggregation) 和圖結合 (Graph Combination) 兩大模塊，為後續的研究提供了簡明的數學表示；此外，他們也提出一連串證明，說明圖神經網路表達能力的上限最多和 WL-Test (Weisfeiler-Lehman Test) 相當，也說明如果要達到該上限，圖神經網路在鄰居聚合和圖結合階段，其選用的運算子 (Operator)，均需要滿足單射函數 (Injective function) 的性質。這樣有力、簡潔的結果和其證明過程，為之後圖神經網路相關的理論研究，提供了很好的基礎；而其提出的圖同構網路 (Graph Isomorphism Network) 雖然不能說不優秀，但在圖分類 (Graph Classification) 以外的領域，反而沒有受到太多的關注。



第三章

研究方法

3.1 問題定義

3.1.1 符號說明

這裏列出我們後續在本文中使用的數學符號。

符號	定義
G	定義使用者和物品之間關係的無向圖
V	定義使用者和物品之間關係的無向圖中的點集
E	定義使用者和物品之間關係的無向圖中的邊集
U	使用者的集合
I	物品的集合
E	使用者和物品代表向量的集合
S	源域
T	目標域
U^S	源域上的使用者
U^T	目標域上的使用者
I^S	源域上的物品
I^T	目標域上的物品
e_u	使用者 u 的代表向量
e_i	物品 i 的代表向量
e_u^{agg}	使用者 u 的經過鄰居聚合後的代表向量
$\langle e_u, e_i \rangle$	使用者 u 和物品 i 代表向量的內積

表 3.1: 本論文中使用的符號列表

3.1.2 推薦系統

考慮一個具有使用者集 U 和物品集 I 的情況，我們可將其中使用者和物品的互動關係，以一張無向的二分圖 $G = (V, E)$ 進行表示：

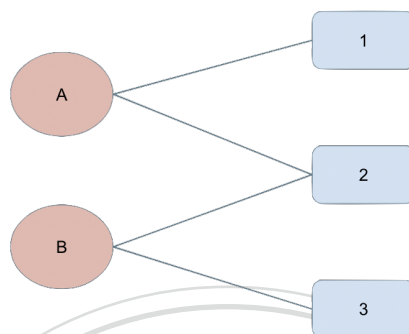


圖 3.1: 無向的二分圖

其中，

$$V = U \cup I, \forall (u, i) \in E, u \in U, i \in I$$

而推薦演算法的目的，在於學習一個特殊的向量集 $E \in \mathbf{R}^{(|U|+|I|) \times d}$ ，使得對於任意 $v \in V$ ，均存在一個向量表示 $e_v \in \mathbf{R}^d$ ，使其推薦任務的表現最大化。

3.1.3 跨領域推薦

在本文中，我們將考慮在兩個不同領域下實現推薦系統的情況：源域 S 是我們具備較多資訊的領域，擁有充分的使用者 \leftrightarrow 物品連結；比如作為一間音樂串流公司為例，音樂推薦便可以視為該公司的源域，是公司相對成熟的市場/場域。而目標域 T 則是資訊相對稀缺的領域，擁有部分的使用者 \leftrightarrow 物品連結，但不足以據此構建出一套相對有效的推薦系統；比如對上述該音樂串流公司而言，商品零售可能就是他們發展中的目標域。其中，我們將源域和目標域的物品集以 I^S 和 I^T 表示，而它們的使用者集則以 U^S, U^T 分別代表。這類的跨領域推薦方法，多半建立在不同的假設上，並據此開發出特性不同的演算法。而本文關注的重點在於，能不能在源域和目標域的使用者存在一定交集的情況下，從源域現有的使用者 \leftrightarrow 物品連結中，擷取出隱藏的特徵/資訊，再將其應用到目標域的推薦上。即採用 $U^S \cap U^T \neq \emptyset$ 的假設。

從這樣的假設出發，我們可以得到源域和目標域聯集的使用者集 $U = U^S \cup U^T$ 和物品集 $I = I^S \cup I^T$ ，以及兩域之間共通的使用者 $U^{\text{share}} = U^S \cap U^T$ 和目標域上

的冷啟動使用者 $U^{\text{cold}} = U^S \setminus U^T$ ；而本文的最終目的便是透過表示法學習，得到一組夠好的向量表示 $E = \mathbf{R}^{(\|U\|+\|I\| \times d)}$ ，以達到下列目的：

- 目標域上最好的推薦任務表現。
- 針對冷啟動(cold-start)，即源域上未出現的使用者，也能起到資訊補充，提升推薦任務表現的作用。

3.2 貝葉斯個人化推薦排序

在個人化推薦排序的方法中，貝葉斯個人化推薦排序(Bayesian Personalized Ranking, BPR) [24]，無疑是其中最經典有效的做法之一。其核心概念在於，在給定任一使用者 u 的情況，隨機抽取其正負面的物品樣本 i, j (此處的正負面意指與使用者 u 的連結與否)，透過最大化目標函數 $\langle e_u, e_i \rangle - \langle e_u, e_j \rangle$ ，使得使用者 u 的代表向量 e_u 與正面物品樣本 i 的代表向量 e_i 互相靠近，而其與負面物品樣本 j 的代表向量 e_j 互相拉遠。BPR 方法的核心價值在於，透過引進負面樣本 j 進行負採樣 (negative sampling)，讓推薦模型除了能觀察使用者的喜好的物品本身，也能得到使用者「相對不喜歡」的物品相關資訊。透過引進這樣隱性的排序資訊，模型將能更有效地運用使用者 \leftrightarrow 物品連結中的隱性回饋 (implicit feedback)，以達到更好的推薦成果。

3.3 訊息傳遞模型與鄰居聚合

訊息傳遞的概念由 [7] 提出，透過遞迴地將任意節點 n 與其鄰居節點之特徵向量進行聚合，以得到該節點的聚合向量表示；而當此過程實行 k 次之後，該聚合向量便會包含原節點 n 及 n 的 k -hop 鄰居的特徵性質及其圖結構，從而達到更好的下游任務效果。

其中，聚合方法可分為平均聚合 (Average Aggregation) [14]，連接聚合 [9]、或是以神經網路模型 (LSTM, Attention, ...) 進行聚合 [9], [32] 等等方法，可視特徵形式及下游任務等進行修改。

訊息傳遞模型在推薦演算法的意義在於，透過讓相鄰節點之間的向量進行資訊的結合，除了能夠讓相鄰節點的代表向量，在高維度的向量空間中能夠更靠近之外；在進行如貝葉斯個人化排序 (BPR) 時，和一般單純抽樣並優化一對節點的做法相比，經過訊息傳遞聚合的模型由於在優化過程中，每個節點的代表向量均與相鄰節點有關，將一次優化一群節點，更能夠有效達到協同過濾 (Collaborative

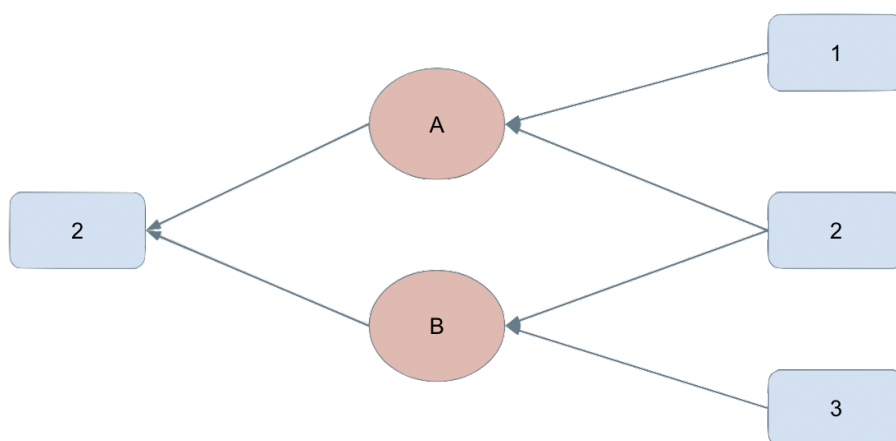


圖 3.2: 訊息傳遞模型

Filtering) 的效果；且由於訊息傳遞的階層化 (hierarchical) 結構，能夠讓模型針對不同距離的節點，給予大小不等的權重，從而將所有節點形成之圖結構 (Graph Structure) 納入考慮，形成一更細緻的模型。

此外，對於採用貝葉斯個人化排序 (BPR) 等以內積 (Dot-Product) 做為目標函數 (Objective Function) 的推薦演算法而言，訊息傳遞/鄰居聚合的過程本身便能一定程度提升其推薦任務表現。我們用定理 3.3.1 對其進行說明：

Theorem 3.3.1. 給定一使用者 u 及其 n 個相鄰物品 $\{i_1, i_2, \dots, i_n\}$ ，而其表示向量為 e_u 和 $\{e_{i_1}, e_{i_2}, \dots, e_{i_n}\}$ ，且 $e_u, e_{i_1}, e_{i_2}, \dots, e_{i_n} \in \mathbf{R}^d$ 。在其向量長度均固定的情況（比如固定為 1），若對使用者之表示向量 e_u 進行平均聚合，則其與相鄰物品表示向量之內積總和，會大於等於聚合之前，即：

$$\sum_{k=1}^n \langle e_u^{agg}, e_{i_k} \rangle \geq \sum_{k=1}^n \langle e_u, e_{i_k} \rangle$$

當

$$e_u^{agg} = \frac{1}{n+1} (e_u + \sum_{k=1}^n e_{i_k})$$

Proof. 考慮 $\sum_{k=1}^n \langle e, e_{i_k} \rangle, e \in \mathbf{R}^d$ ，且 $\|e\| = 1$ 的極大值。

$$\sum_{k=1}^n \langle e, e_{i_k} \rangle = \langle e, \sum_{k=1}^n e_{i_k} \rangle$$

由於 e 長度固定且 $\|e_{i_1}\| = \|e_{i_2}\| = \dots = \|e_{i_n}\| = 1$

$$\implies e = \frac{1}{n} \sum_{k=1}^n e_{i_k} \quad (3.1)$$

$$\implies \sum_{k=1}^n \left\langle \frac{1}{n} \sum_{k=1}^n e_{i_k}, e_{i_k} \right\rangle \geq \sum_{k=1}^n \langle e, e_{i_k} \rangle \quad \forall e \in \mathbf{R}^d \quad (3.2)$$

給定 e_u 和 e_{i_k} ，則 e_u 之平均聚合結果 e_u^{agg} 會滿足

$$\begin{aligned} \sum_{k=1}^n \langle e_u^{\text{agg}}, e_{i_k} \rangle &= \sum_{k=1}^n \left\langle \frac{1}{n+1} \left(e_u + \sum_{k=1}^n e_{i_k} \right), e_{i_k} \right\rangle \\ &= \frac{1}{n+1} \sum_{k=1}^n \langle e_u, e_{i_k} \rangle + \sum_{k=1}^n \left\langle \frac{1}{n+1} \left(\sum_{k=1}^n e_{i_k} \right), e_{i_k} \right\rangle \\ &\geq \frac{1}{n+1} \sum_{k=1}^n \langle e_u, e_{i_k} \rangle + \sum_{k=1}^n \left\langle \frac{n}{n+1} e_u, e_{i_k} \right\rangle \quad (\text{by 3.2}) \\ &= \sum_{k=1}^n \langle e_u, e_{i_k} \rangle \end{aligned} \quad (3.3)$$

故當 $k \in 1, 2, \dots, n$

$$\langle e_u^{\text{agg}}, e_{i_k} \rangle \geq \langle e_u, e_{i_k} \rangle \quad (3.4)$$

□

雖然上述定理的假設，是建立在向量長度固定的情況，但由於向量的標準化 (Normalization) 在機器學習模型中屬於時常使用的技巧，故進行這樣的假設應也在合理的範圍之內。

3.4 UET-CPR

本文中所提出的 UET-CPR 方法，將利用跨領域推薦的性質，同時操作來自源域和目標域的特徵，以得到能同時考量不同領域資訊的使用者表示向量；並結合鄰居聚合的方法，有效取得不同領域間圖架構的資訊；最後利用 BPR 作為目標函數，將使用者和其喜愛的物品在向量空間上拉近，並拉遠其和相對不喜愛物品之距離。

3.4.1 基於使用者向量轉換的跨領域鄰居聚合

在前一個小節中，我們說明了訊息傳遞模型和鄰居聚合的技術，主要目標在於讓各節點能夠接收其鄰居的資訊，並透過訊息傳遞模型階層化的設計，讓圖結構相關的特徵能夠被考慮。此外，針對貝葉斯個人化排序等以內積為目標函數的推薦演算法，僅透過單純的平均聚合，便能有效提高其最終表現。

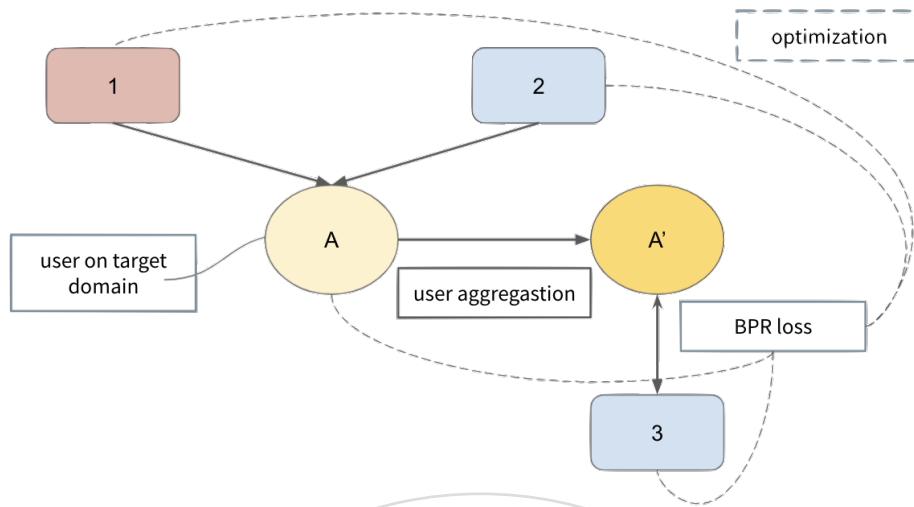


圖 3.3: UET-CPR 的模型

而在跨領域推薦中，我們對模型的主要需求在於，其能夠有效從源域和目標域擷取必要的資訊，並將其回饋給後續的目標域推薦。因此，我們提出了基於使用者向量轉換的跨領域鄰居聚合結構，來實現上述的需求：

$$e_u^{\text{agg}} = e_u + \frac{1}{\|N_u^S\|} \sum_{i^S \in N_u^S} e_{i^S} + \frac{1}{\|N_u^T\|} \sum_{i^T \in N_u^T} e_{i^T}, \forall u \in U \quad (3.5)$$

其中，

$$i \in N_u^S \quad \text{if} \quad i \in I^S, (u, i) \in \mathbf{E}$$

$$i \in N_u^T \quad \text{if} \quad i \in I^T, (u, i) \in \mathbf{E}$$

這樣設計的好處在於，透過均等地讓使用者的代表向量，針對源域和目標域共同進行平均聚合，將能夠讓來自源域和目標域的資訊進行結合，並互相影響，從而達到跨領域推薦的效果。此外，在後續的優化過程中，由於使用者向量的形成與源域和目標域的物品向量有關，在對其進行隨機梯度下降法 (Stochastic Gradient Decent, SGD) 優化的同時，也能夠同時優化其相鄰的源域和目標域的物品向量。

我們會在後續的小節中對此進行詳細討論。

3.4.2 模型與實現方法說明

UET-CPR 方法之數學表示式如下：

給定兩張二分圖 $G^S = (V^S, E^S), G^T = (V^T, E^T)$ 代表源域和目標域中，使用者和物品之間的關聯性，且滿足：

$$V^S = U^S \cup I^S, \forall (u, i) \in E^S, u \in U^S, i \in I^S$$

和

$$V^T = U^T \cup I^T, \forall (u, i) \in E^T, u \in U^T, i \in I^T$$

且

$$U^S \cap U^T \neq \emptyset$$

而對於其表示向量集

$$E = \mathbf{R}^{(\|U\| + \|I\| \times d)} (U = U^S \cup U^T, I = I^S \cup I^T)$$

則

$$\forall u \in U, i \in I, \exists e_u, e_i \in E$$

爾後，針對所有的使用者表示向量進行跨領域鄰居聚合，即：

$$e_u^{\text{agg}} = e_u + \frac{1}{\|N_u^S\|} \sum_{i^S \in N_u^S} e_{i^S} + \frac{1}{\|N_u^T\|} \sum_{i^T \in N_u^T} e_{i^T}, \forall u \in U \quad (3.6)$$

其中，

$$i \in N_u^S \quad \text{if} \quad i \in I^S, (u, i) \in E$$

$$i \in N_u^T \quad \text{if} \quad i \in I^T, (u, i) \in E$$

並以聚合後的使用者向量，和原本的物品向量，在目標域計算 BPR 之目標函數並最大化之：

$$\text{Maximize } O_{\text{BPR}} = \sum_{u \in U^T} \sum_{i \in N_u^T} \sum_{j \in I^T \setminus N_u^T} \ln \sigma(\langle e_u^{\text{agg}}, e_i \rangle - \langle e_u^{\text{agg}}, e_j \rangle) \quad (3.7)$$

基於上述方法及目標函數，在給定使用者 $u \in U^T$ 及其正/反面物品樣本 $i, j \in I^T$ 的情況下，由於 (3.9)，主要的優化目標有三：

1. 最大化 $\langle e_u, (e_i - e_j) \rangle$

2. 最大化 $\langle e_{i_u^S}, (e_i - e_j) \rangle, i_u^S \in N_u^S$

3. 最大化 $\langle e_{i_u^T}, (e_i - e_j) \rangle, i_u^T \in N_u^T$

可以視為同時對 $e_u, e_{i_u^S}, e_{i_u^T}$ 同時進行貝葉斯個人化推薦排序的優化。

而在模型實際訓練的過程中，考慮到計算效率和可擴展性的因素，在訓練過程中，我們利用類似於 Cluster-GCN [4] 中的設定，以抽樣方式來處理 e_{i^S} 和 e_{i^T} 的平均值表示，即在每次訓練週期中，針對每個使用者，均各從源域和目標域中抽樣出一個物品 i_S, i_T ，並以此兩物品的代表向量 e_{i^S} 和 e_{i^T} 作為上述平均值，即以抽樣方式表達平均的概念：

$$e_u^{agg} = e_u + \frac{1}{2}(e_{i^S} + e_{i^T}), \forall u \in U \quad (3.8)$$

但在實際進行推薦時，我們 e_u^{agg} 的計算則依然使用 (3.9)。

值得一提的是，由於我們主要對目標域的物品偏好進行建模，因此，本研究在實現上，進行了三個重要的設定調整：

- 在優化過程中，只針對「目標域」的使用者進行抽樣。
- 對於只和目標域中物品有交互關係的使用者 u_t ，即源域上未出現的使用者，以其自身的代表向量 e_{u_t} 和其在目標域上的相鄰物品向量 $\{e_{i^T} | i^T \in N_u^T\}$ 進行鄰居聚合，即 $e_{u_t}^{agg} = e_{u_t} + \frac{1}{\|N_u^T\|} \sum_{i^T \in N_u^T} e_{i^T}$ 。
- 針對冷啟動 (cold-start) 使用者，即目標域上未出現的使用者 u_c ，以其在源域上的相鄰物品向量之平均值 $\frac{1}{\|N_u^S\|} \sum_{i^S \in N_u^S} e_{i^S}$ 作為其鄰居聚合後的代表向量 $e_{u_c}^{agg}$ 。

我們將針對上述設定的合理性進行討論。

首先是僅對目標域進行抽樣的部分。這樣做的原因在於，由於我們主要關心的是模型在目標域的表現，因此優化的對象以目標域的使用者和物品，以及和目標域使用者在源域的相鄰物品為主，可以得到比較好的結果。而僅出現在源域上的使用者和物品在這個情況下來說，額外對它們的代表向量做優化顯得缺乏意義，也可能會讓它們成為類似雜訊的角色，這點也可以在後續的實驗結果中，發現類似的情況。

而對於只和目標域中物品有交互關係的使用者 u' ，由於 $N_{u'}^S = \emptyset$ ，我們無法取得其在源域上的資訊，故簡單地以目標域上的相鄰物品為主，應該是相對合理的做法。

Algorithm 1 UET-CPR

```
1: Result:  $E \in \mathbf{R}^{(\|U\|+\|I\|)\times d}$ 
2: Randomly assign a  $d$ -dimension vector  $e_u, e_i$  to each  $u \in U, i \in I$ ;
3: Define the number of negative samples as  $n$ ;
4: for  $u \in U^T$  do
5:   sample  $i^S \in N_u^S$ ;
6:   sample  $i^T \in N_u^T$ ;
7:    $e_u^{agg} = e_u + \frac{1}{2}(e_{i^S} + e_{i^T})$ 
8:   sample  $i_u^T \in N_u^T$ ;
9:   for  $k \leftarrow 1$  to  $n$  do
10:    sample  $j_k^T \in I^T \setminus N_u^T$ ;
11:    Maximize  $\langle e_u^{agg}, e_{i_u^T} \rangle - \langle e_u^{agg}, e_{j_k^T} \rangle$ 
12:   end for
13: end for
```

另一方面，對於冷啟動使用者 (Cold Start Users)，即 $u \in U^S \setminus U^T$ 且因此 $e_u = 0, N_i^T = \emptyset$ ；我們在推理階段，會直接讓 $e_u^{agg} = \frac{1}{\|N_u^S\|} \sum_{i^T \in N_u^S} e_{i^S}$ ，用於和物品表示向量計算內積，以獲得推薦結果。這樣做的目的在於，雖然我們無法取得該使用者在源域上的資訊，但由於與其相鄰的物品可能在目標域使用者的優化過程中，一併被優化到，故利用這些物品來代表該冷啟動使用者的代表向量，將能夠建立起它和目標域的連結，也符合鄰居聚合的概念。

上述演算法中，和前面的模型敘述稍有不同的地方在於，在使用者的跨領域鄰居聚合部分，我們利用兩者的平均值來控制最終的向量長度，避免其毫無限制地增長；而所謂的「負採樣數」 n 的設計，也是近期推薦相關演算法常用的作法之一，可以透過調整 n 的大小，來調整並提升模型在不同資料集上的表現，屬於合理的超參數設定範圍。

3.5 UET-CPR 的推薦方法

在經過以 (3.7) 為目標函式的優化之後，我們可以得到一組針對源域和目標域所有使用者/物品的代表向量 $E \in \mathbf{R}^{(\|U\| + \|I\|) \times d}$ ，並以此為基礎進行推薦。

我們首先對目標域使用者的代表向量進行跨領域鄰居聚合。

$$e_u^{agg} = e_u + \frac{1}{\|N_u^S\|} \sum_{i^S \in N_u^S} e_{i^S} + \frac{1}{\|N_u^T\|} \sum_{i^T \in N_u^T} e_{i^T}, \forall u \in U \quad (3.9)$$

其中，

$$i \in N_u^S \quad \text{if} \quad i \in I^S, (u, i) \in \mathbf{E}$$

$$i \in N_u^T \quad \text{if} \quad i \in I^T, (u, i) \in \mathbf{E}$$

(這裏和 (3.9) 是一致的)

之後，針對所有的「目標域」使用者和欲進行推薦的物品，以內積方式估計使用者 u 對物品 i 的偏好 $\hat{y}_{\text{UET-CPR}}(u, i)$:

$$\hat{y}_{\text{UET-CPR}}(u, i) = \langle e_u^{\text{agg}}, e_i \rangle$$

3.6 UET-CPR 的優化

為了說明 UET-CPR 方法設計的合理性，我們將實際展開其在貝葉斯個人化推薦排序的優化過程，並比較其和傳統的貝葉斯個人化推薦排序優化之間的差異。

3.6.1 優化貝葉斯個人化推薦排序

給定使用者 u ，及 u 的鄰居物品 i 和負採樣物品 j 。在一般推薦系統的設定中，我們的優化目標是要讓模型在針對使用者 u 進行推薦時，比起負採樣物品 j ，更傾向於推薦物品 i 。我們可以用 $p(i \geq_u j | \theta)$ 來代表模型在參數集為 θ 的情況下，模型比起 j ，更傾向於推薦 i 給 u 的機率，並試圖找到能將它最大化的參數集 θ 。

而在優化貝葉斯個人化推薦排序時，我們一般會用 $\hat{x}_{uij}(\theta)$ 代表某個任意的實數函數，用以表示其考量了使用者 u ，物品 i 和負採樣物品 j 之間的關係，並以此來估計後續的機率值(為了方便起見， $\hat{x}_{uij}(\theta)$ 之後將省略寫作 \hat{x}_{uij})。現在則一般採用內積的形式來構建函數 $\hat{x}_{uij}(\theta)$ ：

$$\begin{aligned} p(i \geq_u j | \theta) &= \sigma(\hat{x}_{uij}(\theta)) \\ &= \sigma(\langle e_u, e_i \rangle - \langle e_u, e_j \rangle) \\ &= \sigma(\langle e_u, e_i - e_j \rangle) \end{aligned} \tag{3.10}$$

其中 e_u, e_i, e_j 為模型學習到的 u, i, j 各自的代表向量， σ 代表的是 sigmoid 函數：

$$\sigma(t) = \frac{1}{1 + e^{-t}}$$

這裏使用 sigmoid 函數的理由在於，由於其值域正好落在 $(0, 1)$ ，可以直接以這個輸出值代表我們想要的「模型比起 j ，更傾向於推薦 i 給 u 的機率」。

因此，在利用梯度下降法 (gradient decent) 優化模型參數 θ 時，可以將其更新過程寫成：

$$\theta \leftarrow \theta + \alpha \left(\frac{e^{-\hat{x}_{uij}}}{1 + e^{-\hat{x}_{uij}}} \cdot \frac{\partial}{\partial \theta} \hat{x}_{uij} \right)$$

其中 α 是學習率，控制模型更新的步長，並可以隨不同的優化演算法進行動態調整。

將內積形式代入並展開的話會像這樣：

$$\theta + \alpha \left(\frac{e^{-\hat{x}_{uij}}}{1 + e^{-\hat{x}_{uij}}} \cdot \frac{\partial}{\partial \theta} \hat{x}_{uij} \right) = \theta + \alpha [\sigma(-\langle e_u, e_i - e_j \rangle) \cdot \frac{\partial}{\partial \theta} \hat{x}_{uij}] \quad (3.11)$$

其中 $\frac{\partial}{\partial \theta} \hat{x}_{uij}$ 的值對 u, i, j 會有不同的情況：

$$\frac{\partial}{\partial \theta} \hat{x}_{uij} = \begin{cases} (e_i - e_j) & \text{if } \theta = e_u \\ e_u, & \text{if } \theta = e_i \\ -e_u, & \text{if } \theta = e_j \\ 0 & \text{otherwise.} \end{cases} \quad (3.12)$$

從上述的 (3.11) 和 (3.12) 中，我們可以得到幾個關於採用貝葉斯個人化推薦排序的優化的重要結論：

- 使用者 u 的代表向量 e_u 在向量空間中，會靠近相鄰物品 i 的代表向量 e_i ，並遠離負採樣物品 j 的代表向量 e_j 。
- e_i 會靠近 e_u ；而 e_j 會遠離 e_u 。
- 這個靠近和遠離的步長，除了由和 u, i, j 無關的學習率 α 來決定之外，也會和 e_u, e_i, e_j 的內積值有關。如果 e_u 和 e_i 越靠近，或是 e_u 和 e_j 越遠，代表這次抽樣後需要的模型調整並不大，這個步長就會縮小；相反地，如果 e_u 和 e_i 越遠，或是 e_u 和 e_j 越靠近，代表這次抽樣後需要的模型調整就比較多，步長就會增大。

雖然這些結論在過去的研究中已經被完整討論，但由於這個過程對我們後續探討 UET-CPR 的設計合理性相當重要，因此基於邏輯自洽和論文完整性的緣故，依然在這裏重新陳述並簡單推導。

3.6.2 UET-CPR 模型設計的合理性

給定使用者 u ，及 u 的鄰居物品 i 和負採樣物品 j ，以及使用者 u 的源域相鄰物品 i_u^S 和目標域相鄰物品 i_u^T 。

我們在前一個小節中提及，根據 (3.10)，貝葉斯個人化推薦的目的，是要最大化：

$$\sigma(\langle e_u, e_i - e_j \rangle)$$

而基於 (3.9) 和 (3.8)，UET-CPR 在訓練中則是試圖最大化：

$$\sigma(\langle e_u + \frac{1}{2}(e_{i_u^S} + e_{i_u^T}), e_i - e_j \rangle)$$

通過我們在 3.6.1 的推導與簡單的微積分連鎖率，我們可以輕易得到 3.4.2 提及的優化目標，並在 UET-CPR 的設定下延伸 3.6.1 的結論：

- 使用者 u 以及使用者 u 的源域相鄰物品 i_u^S 和目標域相鄰物品 i_u^T 的代表向量 $e_u, e_{i_u^S}, e_{i_u^T}$ 在向量空間中，會靠近相鄰物品 i 的代表向量 e_i ，並遠離負採樣物品 j 的代表向量 e_j 。
- e_i 會靠近 $e_u, e_{i_u^S}, e_{i_u^T}$ ；而 e_j 會遠離 $e_u, e_{i_u^S}, e_{i_u^T}$ 。
- 這個靠近和遠離的步長，除了由和 u, i_u^S, i_u^T, i, j 無關的學習率 α 來決定之外，也會和 $e_u, e_{i_u^S}, e_{i_u^T}, e_i, e_j$ 的內積值有關。如果 $e_u, e_{i_u^S}, e_{i_u^T}$ 和 e_i 越靠近，或是 $e_u, e_{i_u^S}, e_{i_u^T}$ 和 e_j 越遠，代表這次抽樣後需要的模型調整並不大，這個步長就會縮小；相反地，如果 $e_u, e_{i_u^S}, e_{i_u^T}$ 和 e_i 越遠，或是 $e_u, e_{i_u^S}, e_{i_u^T}$ 和 e_j 越靠近，代表這次抽樣後需要的模型調整就比較多，步長就會增大。

基於上述結論，我們發現在 UET-CPR 跟傳統貝葉斯個人化推薦排序在跨領域推薦上的優化差異在於，從使用者 u 的角度而言，我們額外優化了它在源域和目標域相鄰物品 i_u^S, i_u^T 的代表向量；而就相鄰物品 i 和負採樣物品 j 的角度來說，我們額外將其與 u 的其他鄰居 i_u^S, i_u^T 的關係納入考量，並引入了源域的資訊作為補充。

第四章

實驗結果與討論

在本章節中，我們將每個資料集基於三種不同的評估目標，對我們提出的 CPR 方法進行驗證：

- 推薦模型對整個目標域上的使用者的模型表現。
- 推薦模型針對目標域上冷啟動使用者 (Cold-Start User) 的表現。
- 推薦模型針對目標域和源域上共同的使用者 (Shared User) 的表現。

這些實驗的目的在於，評估以 CPR 方法，使用並操作源域資訊，到底是否對目標域的推薦結果產生正面影響，此外，透過分別評估使用者中，性質不同的各個群體，將能夠讓我們更詳細地理解 CPR 能夠對跨領域推薦進行改善的部分。

我們將在三種不同的資料集上驗證我們的模型，並以單領域推薦 (Single Domain Recommendation) 和跨領域推薦 (Cross-Domain Recommendation) 中的數項先進演算法，和我們提出的 CPR 進行比較。

4.1 實驗設定

4.1.1 模型比較對象

為了能夠得到更有效的結果，並驗證設計跨領域推薦方法的必要性，除了和過往的跨領域推薦方法如 CMF [29], EMCDR [18] 等進行比較之外，也針對現今先進或過往經典的單領域推薦方法，如 LightGCN [10]、BPR [24] 進行比較。

以下將我們選用作為比較基準的模型，分成單領域和跨領域兩種推薦演算法進行說明：

單領域推薦演算法：

- BPR [24]：這是推薦領域基礎且持續被證明有效的模型，利用矩陣分解 (Matrix Factorization) 的概念，操作使用者 \leftrightarrow 物品的交互關係，加上負採樣 (Negative Sampling) 來利用資料中的隱性回饋 (Implicit Feedback)。
- LightGCN [10]：在 NGCF [36] 的基礎上進行改善的模型，移除了原始 NGCF 中的全連接神經網路 (Fully-Connected Neural Network) 和非線性層，並同樣以卷積圖神經網路 (GCN) 操作高階鄰居資訊，取得更好的效能和結果。

跨領域推薦演算法：

- CMF [29]：集體化矩陣分解 (Collective Matrix Factorization) 是一種以關係形式進行跨領域推薦的方法，而這裡的關係指涉的是使用者 \leftrightarrow 物品的交互關係。
- EMC DR [18]：基於表示向量和映射的跨領域推薦 (Embedding and Mapping framework for Cross-Domain Recommendation, EMC DR) 是一種基於表示向量的方法。它以矩陣分解 [15] 和 BPR [24] 訓練的使用者和物品的表示向量為基礎，額外訓練一個映射函數 (Mapping Function) 將源域的資訊映射到目標域上；其中這個映射函數可以是線性映射 (Linear Mapping) 或是多層感知器 (Multi-Layer Processon, MLP)。
- Bi-TGCF [17]：利用雙向傳輸圖的協同過濾網絡來進行跨領域推薦 (Cross Domain Recommendation via Bi-directional Transfer Graph Collaborative Filtering Networks, Bi-TGCF) 是以 NGCF [36] 的架構為基礎，先以神經網路進行源域和目標域其中之一的表示法學習，再同時訓練兩者的向量表示，從而達到跨領域推薦的效果。

4.1.2 資料集

本研究為了能夠驗證 CPR 能夠在現實的情況下進行跨領域推薦，選擇了六個工業級資料集進行比較。其中，亞馬遜的書籍/影視/服飾和珠寶/居家 (Amazon-Books/Movies and TV/Clothing, Shoes and Jewelry/Home and Kitchen) 為公開資料集^{1 2}，而電視 (TV) 和視訊 (VOD) 為非公開的資料集。

¹<https://jmcauley.ucsd.edu/data/amazon/>

²<https://reurl.cc/Q6pdZp>

其中，選擇此六個資料集的理由在於，跨領域推薦的資料集由於需要同時兩個領域的完整推薦資料，一般而言並不容易取得，因此我們首先和合作廠商取得了 TV-VOD 資料集作為基礎。接著，針對其他的跨領域資料集，我們利用了亞馬遜評論資料集 (Amazon Review Dataset)。該資料集由亞馬遜的使用者針對不同商品的評論和評分所組成，並依據商品性質分為不同領域，我們發現這樣的設定和跨領域推薦不謀而合，因此便從中節選了內容高度相關的兩組共四個資料集，組成我們進行比較的另兩個跨領域推薦資料集。

我們將其兩兩配對，分別作為源域和目標域進行推薦，並整理其統計列表如下：

	TV-VOD		CSJ-HK		MT-B	
	Source: TV	Target: VOD	Source: CS	Target: CJK	Source: MT	Target: B
#User	22,967	24,915	850,444	564,157	118,061	1,095,159
% of shared user	96.06	88.58	26.04	39.25	34.84	3.75
#Item	781,241	21,814	265,224	152,149	45,060	494,737
Interactions	18,277,659	5,346,812	3,970,542	2,453,346	497,408	6,396,932
Density	0.1018%	0.9837%	0.0017%	0.0028%	0.0093%	0.0011%

表 4.1: 資料集統計數據

4.1.3 實驗設定與方法說明

本研究主要目的在於利用提出的跨領域推薦演算法 (UET-CPR) 同時操作源域和目標域的資訊，進而達到補充冷啟動使用者過往缺乏的目標域資訊，進而改善其最終推薦成果。因此在實驗中，必須證明 UET-CPR 能達到下列兩個條件。

- UET-CPR 能透過利用源域的資訊，對目標域缺乏的資訊進行補充，從而達到比一般單領域推薦演算法更好的成果。
- UET-CPR 由於其模型的设计，能比一般的跨領域推薦演算法效果更好，且對比於補充了源域資訊的單領域推薦演算法，也能夠達到同等甚至更好的結果。

為了能夠與上述第二點中提及的，「補充了源域資訊的單領域推薦演算法」進行比較，我們將前述提及的兩個單領域演算法 (BPR, LightGCN) 分成兩種設定，其一為僅提供目標域資訊，並以此進行建圖和推薦的一般設定，以及將目標域和源域的資訊相互串接，形成一張大圖之後，針對目標域進行推薦的特殊設定。其中，採用第二種設定的單領域推薦演算法，將在方法名稱後方以「+」註記。

此外，爲了能對結果進行更細緻的觀察，我們將使用者依照其在源域和目標域出現的情況分爲三個群體，並針對模型在這三個使用者群體上的表現進行評估：

- 目標域使用者
- 冷啓動使用者
- 共同使用者

其中，目標域使用者指的是所有在目標域資料上出現過的使用者；冷啓動使用者則是在源域上出現，但未在目標域上出現的使用者，故稱爲冷啓動 (cold-start) 使用者。而共同使用者則是在源域和目標域均出現過的使用者。值得一提的是，爲了能夠實際評估冷啓動使用者在目標域的表現，我們這裏所使用的冷啓動使用者，是從共同使用者中選取，並刪除掉其在目標域上的活動紀錄的使用者，使模型在無法看到其目標域資訊的情況下進行訓練和推薦，且也能進行相關的結果評估。

4.2 研究問題說明

本研究想藉由上述的實驗設計，來回答下列的研究問題：

- 在考量所有目標域使用者的情況下，對於一般的單領域推薦演算法而言，同時給予源域和目標域資訊，能不能比單純給予目標域資訊，達到更好的成果？而我們的 UET-CPR 又是因爲什麼因素能比一般的單領域推薦演算法做的更好？
- 在跨領域的演算法之中，它們又各自因爲什麼樣不同的因素，而導致了其模型最終表現的差異？
- 冷啓動使用者的設定會造成什麼問題？又可以透過什麼樣的設計來解決呢？

4.3 實驗結果

4.3.1 目標域上的使用者

我們在這裏評估模型針對目標域上所有使用者的推薦表現 (見 4.2)。

	TV-VOD		CSJ-HK		MT-B	
	Recall@10	NDCG@10	Recall@10	NDCG@10	Recall@10	NDCG@10
BPR	0.9230	0.8329	0.2717	0.1718	0.5951	0.4680
BPR+	0.9060	0.7386	0.3045	0.1895	0.5986	0.4794
LightGCN	0.7328	0.6617	0.3480	0.2240	0.6202	0.4888
LightGCN+	0.7050	0.5363	0.3685	0.2317	0.6228	0.4923
CMF	0.4425	0.3340	0.2095	0.0950	0.3042	0.2367
EMCDR	0.2085	0.1504	0.2612	0.1555	0.4412	0.3202
Bi-TGCF	-	-	0.2978	0.1773	0.5976	0.4411
UET-CPR	0.9470	0.8406	0.3300	0.2100	0.6339	0.4974

表 4.2: 模型在目標域的整體表現

首先值得注意的是，並不是所有的單領域推薦演算法，都會在引進源域資訊後，提升其跨領域推薦的表現(如 BPR [24]、LightGCN [10] 的表現在 TV-VOD 資料集均有些許降低)，說明引進源域資訊是否對整體推薦表現有所幫助，還需要看模型性質而定。而像 LightGCN [10] 之類引進鄰居聚合的推薦演算法，則能夠有一定程度的提升。將其和 UET-CPR，以及 BPR 等模型的表現相較，可以得到的是，鄰居聚合可以做為一種提升推薦模型表現的技巧引入，對於跨領域推薦也有一定的效果，但和 UET-CPR 相比，顯示 UET-CPR 中，一些針對跨領域推薦問題本身的模型設計，依然是提升最終結果的核心因素。

而就跨領域演算法而言，CMF [29] 和 EMCDR [18] 雖然為跨領域推薦研究中重要的演算法，在這三個工業級資料集的表現卻不佳，甚至不如 BPR 等經典的單領域推薦演算法，原因在於，CMF 和 EMCDR 均不是端到端 (End-to-End) 的推薦模型，而是透過訓練好的源域和目標域表示法(如以 BPR 訓練)，額外訓練一個矩陣轉換函數來達到跨領域推薦的結果，自然不容易和一般端到端的推薦演算法相比；此外，在 CMF 和 EMCDR 的矩陣轉換函數中，主要就是使用源域和目標域中，共同使用者的資訊來進行訓練，而在稀疏性 (Sparsity) 較高的工業級資料中，共同使用者僅占了極小的部分，因此使用這類方法的結果，也自然不會太好。

相反地，Bi-TGCF [17] 由於交互並共同訓練源域和目標域的表示法，從而透過這樣端到端的模型設計，達到更好的成果。然而，由於其模型結構相當複雜繁瑣，因此無法成功地在有效時間和空間內訓練最大的 TV-VOD 資料集，是其主要的缺點。

綜上所述，UET-CPR 透過使用者表示法轉換的方式，成功引入了鄰居聚合的概念，也讓源域和目標域的資訊進行合成，是其在目標域整體表現勝過其餘單領域推薦演算法的原因。此外，由於能夠有效地建立起端對端的訓練模型，並利用前述的資訊合成手法，同時訓練源域和目標域的表示法，是其較其他跨領域模型更優秀的原因。

4.3.2 目標域上的冷啟動使用者

我們在這裏評估模型針對目標域上的冷啟動使用者的推薦表現。

	TV-VOD		CSJ-HK		MT-B	
	Recall@10	NDCG@10	Recall@10	NDCG@10	Recall@10	NDCG@10
BPR	-	-	-	-	-	-
BPR+	0.7040	0.5763	0.1990	0.1224	0.3491	0.2587
LightGCN	-	-	-	-	-	-
LightGCN+	0.6680	0.4778	0.2650	0.1650	0.3836	0.2844
CMF	0.0525	0.0270	0.0000	0.0000	0.0001	0.0000
EMCDR	0.3870	0.2465	0.2385	0.1328	0.0461	0.0219
Bi-TGCF	-	-	0.2170	0.1318	0.3473	0.2447
UET-CPR	0.7325	0.6556	0.2363	0.1492	0.4016	0.3010

表 4.3: 模型在目標域上冷啟動使用者的表現

如前所述，這裏的冷啟動使用者是指僅有在目標域上有活動紀錄的使用者，我們試圖觀察其在目標域上的表現。值得一提的是，由於僅提供目標域資訊的 BPR [24] 和 LightGCN [10] 無法生成這類使用者的代表向量，故不會有其整體表現。

而由於冷啟動使用者因為缺乏目標域資訊，所以不若其他在目標域上活躍的使用者容易推薦。但由於這是許多工業級推薦系統時常遇到的情況，探討不同模型在這類設定下的表現差異和原因，也是跨領域推薦研究中重要的一環。

首先，針對單領域推薦方法，使用了高階資訊 (high-order) 資訊的 [10] 分數多數下降，方法相對單純的 BPR [24] 反而有很不錯的表現。這是因為源域資訊雖然對跨領域推薦相當重要，但針對目標域的推薦而言，其實是同時扮演著補充資訊和雜訊的角色，需要謹慎處理；而由於在前述的目標域推薦中，有充足的目標域資訊，它的雜訊成分將會被稀釋，因此不會構成太大影響；但在冷啟動推薦的設定中，由於目標域資訊缺乏，源域資料扮演了過於重要的角色，因此其雜訊的性質便被放大，從而導致了使用高階資訊的模型，因為使用了過多雜訊而對結果產生負面的影響。

本研究提出的 UET-CPR 則透過訓練時僅抽樣目標域使用者的技巧，有效地針對目標域使用者和與其相關的源域和目標域物品，進行表示法學習，讓源域的物品也能充分得到目標域的資訊，再用這些源域物品來生成冷啟動使用者的表示法，從而避免了吸收過多雜訊的情況，達到更好的成果。

值得注意的是，由於 CMF 僅基於對目標域和源域的共通使用者進行矩陣轉換，再以此結果對其他使用者/物品進行推行，因此其表現很大程度受到共通使用者佔全體使用者的比例影響。然而，在我們的冷啟動設定中，被觀察推薦表現的

冷啟動使用者，均是沒有目標域活動紀錄的，從而使其在冷啟動設定中，表現出奇的低。

而對 EMCDR 而言，由於其有得到一定程度的資訊，從而讓其表現大部分優於 CMF，但依舊受限於非端到端 (Non-End-to-End) 的模型設計，使其效果不若一般的端到端單領域推薦演算法。

4.3.3 目標域和源域上的共同使用者

我們在這裏評估模型針對目標域上和源域的共同使用者的推薦表現。

	TV-VOD		CSJ-HK		MT-B	
	Recall@10	NDCG@10	Recall@10	NDCG@10	Recall@10	NDCG@10
BPR	0.7145	0.6459	0.1972	0.1239	0.3662	0.2748
BPR+	0.6835	0.5487	0.1957	0.1204	0.3537	0.2596
LightGCN	0.7612	0.6615	0.2255	0.1440	0.3761	0.2800
LightGCN+	0.7664	0.5681	0.2727	0.1682	0.3790	0.2767
CMF	0.4305	0.3336	0.2620	0.1199	0.2117	0.1307
EMCDR	0.1395	0.0726	0.1705	0.0965	0.0473	0.0213
Bi-TGCF	-	-	0.2225	0.1394	0.3476	0.2435
UET-CPR	0.7241	0.6383	0.2356	0.1513	0.4089	0.3108

表 4.4: 模型在目標域和源域上共同使用者的表現

可以發現在共同使用者占多數的 TV-VOD 和 CSJ-HK 資料集中，模型間並沒有太顯著的差異。然而，針對共同使用者相對稀少的 MT-B 資料集，UET-CPR 的效果是最好的，說明我們提出的模型的強項，在於有效地從沒有重疊的目標域和源域資訊中，提取資訊供後續使用。

4.4 關於 UET-CPR 模型與其他模型在不同資料集上表現差異的討論

從上述三種不同的實驗中，我們可以發現兩點：

- 在 TV-VOD 和 MT-B 資料集上，本研究提出的 UET-CPR，可以在目標域整體，以及冷啟動兩個設定中，得到比其他模型更好的推薦結果。
- 然而，針對 CSJ-HK 資料集，以及源域和目標域的共同使用者的設定下，UET-CPR 的表現就顯得沒有那麼穩定。

我們將針對希望後續的討論，能夠解答上述兩個問題。

4.4.1 不同資料集對結果的影響

從先前的各項數據中，我們可以發現的是，在不同資料集的情況下，其實並無法從實驗中得到太一致的結論，我們將其歸因於資料集的差異，而為了探討這些差異，這裏將利用 4.1 對資料集的描述統計，先整理三組資料集的各項性質。其中，最值得注意的是，共同使用者的比例，以及源域、目標域的密度 (Density) 差異。

- TV-VOD：總體密度比其他兩個資料及高上許多，且共同使用者佔比極高。
- CSJ-HK：共同使用者佔比稍高，且目標域密度高於源域，其中源域密度相當低。
- MT-B：共同使用者佔比低，且源域密度高於目標域。

以下分段就上述資料集性質，針對較有疑問的部分進行討論。

4.4.2 TV-VOD 資料集

首先，就目標域表現進行討論的話，可以發現雖然 LightGCN [10] 作為當前推薦領域最重要且先進的方法之一，其在 TV-VOD 資料集的表現是遠不如 BPR [24] 的。這裏推測原因是因為 LightGCN 相較於 BPR 方法，使用了額外的高階特徵 (high-order features)，以取得較遠的離中心節點較遠的特徵，這個技巧在相對稀疏的推薦問題上相當有用，能有效補充中心節點缺乏的資訊。然而，針對如此密集的 TV-VOD 資料集，並不依於這些較遠的特徵來進行資訊補充，對其需求相對較低，因此 LightGCN 取回的高階特徵，比起資訊，可能更接近噪聲 (noise) 的角色。而這樣的性質，同樣反映在 BPR+、LightGCN+ 等額外考慮源域資訊的實驗設定，其實並不如只取目標域資料的效果來得好。

4.4.3 CSJ-HK 資料集

接著，對於 CSJ-HK 資料集，由於其目標域的密度高於源域，可說明其實對於大部分模型而言，資訊來源比起源域，目標域將佔更大的比重；且因為共同使用者佔比較高，其實比起跨領域推薦問題，更接近單領域推薦的情況。因此可以得到的是，LightGCN [10] 在類似單領域推薦的設定下，是所有方法中最優秀的。

4.4.4 MT-B 資料集

由於源域較目標域密集，且共同使用者佔比較低，。因為目標域相對稀疏的緣故，估需要從源域進行資訊補充，且因為共同使用者佔比較低，相對的，冷啟動使用者就佔了大多數。而上述提到的問題，才是現實情況下，跨領域推薦亟待解決的問題。而在這樣的設定中可以發現，本研究提出的 UET-CPR 方法均能得到最好的結果，說明本文提出的使用者表示法轉換技術，針對實際上的跨領域推薦問題，能夠提供實質有效的解決方案。

4.4.5 小結

上述針對資料集的說明和討論，可以總結為以下三點：

- 當目標域資訊相對充足時，情境將更接近單領域推薦，而使 LightGCN 佔優。
- 當總體密度足夠高時，對鄰居聚合、高階資訊等方法的使用就要更加謹慎，以免收集過多噪聲。
- 理想的跨領域推薦資料集，應選擇源域密度高於目標域，且共同使用者占比低者。

4.5 個案探討 - 以 TV-VOD 資料集為例

TV-VOD 資料集以日本地區的電視節目和影片作為跨領域推薦中的源域和目標域，本研究將以其冷啟動使用者的相關活動紀錄，進行個案探討，並說明在冷啟動情境中，UET-CPR 較其他方法更優秀的理由。

選擇 TV-VOD 做個案探討的原因在於，其餘的 CSJ-HK 和 MT-B 資料集，因為亞馬遜將其使用者和物品進行有效雜湊的緣故，較難還原至原始的資料形式，故無法知道該物品原本是何種商品，才會選擇了元數據 (metadata) 較為完整的 TV-VOD 資料集進行研究。

其中，作為比較基準，本研究節選了在 TV-VOD 資料集中，冷啟動情境表現較好的 BPR [24] 和 LightGCN [10] 兩種方法進行比較。這兩種方法除了在 TV-VOD 資料集中表現最為優秀，兩個方法本身也分別代表了「經典的推薦演算法」和「當前最好的一般推薦演算法」，相信能透過下列的個案探討及和 UET-CPR 的比較，側面說明兩方法的優勢和不足。

我們首先隨機選定了某使用者，並列出其觀看紀錄中，人氣較高的三部：

- 日劇 - 仁醫
- 動畫 - 名偵探柯南
- 動畫 - 鬼滅之刃

從上數的幾部作品中，可以發現該使用者喜歡的傾向接近 1) 熱門日劇 2) 少年向動畫。且使用者觀看了多集名偵探柯南動畫，說明其對懸疑類作品的偏好。

而 UET-CPR 的推薦清單前三名則是：

- 動畫 - 名偵探柯南劇場版
- 日劇 - 嫌疑人 室井慎次
- 日劇 - 圈套劇場版

從上述的推薦清單來看，即使並沒有納入文字等外部資訊，UET-CPR 依舊很好地捕捉到了《名偵探柯南》這一個關鍵作品，且推薦的其他作品也都是與懸疑案件相關的熱門日劇。說明即使在冷啟動的情境中，UET-CPR 依然能很好地利用源域資訊，學習使用者的喜好。

而 BPR 和 LightGCN 的推薦清單前三名則分別是：

- 動畫 - 名偵探柯南劇場版
- 動畫 - 可愛巧虎島
- 動畫 - 死神

和

- 電影 - 信長協奏曲
- 動畫 - 異世界超能魔術師
- 動畫 - 電王劇場版

從上述兩組推薦清單可以發現，雖然 BPR 和 LightGCN 均捕捉到了少年向動畫的重要特徵，但細究其推薦的作品，卻會發現有些不太合理的地方。比如 BPR 推薦的可愛巧虎島，和該使用者日常觀賞作品的年齡範圍並不相稱；而 LightGCN 則是推薦了大量動畫，卻不太涉及懸疑性質，也不和日劇類別相關。說明其推薦效果實質上並不如 UET-CPR 精準。

第五章

結論

本研究想討論的問題是，在跨領域推薦問題之中，能否透過引進源域資料，補充目標域缺乏的資訊，從而提升整體推薦的結果，並改善冷啟動使用者的問題。

本研究的主要貢獻在於：

1. 提出了一個基於使用者表示法轉換的跨領域推薦演算法。
 - 通過將目標域和源域的資訊進行結合，從而在目標域推薦中使用來自源域的資訊。
 - 利用抽樣方式的設計，試圖減少源域資訊為目標域推薦的訓練過程中，帶來雜訊的可能。
 - 並以理論和實驗證明，其相較過去傳統的跨領域演算法更為有效。
2. 探討不同性質的推薦演算法，在各式跨領域推薦的情境中的表現差異和原因
 - 針對傳統跨領域推薦演算法和當前主流跨領域推薦演算法的比較，發現端到端的設計依然是提升表現的重要結構，從而說明了傳統跨領域推薦演算法在工業級資料及表現不佳的原因。
 - 為了說明引進源域資訊和跨領域架構設計的重要性，本研究也引入了兩個單領域推薦中表現最好和最重要的方法進行比較，證明了在大部分情況下，UET-CPR 引進源域資訊並有效利用的做法，表現都較單領域推薦演算法更為優秀。
 - 針對不同方法在不同資料集表現的差異，說明了在跨領域推薦資料集中，造成表現差異的因素：
 - 資料集總體的密度
 - 源域和目標域密度的差異程度

- 共同使用者佔源域和目標域資料的比例
- 對冷啟動情境中的使用者進行了個案探討，說明了 UET-CPR 在使用源域資訊後，能如何改善最終的推薦結果。

而作為後續研究的方向，則可以往以下兩個方向發展：

- 結合更好的單領域推薦演算法，從而得到更堅實的模型表現，並更深入地探究各項模型設計的技巧和巧思，能夠如何地改變模型在不同情境下的推薦表現
- 針對源域和目標域的資訊融合，加上可訓練權重，從而更細緻地描述跨領域推薦問題，在不同資料集上取得更一致的結果。



參考文獻

- [1] A. Andoni, R. Panigrahy, G. Valiant, and L. Zhang. Learning polynomials with neural networks. In *ICML*, volume 32 of *JMLR Workshop and Conference Proceedings*, pages 1908–1916. JMLR.org, 2014.
- [2] C. Burges, T. Shaked, E. Renshaw, A. Lazier, M. Deeds, N. Hamilton, and G. Huelender. Learning to rank using gradient descent. In *ICML '05: Proceedings of the 22nd international conference on Machine learning*, pages 89–96, New York, NY, USA, 2005. ACM.
- [3] C.-C. Chang and C.-J. Lin. LIBSVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology*, 2:27:1–27:27, 2011. Software available at <http://www.csie.ntu.edu.tw/~cjlin/libsvm>.
- [4] W.-L. Chiang, X. Liu, S. Si, Y. Li, S. Bengio, and C.-J. Hsieh. Cluster-gcn: An efficient algorithm for training deep and large graph convolutional networks. *CoRR*, abs/1905.07953, 2019.
- [5] P. Covington, J. Adams, and E. Sargin. Deep neural networks for youtube recommendations. In *Proceedings of the 10th ACM Conference on Recommender Systems*, New York, NY, USA, 2016.
- [6] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding, 2019.
- [7] J. Gilmer, S. S. Schoenholz, P. F. Riley, O. Vinyals, and G. E. Dahl. Neural message passing for quantum chemistry. *CoRR*, abs/1704.01212, 2017.
- [8] A. Grover and J. Leskovec. node2vec: Scalable feature learning for networks. *CoRR*, abs/1607.00653, 2016.
- [9] W. Hamilton, Z. Ying, and J. Leskovec. Inductive representation learning on large graphs. In *Advances in Neural Information Processing Systems*, page 11, 2017.

- [10] X. He, K. Deng, X. Wang, Y. Li, Y. Zhang, and M. Wang. Lightgcn: Simplifying and powering graph convolution network for recommendation. *CoRR*, abs/2002.02126, 2020.
- [11] X. He, L. Liao, H. Zhang, L. Nie, X. Hu, and T.-S. Chua. Neural collaborative filtering. In *Proceedings of the 26th International Conference on World Wide Web, WWW '17*, page 173–182, Republic and Canton of Geneva, CHE, 2017. International World Wide Web Conferences Steering Committee.
- [12] S. Hochreiter and J. Schmidhuber. Long short-term memory. *Neural computation*, 9(8):1735–1780, 1997.
- [13] K. Hornik, M. Stinchcombe, and H. White. Multilayer feedforward networks are universal approximators. *Neural Networks*, 2(5):359 – 366, 1989.
- [14] T. N. Kipf and M. Welling. Semi-supervised classification with graph convolutional networks. *5th International Conference on Learning Representations*, 2016.
- [15] Y. Koren, R. Bell, and C. Volinsky. Matrix factorization techniques for recommender systems. *Computer*, 42(8):30–37, Aug. 2009.
- [16] Y. LeCun, Y. Bengio, and G. Hinton. Deep learning. *Nature*, 521(7553):436–444, 2015.
- [17] M. Liu, J. Li, G. Li, and P. Pan. Cross domain recommendation via bi-directional transfer graph collaborative filtering networks. In M. d’Aquin, S. Dietze, C. Hauff, E. Curry, and P. Cudré-Mauroux, editors, *CIKM*, pages 885–894. ACM, 2020.
- [18] T. Man, H. Shen, X. Jin, and X. Cheng. Cross-domain recommendation: An embedding and mapping approach. In *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence, IJCAI-17*, pages 2464–2470, 2017.
- [19] T. Mikolov, K. Chen, G. Corrado, and J. Dean. Efficient estimation of word representations in vector space. In Y. Bengio and Y. LeCun, editors, *1st International Conference on Learning Representations, ICLR 2013, Scottsdale, Arizona, USA, May 2-4, 2013, Workshop Track Proceedings*, 2013.
- [20] T. Mikolov, I. Sutskever, K. Chen, G. Corrado, and J. Dean. Distributed representations of words and phrases and their compositionality. *CoRR*, abs/1310.4546, 2013.
- [21] B. Perozzi, R. Al-Rfou, and S. Skiena. Deepwalk: Online learning of social representations. *CoRR*, abs/1403.6652, 2014.

- [22] R. Raina, A. Madhavan, and A. Y. Ng. Large-scale deep unsupervised learning using graphics processors. In A. P. Danyluk, L. Bottou, and M. L. Littman, editors, *ICML*, volume 382 of *ACM International Conference Proceeding Series*, pages 873–880. ACM, 2009.
- [23] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi. You only look once: Unified, real-time object detection, 2015. cite arxiv:1506.02640.
- [24] S. Rendle, C. Freudenthaler, Z. Gantner, and L. Schmidt-Thieme. BPR: bayesian personalized ranking from implicit feedback. *CoRR*, abs/1205.2618, 2012.
- [25] S. Rendle, W. Krichene, L. Zhang, and J. R. Anderson. Neural collaborative filtering vs. matrix factorization revisited. *CoRR*, abs/2005.09683, 2020.
- [26] P. Resnick, N. Iacovou, M. Suchak, P. Bergstrom, and J. Riedl. Grouplens: an open architecture for collaborative filtering of netnews. In *CSCW '94: Proceedings of the 1994 ACM conference on Computer supported cooperative work*, pages 175–186, New York, NY, USA, 1994. ACM Press.
- [27] B. Sarwar, G. Karypis, J. Konstan, and J. Riedl. Incremental singular value decomposition algorithms for highly scalable recommender systems. In *Proceedings of the 5th International Conference in Computers and Information Technology*, 2002.
- [28] K. Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.
- [29] A. P. Singh and G. J. Gordon. Relational learning via collective matrix factorization. In *Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 650–658, 2008.
- [30] J. Tang, M. Qu, M. Wang, M. Zhang, J. Yan, and Q. Mei. LINE: large-scale information network embedding. *CoRR*, abs/1503.03578, 2015.
- [31] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Laiser, and I. Polosukhin. Attention is all you need. In I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, editors, *Advances in Neural Information Processing Systems 30*, page 5998–6008. Curran Associates, Inc., 2017.
- [32] P. Veličković, G. Cucurull, A. Casanova, A. Romero, P. Liò, and Y. Bengio. Graph attention networks, 2018.
- [33] M. Volkovs, G. W. Yu, and T. Poutanen. Content-based neighbor models for cold start in recommender systems. In *Proceedings of the Recommender Systems Challenge 2017*, pages 1–6. 2017.

- [34] M. Volkovs, G. W. Yu, and T. Poutanen. Dropoutnet: Addressing cold start in recommender systems. In I. Guyon, U. von Luxburg, S. Bengio, H. M. Wallach, R. Fergus, S. V. N. Vishwanathan, and R. Garnett, editors, *Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017, December 4-9, 2017, Long Beach, CA, USA*, pages 4957–4966, 2017.
- [35] X. Wang, X. He, Y. Cao, M. Liu, and T.-S. Chua. Kgat: Knowledge graph attention network for recommendation. *CoRR*, abs/1905.07854, 2019.
- [36] X. Wang, X. He, M. Wang, F. Feng, and T. Chua. Neural graph collaborative filtering. *CoRR*, abs/1905.08108, 2019.
- [37] K. Xu, W. Hu, J. Leskovec, and S. Jegelka. How powerful are graph neural networks? *CoRR*, abs/1810.00826, 2018.
- [38] R. Ying, R. He, K. Chen, P. Eksombatchai, W. L. Hamilton, and J. Leskovec. Graph convolutional neural networks for web-scale recommender systems, 2018. cite arxiv:1806.01973Comment: KDD 2018.

