

國立交通大學

資訊科學研究所

碩士論文

圖書館個人化館藏推薦系統

A Personalized Recommender System for Library

研究生：余明哲

指導教授：柯皓仁 博士

楊維邦 博士

中華民國九十二年六月

A Personalized Recommender System for Library

Student: Ming-Che Yu Advisor: Dr. Hao-Ren Ke, Dr. Wei-Pang Yang

Institute of Computer and Information Science

National Chiao Tung University

ABSTRACT

Recommender systems are popularly being used in e-Commerce to encourage consumers to purchase more products. In library, recommender systems can also be used to help patrons find collections to borrow, and increase the value of library. In this paper, we propose a personalized recommender system for library, which combines collaborative filtering and content-based filtering recommendation methods. We hope patrons can find books which fit their preferences through this personalized recommender system. In this recommender system, first, we use association mining to discover association rules between users and between books. Second, we take these associations rules to find user groups which have the same preference for collaborative filtering, and build representation model for content-based filtering. Finally, we use collaborative filtering to find a recommendations list which contains books collecting from other patrons' borrowing history, and we would re-rank this recommendations list by content-based filtering to cater to user's preference. We implemented this recommender system on National Chiao Tung University (NCTU) library's personalized information environment "myLibrary" (<http://mylibrary.e-lib.nctu.edu.tw/>), and the feedback received from users of library shows that most users satisfied with our recommendations.

Keywords: Recommender Systems, Personalization, Data Mining, Association Rules, Collaborative Filtering, Content-Based Filtering, Library.

圖書館個人化館藏推薦系統

A Personalized Recommender System for Library

研究生：余明哲

指導教授：柯皓仁博士，楊維邦博士

國立交通大學資訊科學研究所

摘要

推薦系統是網路商店上常運用的技術，主要用來提高顧客的購買慾望。我們將此技術應用在圖書館上，希望能藉由個人化館藏推薦系統推薦給讀者圖書館中其有興趣的館藏，幫助讀者使用圖書館資源。同時也希望圖書館這個新的館藏推薦服務能增加館藏的利用率，並提高圖書館的價值。

在本論文中，將利用資料探勘 (Data Mining) 中關聯規則探勘 (Association Mining) 的技術，從讀者的借閱歷史檔中找出頻繁項目集 (Frequent Itemsets) 和關聯規則 (Association Rules)。從這些探勘的結果中，分析得知讀者與讀者間和館藏與館藏間存在的關係，由此取得讀者的興趣。接著先利用推薦系統中常用的協力式過濾 (Collaborative Filtering) 找出給讀者的推薦書目清單，再以內容導向過濾 (Content-Based Filtering) 的方法將推薦清單依照讀者興趣做個人化的排序，最後能找出合適的館藏推薦給讀者。

個人化館藏推薦系統目前實作於交通大學浩然圖書館的個人化資訊環境 myLibrary 中(<http://mylibrary.e-lib.nctu.edu.tw/>)。並且從收集回來的讀者回應中發現，這套系統的確可以有效地依照讀者的興趣將館藏推薦給讀者。

關鍵字：推薦系統、個人化、資料探勘、關聯規則、協力式過濾、內容導向過濾、圖書館

誌謝

感謝指導教授柯皓仁老師與揚維邦老師這兩年來在各方面的悉心指導。除了學業上的教導外，也培養我獨立思考和研究的能力。常能適時地引導我研究方向，並在我遇到困難時能指引方向，鼓勵我、給予我信心，使我不會徬徨無措、孤立無援。當我有所懈怠時，也會不忘督促的責任給予提醒，使我在研究上能更上層樓。同時，生活上也給我許多的啟示與照顧，帶給我這段人生中難得且珍貴的回憶。

實驗室裡的學長姐、同學和學弟，感謝你們陪我走過這段時間，給予我包容，不論我高興或者不高興的時候伴在我身邊，聽我抱怨，替我加油打氣。在課業上常給予我很多的幫助，而且時時關心我的生活，使我不感到孤單。尤其感謝葉鎮源學長，在論文上提供我許多建議與協助。

還要感謝交通大學圖書館中館合計劃室、數位圖書資訊組和參考諮詢組中曾幫助過我的朋友們，由於你們的幫忙使我能順利完成這篇論文。其中，蔡淑琴小姐提供我圖書館的資料來源，劉玉芝小姐和湯春枝小姐給予我許多圖書館方面的指導和建議，在這邊特別感謝她們。

最後，感謝我的父母親及家人能無條件地支持我，提供我一個能安心休息的避風港，並且長久以來不斷地給我最溫暖的照顧，讓我能無牽無掛地順利完成學業。僅將此篇論文獻給他們。

癸未年 孟夏

目錄

英文摘要	i
中文摘要	ii
誌謝	iii
目錄	iv
表目錄	v
圖目錄	vi
方程式目錄	vii
第一章 緒論	1
第一節 研究動機及目的	1
第二節 研究方法與目標	2
第三節 論文架構	6
第二章 相關研究工作	7
第一節 推薦系統與電子商務	8
第二節 個人化推薦系統	11
第三節 協力式過濾與關聯規則探勘	15
第四節 內容導向過濾與系統模型建立	17
第五節 圖書推薦系統	19
第三章 圖書館館藏推薦	27
第一節 協力式過濾的推薦方法	27
第二節 內容導向的推薦方法	39
第三節 結合協力式過濾和內容導向的推薦方法	43
第四章 系統實作	44
第一節 系統架構	44
第二節 前置處理	46
第三節 找出相關讀者群和建立館藏與讀者的表示法	48
第四節 產生推薦清單	52
第五節 系統在 myLibrary 上的呈現方式	55
第五章 館藏推薦滿意度評估	58
第一節 收集讀者對館藏推薦的回應	58
第二節 計算讀者對館藏推薦的滿意度	59
第三節 結果分析	60
第六章 結論與未來研究方向	63
第一節 結論	63
第二節 未來研究方向	64
參考文獻	67

表目錄

表 1 以一本書為一筆交易的交易集	27
表 2 代表具有相同興趣讀者群的頻繁項目集	28
表 3 以書籍為交易的交易集	30
表 4 聚集過後以類別為交易分別不考慮次數和考慮次數的交易集	30
表 5 項目集{3, 4}在各個交易集中的支持度	31
表 6 屬於相同類別的交易集	32
表 7 將相同類別的書聚集起來的新交易集	32
表 8 以一個類別為一筆交易的交易集	33
表 9 單一項目的出現頻率	33
表 10 經過正規化處理後的交易集	34
表 11 找出項目集{1, 2}的支持度	36
表 12 所有長度為 2 的項目集的支持度	36
表 13 第二層中具有兩個項目的頻繁項目集	37
表 14 找出項目集{1, 3, 4}的支持度	37
表 15 找出項目集{2, 3, 5}的支持度	38
表 16 以書為交易項目的交易集	40
表 17 探勘第三層分類所需的交易集	40
表 18 探勘第二層分類所需的交易集	41
表 19 探勘第一層分類所需的交易集	41
表 20 各階層分類探勘的最小支持度、最小確信值及最後關聯規則數量	49
表 21 各階層所探勘出的關聯規則和確信值	49
表 22 包含讀者 3 的頻繁項目集	52
表 23 讀者借閱歷史	53
表 24 讀者 3 在協力式過濾推薦裡得到的推薦數目	53
表 25 書的屬性向量表示式	54
表 26 推薦書目的新的推薦度和排序	54
表 27 圖書館個人化館藏推薦系統各細部元件	57
表 28 系統滿意度調查結果	61

圖目錄

圖 1 相關研究文獻	7
圖 2 一個兩層的分類樹	18
圖 3 交通大學浩然圖書館館藏查詢使用者介面	20
圖 4 交通大學浩然圖書館館藏查詢結果呈現介面	21
圖 5 科幻圖書館藏的主題推薦	22
圖 6 科幻圖書館藏中的分類推薦 - 以“時空旅行”分類為例	22
圖 7 個人化智慧型館藏查詢結果 - 以“資料庫”為書刊名關鍵字查詢	23
圖 8 個人借閱歷史查詢介面	24
圖 9 推薦相關館藏畫面	24
圖 10 個人化智慧型查詢系統介面	26
圖 11 圖書館個人化館藏推薦系統架構圖	45
圖 12 協力式過濾需要的交易集	47
圖 13 內容導向過濾需要的交易集	47
圖 14 相關讀者群的頻繁項目集	48
圖 15 類別 312 的屬性向量表示式	50
圖 16 讀者的個人興趣向量	51
圖 17 myLibrary 中“我的館藏”畫面	56
圖 18 系統推薦館藏的呈現畫面	56
圖 19 設定個人興趣畫面	57
圖 20 讀者滿意度回應畫面	58
圖 21 Top N 的平均誤差值	61
圖 22 Top N 的排序分數度量值	62
圖 23 Top N 的讀者平均滿意度	62

方程式目錄

方程式 1	文獻[8]中的權重給定公式.....	18
方程式 2	項目集在每個交易中支持度的計算公式.....	30
方程式 3	項目集於單筆交易內的支持度計算公式.....	34
方程式 4	以取最小值為交集運算的支持度計算公式.....	35
方程式 5	乏析資料探勘中項目集的支持度計算公式.....	35
方程式 6	協力式過濾的推薦分數計算公式.....	39
方程式 7	書籍屬性向量表示式.....	41
方程式 8	書籍屬性向量權重給定公式.....	42
方程式 9	讀者興趣向量計算公式.....	42
方程式 10	內容導向過濾的喜好度計算公式.....	42
方程式 11	新的推薦分數計算公式.....	43
方程式 12	讀者興趣向量修正公式.....	51
方程式 13	平均誤差值計算公式.....	59
方程式 14	排序分數度量值計算公式.....	59
方程式 15	排序分數度量值正規化計算公式.....	60

第一章 緒論

第一節 研究動機及目的

圖書館長久以來一直扮演著傳承人類知識的重要角色。每一個圖書館都蘊藏著許多的館藏、許多的資源和無數的智慧心血結晶，不但滿足一般人的求知慾，同時也提供了後人學習及發展新技術的基礎。如果圖書館館藏資料能被妥善利用，一定能帶給人類社會莫大的助益，萬一館藏資料乏人問津，那無疑地就是一種資源的浪費，是件相當可惜的事情。

1979 年美國 Pittsburgh 大學調查報告[7]中指出，圖書館的館藏資源僅有少數被有效利用。原因之一是圖書館和使用者之間互動不足。往往圖書館扮演被動的角色，無法主動提供資訊給使用者。例如當圖書館添購了新的館藏或增加新的服務，常常無法有效地告知使用者相關的訊息，造成使用者不清楚甚至不知道圖書館有哪些資源可以利用。另外，圖書館中的館藏，有部分屬於知名度比較高的熱門書，或者因為編排比較容易被使用者取得的館藏，其使用率一直很高，導致使用者不易借閱到；但是事實上圖書館還有許多相同類別且使用者感興趣的館藏沒有被借閱。由於使用者搜尋的技巧、習慣或找資料的方向不同，使得圖書館中使用者有需要的館藏，沒被找出來利用。結果圖書館裡豐富的館藏和服務因此被束之高閣而浪費掉了。這樣不但是圖書館資源的一種浪費，也是使用者的損失。

圖書館的真正價值不在於優良的硬體設備或豐富的館藏，而在於其館藏的利用率。也因此圖書館莫不致力於提升利用率以增加圖書館的價值。隨著時代的進步，提升利用率的方法從以前較被動式的，如增加採購借閱率較高的館藏類別、設計良好的編目方式和架位排列方式、採用方便有效率的館藏查詢系統等，也逐漸增加主動式的方法，如開設一些改進使用者尋找資料的技巧或指導使用者利用圖書館館藏資源的訓練課程等，這

些無非都是在吸引使用者利用圖書館資源，以提升圖書館的利用率。在[28], [29]這幾篇論文中，也提出以資料探勘 (Data Mining) 的技術和統計的方法，分析使用者的借閱歷史資料，找出一些規則及有用的資訊，使圖書館能根據這些規則和資訊主動吸引使用者到圖書館借閱館藏以提升館藏的使用率和借閱率，並且增強圖書館的經營與服務。而我們在這篇論文中所提出的個人化館藏推薦系統也是一種偏主動式的方法來吸引使用者使用圖書館資料。與[28], [29]比較，本論文提出的這套個人化館藏推薦系統，不僅能從使用者以往的借閱歷史中找出館藏間的關連性，同時也會找出使用者間的關連性，並且判斷出使用者的個人興趣。藉此來找出圖書館中使用者可能有興趣，但是卻還沒借閱過的館藏，然後把這些館藏推薦給使用者，藉此來鼓勵使用者借閱館藏。希望藉由這樣一種新的圖書館服務能提高圖書館的利用率，同時也增加圖書館的附加價值。

第二節 研究方法與目標

本論文提出的個人化圖書館館藏推薦系統，同時採用協力式過濾 (Collaborative Filtering) 和內容導向過濾 (Content-Based Filtering) 這兩個推薦方法。

協力式過濾的方法是一種最典型的推薦方法，自從 1992 年由 Goldberg 等人在[14]中提出後就被廣泛採用在一般的推薦系統中，甚至在 1997 年 Resnick 和 Varian 提出推薦系統 (Recommender System) [23]這個名詞之前，協力式過濾就是推薦系統的代名詞。協力式過濾主要的概念是使用者利用同儕對某物品既有的評價來預期自己對該物品可能的評價。這個想法就好像當我們要做某樣決定之前，會事先詢問一些有經驗的人，以他們的意見做為決定的參考。

本篇論文將協力式過濾的概念應用在圖書館上，希望使用者之間能互相分享資訊，每個使用者都是過濾代理人 (Filtering Agent)，可以幫忙篩選館藏，提供別人借閱書籍意見參考。而圖書館中每個使用者的借閱歷史可視為是給別人的館藏推薦清單，使用者可

以參考這些推薦清單，來挑選自己有興趣的館藏，而不用從圖書館眾多的館藏中盲目地尋找。

不過若一個圖書館有上千名使用者，如果要使用者從這上千份推薦清單中挑選自己感興趣的館藏，事實上的意義跟使用者要直接從館藏中挑選自己需要的館藏一樣困難且麻煩。而且這麼多的推薦清單中，其實並不是每一份都是適合自己需要的，因為每個人的興趣或目的不同，所以會借閱不同的館藏，分別產生有不同的借閱歷史和推薦清單，而其中可能只有某部分清單是真正符合使用者需要的。因此我們這套系統還必須事先幫使用者決定那些推薦清單是使用者需要或感興趣的，而不需要使用者每次都從這上千份推薦清單中挑選。

本論文用的方法是先？使用者挑選和他有相同借閱興趣的其他使用者，稱之為“同好” (Friends)。同好間因為借閱興趣類似，所以他們的借閱歷史，也就是本論文所認為他們給其他同好的推薦清單，對使用者來說有比較高的參考價值。而認定有相同借閱興趣的標準是依照使用者間借閱過相同館藏的數量，相同的館藏愈多時代表相同的借閱興趣就愈高。當使用者相同的借閱興趣高過某一個標準時，我們就認定他們是“同好”。

找出有相同借閱興趣同好的方法，主要是採用資料探勘中的關聯規則探勘 (Association Mining)，並且？了配合圖書館特殊的資料類型，同時也參考了乏析式資料探勘 (Fuzzy Data Mining) [17], [18]的方法。在實作上，則是採用[3]中所提出的以前序樹 (Prefix Tree) 為基本概念實作的 Apriori [2] 演算法來找出頻繁項目集 (Frequent Itemsets)，也就是找出有相同借閱興趣的使用者群。

內容導向過濾事實上比協力式過濾更早被應用在篩選資料上，通常內容導向過濾根據使用者輸入的關鍵字，由系統來篩選出符合這些關鍵字或者符合這些關鍵字語意概念 (Semantic Concept) 的資料。一般內容導向過濾的方法都會運用到資訊擷取 (Information Retrieval) 的技巧來萃取出資料中的內容[27]。顧名思義，內容導向過濾主要是根據資料的內容來篩選資料。這一點和協力式過濾相比是很明顯的不同，因為協力式過濾主要利

用眾人的意見來篩選資料，完全跟資料的內容無關。但是內容導向過濾通常需要收集許多系統中物件的內容資料，再利用一個模型 (Model) 根據物件的內容來替系統內的每個物件建立各自的屬性表示式 (Representation)，此模型通常是向量空間模型 (Vector Space Model)，建出來的屬性表示式以向量表示。內容導向的方法因為實際考慮到物件的內容屬性，所以系統中的物件利用各自的屬性向量表示式來比較彼此間相似度時，通常有相當高的準確率。

但是在推薦系統中要找出一個適合的模型來描述整個系統，並且從每個物件中抽取足夠的資訊來使系統中的每個物件都能各自的表示式並不容易。主因是在推薦系統中，通常系統中的每個物件能提供的資訊並不充足。例如推薦書籍，可以獲得的資料只有作者、出版社、出版日期、摘要等簡單的摘要資訊。只根據這些資料獲得的資訊很難建出一個可以描述整個系統的模型，所以內容導向過濾的方法相對的比較難被應用在推薦系統中。

在本論文提出的個人化圖書館館藏推薦系統中，採用的內容導向過濾是利用 Lawrence 等人在[8]中所提出來的的方法來建立整個系統模型。[8]這篇論文主要目的是建立一個超級市場中的推薦系統，每個消費者建立個人化的推薦清單當做購物時的參考。[8]在建立內容導向過濾模型的作法是利用超級市場中的商品分類樹 (Product Taxonomy) 為模型基礎，再探勘消費者的消費歷史資料，以找出分類樹中各個分類間的關聯規則，藉此來替市場中每個商品和每個消費者建立各自的表示式。

超級市場的情境和圖書館頗為類似，在本論文中，我們利用圖書館中原本就有的書目分類表為內容導向過濾的模型，再探勘使用者在圖書館的借閱歷史檔，找出書目分類表中各個分類之間的關聯規則。這樣利用[8]中所提出的方法就可以替圖書館館藏推薦系統中每本書和每個使用者找出各自的表示式。

找出每個物件的表示式後，接下來要找出系統內各個物件間的關係就相當容易，因為系統中物件的表示式都在同一個空間中，所以要比較彼此間的關係通常只要一個計算相似度的公式，就可以算出系統中物件間的相關程度。

因此在推薦系統中，若能夠建立使用者和推薦物品的屬性表示式分別代表使用者興趣和物品屬性，就可以輕易地計算出使用者對推薦物品的喜好度，而不需要像協力式過濾要參考其他人的意見。同時內容導向過濾也沒有像協力式過濾參考資料不足的問題，因為在協力式過濾中需要的是眾人對推薦物品的評價，如果某物件太少人甚至根本沒有人對其有評價，系統將無法為其做推薦。這個問題在內容導向過濾的推薦系統中不存在，因為內容導向過濾的推薦系統中每個物件都是依照自己的屬性內容來建立表示式，而要？每個人找出推薦只要找出與使用者興趣符合度比較高的物件即可。而且因為以內容導向過濾為主的推薦系統中每個人是以自己的興趣來建立個人的興趣表示式(興趣向量)，所以每個人的興趣向量都不盡相同，各有各的特性，因此系統可以找出每個人不同而只適合某個人的推薦，這使得採用內容導向過濾的推薦系統可以更具備個人化的功能，推薦的東西更能符合個人的需求。

所以在本論文提出的個人化圖書館館藏推薦系統中，我們結合了協力式過濾和內容導向過濾來產生推薦。利用協力式過濾？使用者收集其他人的意見，幫助使用者篩選圖書館的館藏，讓使用者可以更快速地找到有興趣的、甚至是以前從來沒想到或沒接觸過的館藏。內容導向過濾則是用來進一步篩選之前以協力式過濾找出的推薦清單，藉以提高推薦的準確度和使推薦具有個人化的功能；並且因為採用內容導向過濾的方法，所以也能解決協力式過濾無法有效推薦新書[20]或冷門書的問題 (Cold-Start) [25]。

本論文提出的個人化圖書館館藏推薦系統的目標是希望能依據使用者的借閱歷史來找出他們潛在的興趣，並依照興趣來替每個使用者推薦各自適合的館藏。期待經由這樣一個新的圖書館服務，能吸引讀者更有效利用圖書館的資源，增加圖書館資源的利用率，並且提高圖書館的價值。

第三節 論文架構

在這篇論文中，第二章將會介紹目前推薦系統發展的情況和推薦系統目前常應用到的技術。第三章會把個人化圖書館館藏推薦系統所用到的協力式過濾和內容導向過濾分別做詳細的介紹，並說明如何結合這兩種方法來提升推薦的準確度。第四章介紹交通大學浩然圖書館的 myLibrary 上個人化圖書館館藏推薦系統實作的狀況。第五章則是評估推薦系統效能，計算讀者對這個推薦系統所推薦的館藏滿意度。第六章為結論及探討個人化圖書館館藏推薦系統未來可能發展的方向。

第二章 相關研究工作

推薦系統由於電子商務 (e-Commerce) 的發展而逐漸受到重視，各式各樣的推薦系統為因應各種不同的需要而產生。目前推薦系統新的研究方法是加入資料探勘的技術，其中最常用的是以資料探勘中關聯規則探勘 (Association Mining) 的方法找出交易集 (Transaction Set) 中隱含的知識 (Knowledge)、有用的資訊 (Useful Information) 和關聯規則 (Association Rules)，例如顧客與顧客間的關係、產品與產品間的關係或顧客與產品間的關係等。圖 1 為這方面相關研究的文獻，主要分成推薦系統相關和關聯規則探勘相關這兩部分。其中，推薦系統中的 Law01[8]和 LAR02[9]及關聯規則中的 Hong99[17]、Han00[16]和 Tai02[28]這幾篇為本論文中特別有參考到的文獻。

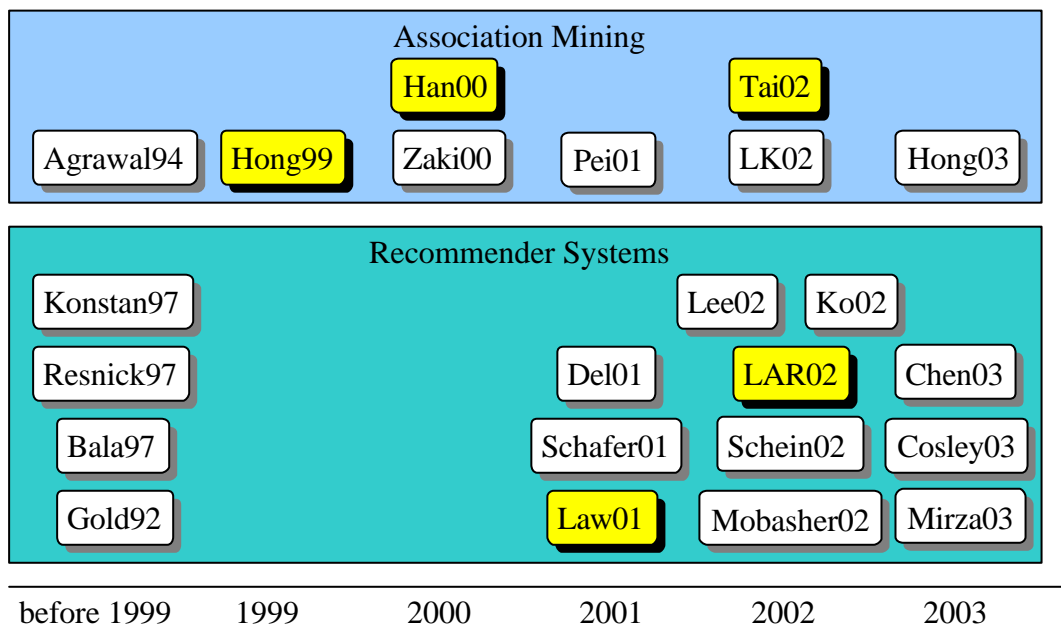


圖 1 相關研究文獻

在本章中，我們將介紹目前推薦系統發展的概況。第一節中會先介紹推薦系統與電子商務間的關係。第二節說明一般推薦系統的主要架構及設計考量。第三節和第四節分別描述推薦系統常用到的協力式過濾 (Collaborative Filtering) 和內容導向過濾 (Content-Based Filtering) 這兩種推薦方法。第五節介紹國立交通大學浩然圖書館中目前已有且屬於推薦系統方面的功能。

第一節 推薦系統與電子商務

推薦系統目前應用最廣的領域在於電子商務上，主要用來提升消費者回店消費的機率以增加其營業額。

由於近年來網際網路的興起，帶來了網路零距離的概念和前有未有的便利性，網路逐漸成為另一個人們用來交換訊息和溝通的方式。尤其在網路技術成熟和基礎建設完成之後，網路愈來愈深入人們的生活，使得愈來愈多的人開始學習使用網路，並且逐漸將生活帶到網路上。在此同時很多商業行為也開始在網路上進行，各種電子網路商店林立，消費者只要輸入網址馬上就被帶到網路商店裡，不但方便而且迅速，這創造了一個新的且龐大的商機。網路不但縮短了消費者與商店間的距離，也縮短了消費者口袋和業者口袋的距離。

網路市場這一塊大餅很快就被注意到並且激烈競爭，因為網路商店的設立成本低，所以許多實體商店紛紛成立網路商店，增加一個窗口服務顧客同時多一個收入來源；甚至更多的網路商店是沒有實體商店的純網路商店，只要少少的資本就可以在網路上租個網頁空間做起生意來。而且網路商店因為成本低，所以有辦法壓低商品價格來跟實體商店競爭，加上網路商店具有時間和空間上兩個最大的優點：在時間上的優勢是網路商店可以 24 小時全天候甚至也是一年 365 天全年無休的營業。在空間上的優勢是其消費群涵蓋全世界任何有網路的地方。所以在任何時間、任何地點，消費者只要上網就可以購物。這些優點使得網路商店同時受到業者和消費者的喜愛。

不過網路商店雖然有那麼多的好處，但對於顧客和業者也各有其特有的缺點。對顧客而言，網路交易的安全性是其在網路上消費的一個障礙；對業者而言，則是網路商店太多、競爭太激烈，而且消費者的流動率很高。在這麼多的網路商店裡，如何吸引消費者來你這家商店消費，甚至持續消費就是個重要的議題。而且實體商店和網路商店有一點重要的不同在於網路商店沒有店員，沒有店員固然減少了人事上的費用，可是當顧客

上門時沒有店員幫忙介紹商品、推銷商品、回答顧客的問題和給與顧客建議，商店和顧客之間沒有互動，這無形中就減少了很多交易的機會。如果網路商店也僱請店員在網路上跟顧客互動，說不定所花費的人事費反而比實體商店高。因為店員必須隨時在線上待命，且網路上的客源數量不定，店員的數量難以拿捏。面對這個問題，其中一個解決的方法就是建立一套具個人化的推薦系統，個人化的推薦系統可以取代店員部分的功能，其中最重要的功能就是推銷商品、給予顧客購買上的建議，因此推薦系統逐漸被應用在網路商店上，幫助網路商店的經營。

亞馬遜網路書店 (Amazon.com) 的執行長 Jeff Bezos 曾經說過：『假如我在網路上有三百萬個顧客，那我就應該有三百萬個網路上的商店。』(“If I have 3 million customers on the Web, I should have 3 million stores on the Web”) [24]，這句話突顯了網路商店個人化的重要性。因為在實體商店中，當顧客找不到需要的商品或對商品有疑問時，都有店員會給予協助，店員可以把店中符合顧客需要的商品展示給顧客選擇，並協助顧客完成交易。但在網路商店中沒有店員的協助，當顧客瀏覽幾頁網頁找不到需要的商品時，馬上會連上其他網路商店的網頁購買，顧客和交易就這樣輕易地流失了。所以每一家網路商店都希望能把顧客需要的商品放在網頁上容易查詢到的地方，讓顧客覺得整家商店裡的商品都是他所需要的，彷彿這家商店是專為他所開設的一般。個人化 (Personalization)，又稱客製化 (Customization)，剛好可以滿足這個要求。建立一個個人化的資訊環境 (Personalized Information Environment) 可以收集顧客的喜好 (Preferences) 及習慣，有助於商店了解顧客的需求。當顧客留下愈多的個人資訊，商店就愈能提供符合顧客需求的商品和服務以吸引顧客下次再度光臨，藉以提高顧客的回店率和忠誠度。顧客一旦再度光臨又將會留下更多的個人資訊，讓商店更能滿足顧客的需要。如此可以提高顧客與商店間的互動，之間形成一個良性循環。所以現在網路商店的一個重要資產就是顧客的個人資料和興趣，誰掌握最多的顧客資訊，能提供最符合顧客需要的商品，誰就掌握了客源。

所以推薦系統和個人化的功能常被運用在網路商店上，以提高業績收入。推薦系統也逐漸加入個人化的功能，希望推薦出來的商品能更符合顧客個人的需求。目前比較有名且有個人化推薦系統的網路商店有 Amazon.comTM (<http://www.amazon.com>)、CDNOWTM (<http://www.cdnw.com>) 和 eBay.comTM (<http://www.ebay.com>) 等。

目前應用於網路商店上的推薦系統通常可分成兩種[11]。第一種是推薦購買頻率較高的商品，例如書或影片之類，這些商品因為顧客購買的頻率較頻繁，所以主要用來推薦的方法是由分析顧客購買的歷史紀錄，進而來推斷顧客的興趣、喜好，再由此來推薦顧客喜歡或感興趣的商品。而這種推薦方式主要的技術在於分析顧客的興趣，通常用的方法是資料探勘並且加入個人化的功能。

第二種系統則主要是推薦顧客購買頻率較低的商品，例如汽車、電腦、家庭劇院組合之類的商品。針對通常購買這類商品的顧客，較無法得知其之前長期以來的興趣，而且往往購買這些商品時都有顧客當時的特殊原因，如果採用以前的購買記錄來做推薦往往無法達到預期的效果。所以這類推薦系統主要是用事先設定好的一些專家知識，依據顧客購買當時的特殊需求，來推薦顧客當時最適合的商品。例如購買電腦時，對一個需要做 3D 繪圖、影像處理的顧客，我們必須推薦一台有高階中央處理器和顯示卡的電腦，而對一個主要用電腦來上網、做一些簡單的文書處理的顧客，則我們須推薦穩定性高或價格較平價的電腦比較合適。在推薦這種購買頻率較低的商品時，主要採用專家系統 (Expert System) 的技術。

根據圖書館的特性，圖書館館藏的流通相當頻繁，而且來借閱圖書館館藏的讀者通常都是有長期的借閱習慣，所以比較接近於“購買頻率較高的商品”這一類的屬性。也因此在本論文中我們主要用資料探勘的技巧來分析處理讀者的借閱歷史紀錄，並判斷讀者的興趣所在，推薦給讀者感興趣的館藏資料。

第二節 個人化推薦系統

推薦系統隨著電子商務而蓬勃發展，有相當多的研究和各式各樣的實作系統，同時也有很多推薦系統已經實際在線上運作了。在[24]這篇論文中整理出了一般電子商務上的推薦系統架構，主要把推薦系統分成三大部分，分別為輸入和輸出 (Functional I/O)、推薦方法 (Recommendation Method) 和其他設計考量 (Other Design Issues)，依序介紹如下：

2.2.1 輸入和輸出

首先，第一部分的輸入和輸出指的是系統需要哪些關於顧客的資訊，以及推薦結果如何呈現給顧客。在輸入這方面包含目標顧客的輸入 (Target Customer Inputs) 和社群資訊的輸入 (Community Inputs) 這兩種。

目標顧客的輸入指的是顧客的興趣輸入，從顧客特意或不經意的情況下，系統可以取得顧客的興趣。有幾種方式可以得到關於顧客興趣的資訊：

1. 直接詢問顧客興趣 (Explicit Navigation)：當顧客在網路商店上瀏覽時，直接要求顧客填入或選擇他們有興趣的商品類別，明確指出他們的興趣所在。通常是在顧客加入會員時，會要求顧客填入一些自己的興趣。
2. 網頁瀏覽記錄 (Implicit Navigation)：由顧客瀏覽過的商品網頁有助於得知顧客感興趣的商品種類。所以分析顧客在網站上的網頁瀏覽記錄，也可以判斷出顧客的興趣。
3. 查詢關鍵字 (Keywords and Item Attributes)：顧客查詢商品時所使用的關鍵字或商品種類。
4. 商品評分 (Rating)：顧客對使用過或購買過的商品的評價。

5. 購買記錄 (Purchase History)：顧客在商店裡曾購買過的商品和時間。

這些顧客所留下來的資訊，都透露著顧客的興趣和商品偏好。善用這些資訊，則推測出來的使用者興趣將更貼近使用者的真正興趣。

社群資料的輸入指的是一般大眾對商品的喜好，在目標顧客的輸入中主要目的是取得顧客的資訊，而社群資料的輸入則是想取得商品的屬性資訊。社群資料包含以下幾種：

1. 商品屬性 (Item Attribute)：包含商品所擁有特定的顧客群，和商品所屬的分類。
2. 熱門商品 (External Item Popularity)：銷售量較高的商品，通常為一般大眾普遍喜歡或在某些時節有特殊需要的商品。
3. 社群購買記錄 (Community Purchase History)：由結合所有個人的購買記錄所組成，可以找出商品銷售的趨勢和商品間的相似度，也是一種隱性的商品評分。
4. 商品評分和心得感想 (Rating and Text Comments)：社群對商品的喜好程度及其評論，有助於突顯商品價值。

收集個人和社群的輸入資訊加以分析，對了解顧客的興趣和分析商品的屬性有很大的幫助，這些收集來的相關資訊都有助於使推薦系統能做出最合適的推薦。

關於輸出部分，指的是要如何把推薦的結果呈現給顧客，通常是推薦系統中最後的步驟，但也是最關鍵性的步驟。因為如果沒有用適當的方式把推薦呈現給顧客，那可能會使顧客沒機會接觸到這些推薦，即使顧客得到推薦也會影響顧客對這些推薦的印象，使得推薦無法達到預期的功效。推薦的輸出會根據商品不同的性質 (Type)、推薦的數量 (Quantity) 和呈現給顧客看的形式 (Look of Information) 而有所不同。通常輸出的方式可分為建議 (Suggestion) 和預期 (Prediction) 這兩種。建議是指給顧客一個未排序過的推薦清單，讓顧客檢閱其中是否有符合需要的商品，避免因為清單中前幾個推薦不滿意而使後面的推薦不被參考。而預期這種方法所給的推薦清單則是有排序過的推薦，甚至會列出系統預期顧客可能的喜好程度，使顧客知道各個商品的推薦強度。

2.2.2 推薦方法

第二部分推薦方式指的是如何找出顧客感興趣的商品並列出推薦清單。推薦方法是推薦系統的核心，方法的好壞直接影響了推薦結果的優劣。而因為需要的不同發展出多種不同的推薦方法，常利用的方法如下：

1. 基本檢索 (Raw Retrieval)：網路商店上通常有基本的商品搜尋引擎，讓顧客輸入關鍵字來找尋有興趣的商品。搜尋引擎可由顧客輸入的關鍵字列出所有相關的商品，但通常只是純粹的字串比對，所以在查全率 (Recall) 和準確度 (Precision) 上都有所不足。
2. 人工選粹 (Manually Selected)：由專家以人工方式建立推薦清單或分類清單，使顧客可以依照自己的興趣選擇需要的清單參考。這種方式有相當高的準確度，可是需要長期人工維護更新，無法及時提供資訊。
3. 統計分析 (Statistical Summaries)：以統計的方法分析一般顧客的購買趨勢來做推薦，也可以整理出熱門商品清單讓顧客參考。
4. 屬性導向 (Attribute-Based) 推薦：利用商品的屬性和顧客的興趣，比較之間的關係，依此來判斷出顧客可能需要或喜歡的商品。
5. 商品間的關聯性 (Item-to-Item Correlation)：商品間常會有某些關係存在，例如可以找出常被一起購買的商品，當顧客購買某樣商品時，可以推薦給他其它常被一起購買的商品。
6. 顧客間的關聯性 (User-to-User Correlation)：有些顧客因為購買習慣或興趣相同，可以被歸類在同一個社群中。同社群內的顧客彼此之間可以分享資訊，互相推薦商品。

這些推薦方法並無絕對的好壞，依照需求不同而各有存在的價值。這些方法除了可以獨立運作外，也可以互相搭配使用來提升推薦的效果。例如在[8]中的推薦方法結合了

屬性導向推薦和顧客間的關聯性兩種方法，首先利用顧客間的關聯性來找出顧客可能感興趣的推薦清單，再利用屬性導向推薦來對推薦清單排序，這樣找出來的推薦清單不但參考了社群的意見，也相當符合個人的興趣。而在[9]中則提出了結合商品間的關聯性和顧客間的關聯性的推薦方法以互補這兩種方法的缺點，當要對某位顧客做推薦時，如果找不到相同興趣的顧客，無法用顧客間的關聯性做推薦時，就轉換成用商品間的關聯性，從顧客的購買紀錄中判斷哪些商品顧客目前可能感興趣。

目前推薦系統在電子商務上的應用，比較偏重於個人化的功能，希望能根據每個使用者不同的興趣分別給予使用者符合興趣的推薦。個人化推薦系統主要的方法分成兩種，分別為協力式過濾和內容導向過濾。而在之前提到的方法中，商品間的關聯性和使用者間的關聯性是協力式過濾推薦常用到的方法，而屬性導向推薦其實就是內容導向過濾推薦的別名。協力式過濾和內容導向過濾這兩種方法將分別在本章第三節和第四節中介紹。

2.2.3 其他設計考量

最後，設計一個推薦系統還必須考量到一些特別需求，依照各個網路商店不同的需要分別做不同的設計，通常考量到的主要有個人化的程度 (Degree of Personalization) 和如何把推薦送達 (Delivery) 給顧客這兩個方面。這些設計上的考量甚至會直接影響推薦方法的選擇。

個人化的程度方面可分成三種，主要因為產生推薦的方法不同而有所差異。第一種是無個人化 (Non-Personalized)，例如只用一些統計摘要的方法所產生的推薦，每個人所得到的推薦都是相同的。第二種是短期的個人化 (Ephemeral Personalization)，這種個人化主要是依據顧客最近的網頁瀏覽紀錄或選取了哪些商品，來推薦可能需要的商品，這種方式大多採用內容導向過濾或利用資料探勘所找出的物品間的關係來做推薦。第三種則是長期的個人化 (Persistent Personalization)，這種方式會長期追蹤顧客的興趣，主

要的推薦方式是利用顧客的歷史紀錄檔找出使用者和使用者間的關係，再利用協力式過濾推薦。

而如何把推薦清單送達給顧客也是一個重要的議題，主要分成主動送出 (Push)、等候選取 (Pull) 和被動產生 (Passive) 三種。主動送出顧名思義是主動將推薦送給顧客，通常採用電子郵件的方式。主動送出的優點是是可以直接且及時地把訊息送給消費者，吸引顧客光臨，但也可能會被當成廣告信，引起顧客反感。等候選取則是可由顧客自己決定何時要查看推薦。顧客可在瀏覽網站時，可自行點選檢視商店所推薦的商品，讓顧客享有絕對的自主性。等候選取的困難在於需要引起顧客的興趣來點選瀏覽，所以往往這種方法在應用上造成的效果不彰。第三種被動產生的方法則是當顧客的購物車裡已經有某些商品時，可以根據這些商品來做即時的推薦。被動產生的方法因為直接跟顧客互動，往往推薦的東西和顧客當時的需要接近，所以應用上的效果不錯。但是因為要線上作業，所以特別要考量推薦執行的速度，避免為了推薦而影響顧客瀏覽網頁的流暢度，使顧客失去耐心。當然以上這三種方法可以互相搭配使用，主要還是依照商品的特性、顧客群習慣和實際施行的效果。

第三節 協力式過濾與關聯規則探勘

協力式過濾可以說是推薦系統的前身，1992 年首先由 Goldberg 在[14]中提出，應用於電子郵件分類系統 TAPESTRY 上。協力式過濾主要的想法是使用者之間可以利用彼此已知的資訊，互相幫忙過濾出自己有用的資訊，就好像在我們做決定之前常會事先詢問有經驗的人意見來當做決定的參考一般。協力式過濾在實作上常利用的方法是建立一個顧客 - 商品的矩陣 (Customer-Product Matrix)，或稱為評分矩陣 (Rating Matrix) [6]，矩陣中的值為顧客對商品的評分。從這個矩陣中可以得到顧客間的興趣相似度，藉著顧客間的相似度和其他顧客對商品的評分可以預測出顧客對未評分商品的喜好程度。

而協力式過濾推薦近來常利用關聯規則探勘 (Association Rules Mining) [1] 的方法, 找出有用的關聯規則來當作推薦的依據。例如在[9]中就主要以關聯規則探勘所找出的物品之間的關係 (Item Association) 及使用者間的關係 (User Associations) 這兩種關聯規則互相搭配來做推薦。物品間的關係可以表現出哪些商品常常被一起購買, 這些商品之間可能隱含著某種關係而造成常被一起購買的現象, 因此當顧客購買其中某樣商品時, 我們可以推薦給他其它也常常一起被購買的商品。而使用者間的關係則是要找出購買習慣類似的顧客群。利用這些購買習慣類似的顧客群, 當要對群中某個顧客做推薦時, 系統可以推薦給他同一群中其他人常購買的商品。例如若我們發現一群顧客常常購買古典樂 CD, 如果某天這群顧客中的某位來購買 CD, 那我們可以將其所屬的顧客群中其他人所購買過的古典樂 CD 推薦給他。

關聯規則探勘是資料探勘中的方法之一, 經常運用於商品交易記錄資料庫上, 針對使用者的交易行為做分析, 找出商品間的關聯規則。再根據這些關聯規則來決定搭配促銷商品和商品架位等行銷策略, 藉以提高交易量和營業額。例如: 80%購買牛奶和果醬的顧客也會同時購買吐司, 就是一個典型的關聯規則。從 1993 年 Agrawal 提出從交易資料庫中探勘出關聯規則的演算法[1]後, 陸續有學者將關聯規則的概念應用到其他領域, 提出適用於該領域的演算法。

關聯規則探勘主要分成兩個步驟: 一、找出頻繁項目集 (Frequent Itemsets, Large Itemsets)。二、由頻繁項目集中歸納出關聯規則。由於第二個步驟, 歸納出關聯規則, Agrawal 已經在[2]中提出有效率的演算法解決了, 所以目前關聯規則探勘效率的瓶頸在於第一步的找出頻繁項目集。早期的演算法找出頻繁項目集時常需要大量時間來產生和檢查候選項目集 (Candidate Itemsets), 這造成探勘的效率低落。因此? 了提高探勘的效率解決這個問題, 不需產生候選項目集的演算法便因應產生。

後來被提出的演算法大多主要著重在探勘效率的提升上, 所以關聯規則探勘可粗略依照是否產生候選項目集來區分成兩種[16]。需要產生候選項目集的有 Apriori [2]、

Pincer-Search[10]和[26]提出可處理大量資料的演算法 (Scalable Algorithms) 等，而不需產生候選項目集的演算法有 FP-growth[19]和 H-mine [22]等。另外在[17], [18]中提出乏析資料探勘 (Fuzzy Data Mining), 專門處理量化 (Quantitative) 的資料。[15]則以 AprioriTid 演算法來實作乏析資料探勘。

第四節 內容導向過濾與系統模型建立

內容導向過濾的推薦方式，主要是從資訊擷取 (Information Retrieval) 的領域延伸過來。這種方式會以物件的內容或屬性來替每個物件建立屬性向量，在判斷物件是否相關時，常會以計算物件其向量間的餘弦 (Cosin) 值，也就是兩向量間的夾角大小來代表相似度 (Similarity)，如果代表這兩個物件的向量之間夾角小，則代表這兩個物件相關程度高，反之則相關程度低。若以顧客和商品間的關係來看，我們可以把顧客和商品分別用一物件表示，兩個物件間的向量夾角大小就是顧客對商品可能的喜好度。

不過屬性導向的方法最大的困難在於如何建立一個模型 (Model) 來表示系統中所有的物件。在[8]這篇論文中，提出了一種模型來替商品和顧客建立分別其屬性向量表示式和興趣向量表示式。首先將所有的商品分成數類，建成一個有階層的分類樹，每個商品屬於分類樹分類中的其中一類。假設最底層的分類有 1000 類，則這 1000 類事實上也就是屬性向量中的 1000 個屬性，而每個物件其屬性向量中的值代表著它在各分類中的權重。？ 了要分析分類樹中每個分類之間的關係，所以利用關聯規則探勘找出分類樹中每一層中分類之間的關聯規則。利用這些關聯規則，和權重給定公式，可以？ 每一個商品定義出其屬性向量表示式。

假設有一個分類樹如圖 2，第一層有 A, B, C 三個分類，而每個分類都有三個子分類，所以第二層有九個分類，而所有系統中的物件都會以具有九個屬性的向量來表示，而這九個屬性分別代表 A.1~C.3 這九個分類，各個屬性的值也就代表物件在各個分類的關係權重。

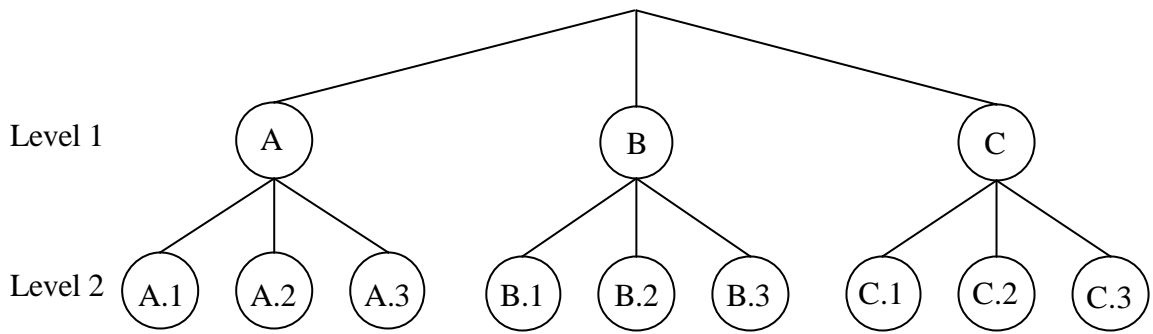


圖 2 一個兩層的分類樹

之前顧客的購買記錄可以利用關聯規則探勘找出各階層分類間的關聯規則。假設找到 $A \Rightarrow B$ 和 $A.2 \Rightarrow C.2$ 這兩條規則，而[8]定義的權重給定公式如下：

$$P_s^{(n)} = \begin{cases} 1.0 & \text{if } s = S(n) & (\text{within same subclass}) \\ 1.0 & \text{if } S(n) \Rightarrow s & (\text{within associated subclass}) \\ 0.5 & \text{if } C(s) = C(n) & (\text{subclass within same class}) \\ 0.25 & \text{if } C(n) \Rightarrow C(s) & (\text{within subclass of associated class}) \\ 0 & \text{otherwise} \end{cases}$$

方程式 1 文獻[8]中的權重給定公式

其中 $P_s^{(n)}$ 代表商品 n 在第二層分類 s 上的權重, $S(n)$ 代表 n 所在的第二層分類, $C(s)$ 和 $C(n)$ 代表第二層的分類 s 和商品 n 所在的第一層分類。

從公式中我們可以知道如何去找出一個商品的屬性向量。首先將商品 n 所在的分類 $S(n)$ 之權重設為 1；如果有關聯性規則表示可以從 $S(n)$ 推廣到其他的分類 s ，則 n 在這些分類 s 的權重也是 1；此外如果商品 n 和分類 s 都同屬於某個第一層的分類下，則 n 在這個分類 s 下的權重為 0.5；如果商品 n 和分類 s 不屬於某個第一層的分類下，但是商品 n 所在的第一層分類 $C(n)$ 和分類 s 所在的第一層分類 $C(s)$ ，有 $C(n) \Rightarrow C(s)$ 這樣的關聯規則存在，則 n 在這個分類 s 的權重設為 0.25。至於其他的分類， n 在這些剩下分類中的權重皆設為 0。

以一個屬於 A.2 這個分類的商品 k 為例，建立 k 的屬性向量之步驟如下：

1. 一開始 k 的表示式是一個具有九個屬性的向量，其中每個屬性的值都為零。

$$k = (0, 0, 0, 0, 0, 0, 0, 0, 0)$$

2. k 屬於 A.2 這個分類，所以 k 在 A.2 分類上的權重設為 1。

$$k = (0, 1, 0, 0, 0, 0, 0, 0, 0)$$

3. k 所在的分類 A.2 有一條關聯規則為 $A.2 \Rightarrow C.2$ ，所以 k 在 C.2 分類上的權重也設為 1。

$$k = (0, 1, 0, 0, 0, 0, 0, 1, 0)$$

4. A.1 和 A.3 跟 k 都屬於 A 這個第一層分類之下，所以 k 在 A.1 和 A.3 這兩個分類上的權重都為 0.5。

$$k = (0.5, 1, 0.5, 0, 0, 0, 0, 1, 0)$$

5. 又 k 所在的第一層分類 A 有一條關聯規則為 $A \Rightarrow B$ ，所以 k 在第一層分類 B 之下的分類 B.1、B.2 和 B.3 這三個分類上的權重都可設為 0.25。

$$k = (0.5, 1, 0.5, 0.25, 0.25, 0.25, 0, 1, 0)$$

所以最後商品 k 的向量表示式為 $(0.5, 1, 0.5, 0.25, 0.25, 0.25, 0, 1, 0)$ ，代表商品 k 的屬性。將商品的表示式找出來之後，可據以建立顧客的表示式，最簡單的方式就是依照顧客的購買記錄統計顧客消費過的商品，依照這些商品的表示式來建出顧客的興趣向量表示式。之後可以利用這個模型，透過顧客和商品分別的表示式，來推測顧客對商品可能的興趣程度。

第五節 圖書推薦系統

圖書館中事實上已經有許多推薦系統以各種形式存在，帶給讀者許多的便利性。以下我們將以國立交通大學浩然圖書館為例，來舉出幾種目前圖書館中存在的推薦系統。

2.5.1 關鍵字搜尋

關鍵字查詢是一般圖書館使用率最高的功能，通常使用者利用這個功能來找尋圖書館中需要的館藏。圖 3 是浩然圖書館館藏查詢系統的使用者介面，其中提供有書刊名、標題、作者、出版商等多達 11 種欄位來供使用者填入關鍵字來查詢。查詢的結果會呈現出所有含有使用者輸入關鍵字的館藏，這些回傳的資料就相當於是查詢系統給予使用者的推薦館藏，如圖 4 為以書刊名關鍵字“圖書館”的查詢結果。

但是關鍵字查詢通常只是單純的字串比對，若使用相同的關鍵字查詢，找出來的資料都是一樣的。而且萬一使用不恰當的關鍵字搜尋，往往找到的資料不是過多就是過少，無法達到讀者預期的需要。

圖 3 顯示了國立交通大學圖書館館藏查詢系統的使用者介面。該介面包含一個搜尋欄位，用於輸入關鍵字，以及多個篩選條件，如書刊名、標題、作者、出版商等。此外，還提供了資料類型和排序方式的選擇。介面底部列出了系統首頁、館藏查詢、個人資料查詢、新書通報、指定參考書查詢、讀者意見、公佈欄、推薦書刊、分類號查詢、開放時間、MyLibrary 個人化圖書資訊環境、其他圖書館館藏查詢、冊刊館藏查詢等服務。

圖 3 交通大學浩然圖書館館藏查詢使用者介面

#	書刊名(按此HyperLink可以查館藏) TITLE	作者/出版商 AUTHOR/PUBLISHER	書目索書號 CALL NO	出版年 PUBLISH YEAR	資料來源 TYPE
1	21世紀公共圖書館之探索	王元仲/聯經	026.07 1012	2000	一般書/BOOK
2	750種方法讓地球更乾淨	MacEACHERN, DIANE/李麗潔/台灣地球日	412.7 2265 v.9	1993	一般書/BOOK
3	Attention!注意力行館	肯·薩可瑞/岳心怡/商周	497 0265 v.12	2002	一般書/BOOK
4	Clone傳奇揭開生命複製的神祕面紗	韓王榮/龔靜/業強	363.019 4419	1998	一般書/BOOK
5	Excel 2000圖說圖書館	程錦/第三波	312.949E9 2686-2	1999	一般書/BOOK
6	Proceedins of the International Conference on National Libraries-Towards the 21st Century(April 20-24, 1993: Taipei)-邁向21世紀的國家圖書館國際研討會論文集	National Central Library/Proceedins of the International Conference o/National Central Library ROC	2674.83 157 1993	1993	一般書/BOOK
7	Tokyo book map : 東京23區書店圖書館簡版/	Tokyo Bukku Mappu Henshu linkai. /書籍情報社	2463.86.T6 T 618 LCC		
8	一五一答客問:以圖書館為主的文化活動	台中國書館/台中國書館	023.6 2565-1	1984	一般書/BOOK
9	二十一世紀公共圖書館發展研討會實錄	王維昭/台灣省立台中國書館	026 1026	1998	一般書/BOOK
10	二十一世紀公共圖書館新論-The Public Library of the 21st Century	姓名/吳建中/胡歐蘭/華文圖書館管理	026 2615	2002	一般書/BOOK

圖 4 交通大學浩然圖書館館藏查詢結果呈現介面

2.5.2 主題式人工推薦

當圖書館中某類館藏資料多到一定的程度，並且有專家的協助，有時候會把這類的館藏以主題分類陳列。如圖 5 是浩然圖書館中一個以科幻圖書為主題的主題式推薦，在這個例子中，主要的推薦是左方已經分門別類過的館藏，對科幻圖書有興趣的人可以很方便地找到各種分類的科幻圖書。而圖 6 是點選左邊的“時光旅行”項目後出現的推薦清單。圖 5 右方的查詢事實上仍是關鍵字查詢，只是在這裡的查詢會把搜尋類別侷限在科幻圖書上。

主題式人工推薦最大的好處就是呈現出來的推薦準確度相當高，加上已經有經過專家的分類，讀者可以很容易取得符合自己興趣的館藏，但缺點是需要人工長期維護以保持更新。

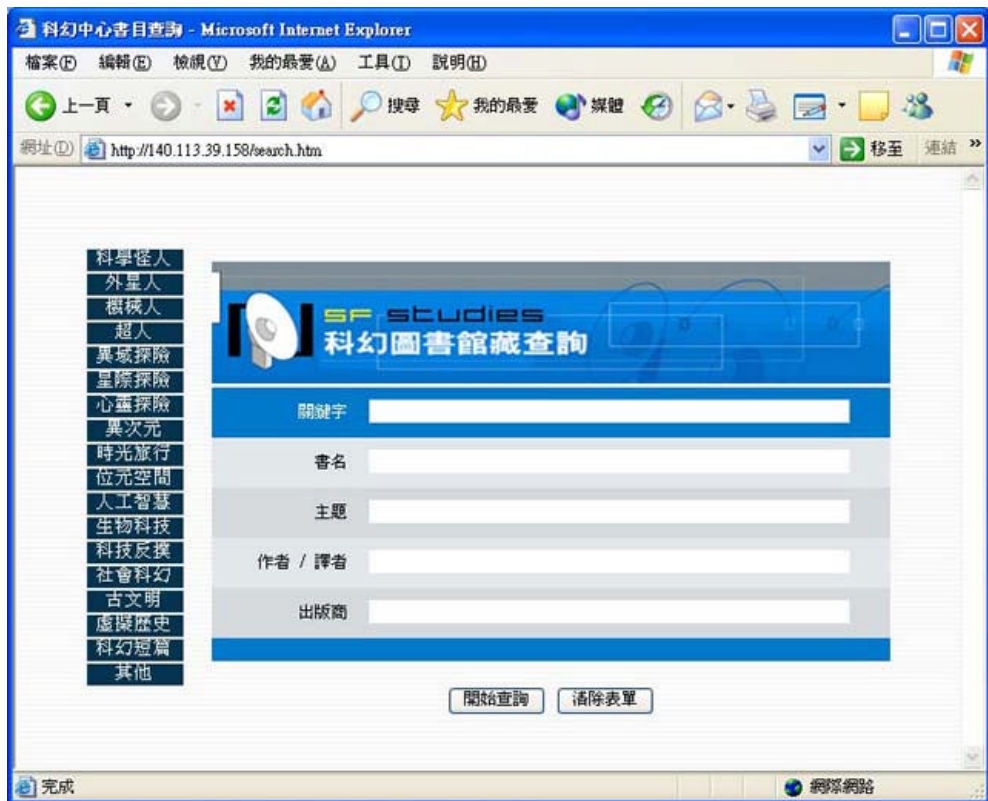


圖 5 科幻圖書館藏的主題推薦

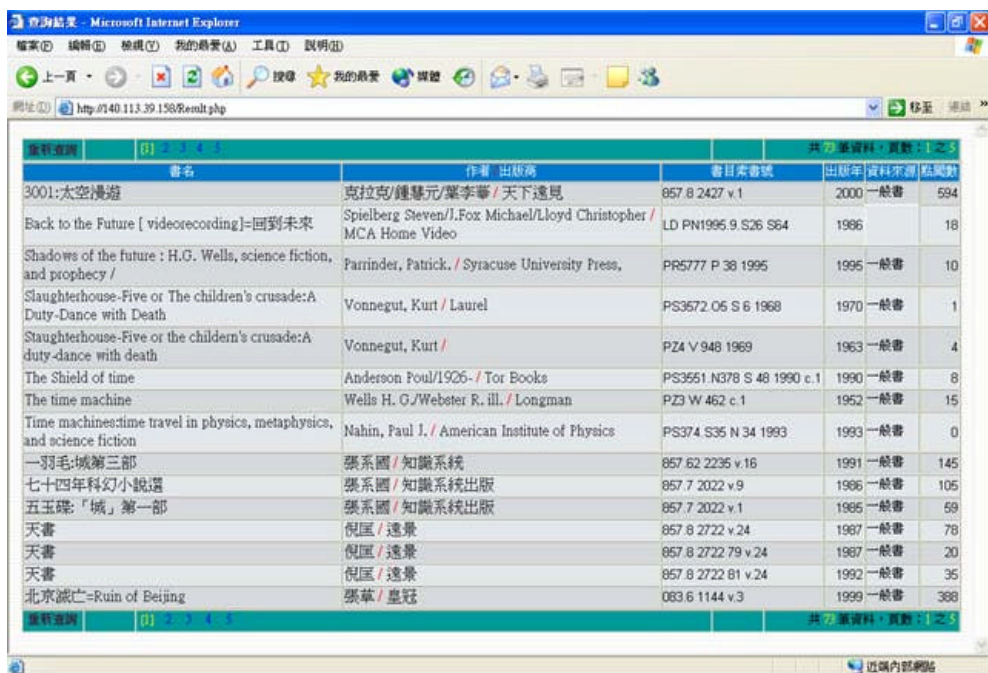


圖 6 科幻圖書館藏中的分類推薦 - 以“時空旅行”分類為例

2.5.3 館藏間的關連性

資料探勘的技術已經普遍應用在各領域，主要用來找尋大量資料中的規則性和關聯性。現在這種技術也逐漸引入圖書館中，在[28], [29]這兩篇論文中就提出以資料探勘的方法來找出館藏間的關聯性，找出哪些館藏經常會被讀者一起借閱。然後當讀者在查詢館藏或者瀏覽自己的借閱記錄時，可以查看還有哪些館藏是常常一起被借閱的，藉此來提高館藏的利用率。如圖 7 是以“資料庫”為書刊名關鍵字查詢館藏資料回傳的結果，圖 8 為某人的借閱歷史查詢結果。其中按下圖 7 中最右邊的“推薦”欄位或圖 8 中最右邊的“感想”欄位，就可以看到該本書的關聯館藏，得知有哪些館藏常常跟那本書一起被借閱，如圖 9 為“溫馨廚房咖啡屋”這本書的相關館藏資料。

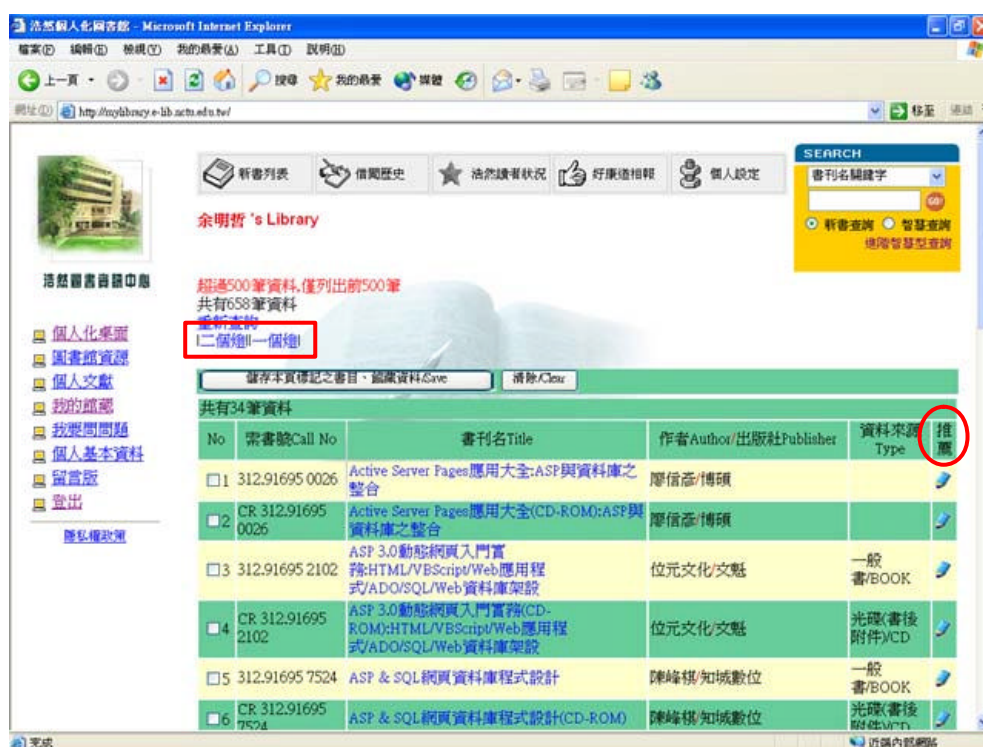


圖 7 個人化智慧型館藏查詢結果 - 以“資料庫”為書刊名關鍵字查詢



圖 8 個人借閱歷史查詢介面

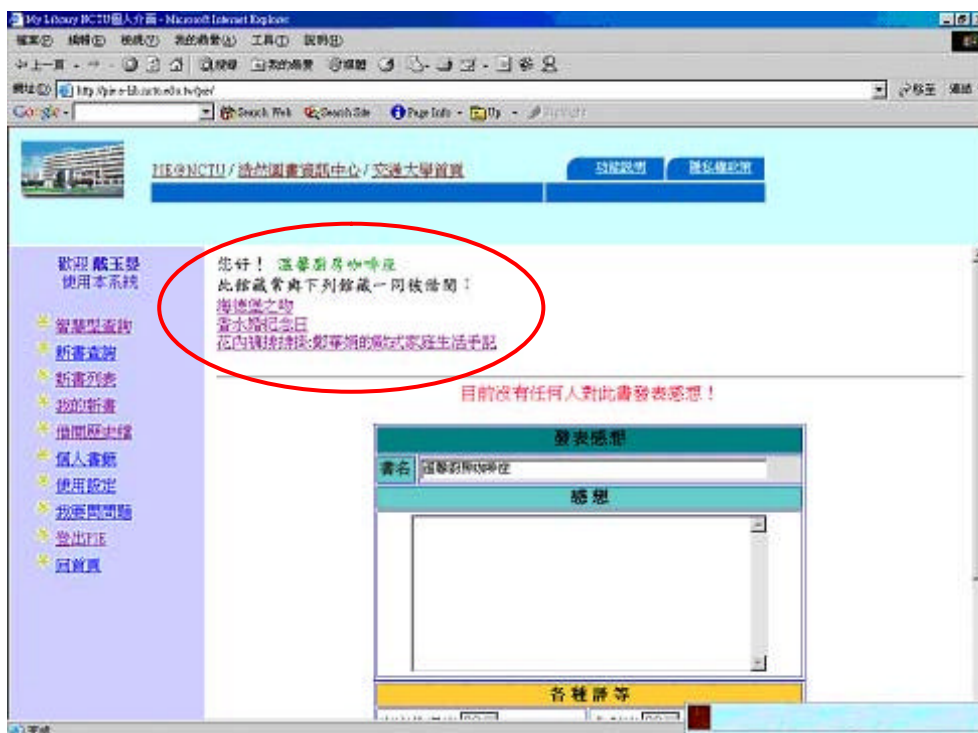


圖 9 推薦相關館藏畫面

2.5.4 個人化的智慧型查詢

在[30]中提到，圖書館面對與日俱增的數位或紙本資源，傳統的搜尋引擎已無法滿足讀者的需求。因此[30]這篇論文將個人化資訊服務技術應用在圖書館上，建立一套個人化數位圖書資訊環境 - PIE@NCTU。

其中 PIE@NCTU 中最重要的改變是將館藏查詢系統採用個人化的智慧型查詢系統 (Intelligent Personal Search Engine) 以契合使用者的需求。這套具個人化的館藏查詢系統會收集使用者搜尋時所用的關鍵字來判斷使用者的興趣，並且依此產生一些系統關鍵字來代表使用者的興趣。另外使用者也可以自己增加關鍵字來告知系統自己的興趣所在，或者自己選定有興趣的館藏類別。當使用者查詢館藏資料時，系統會利用系統判斷關鍵字、使用者自訂關鍵字和使用者自選類別這三種資訊，來判斷使用者最需要的館藏，把查詢結果先加以過濾和排序後再呈現給使用者。

另外一個重要的功能是新書通告 (New Arrival Notice)，當圖書館有新進館藏時，系統可以根據使用者之前留下的個人興趣來判斷哪些新進館藏是使用者可能會感興趣的，並將這些新進館藏的消息通知使用者。

圖 10 為目前交通大學浩然圖書館的個人化智慧型查詢系統介面，查詢結果呈現方式如圖 7 所示，其中上方的“二個燈”和“一個燈”表示和讀者興趣的符合度，燈數愈高代表愈契合讀者興趣，最高為五個燈。

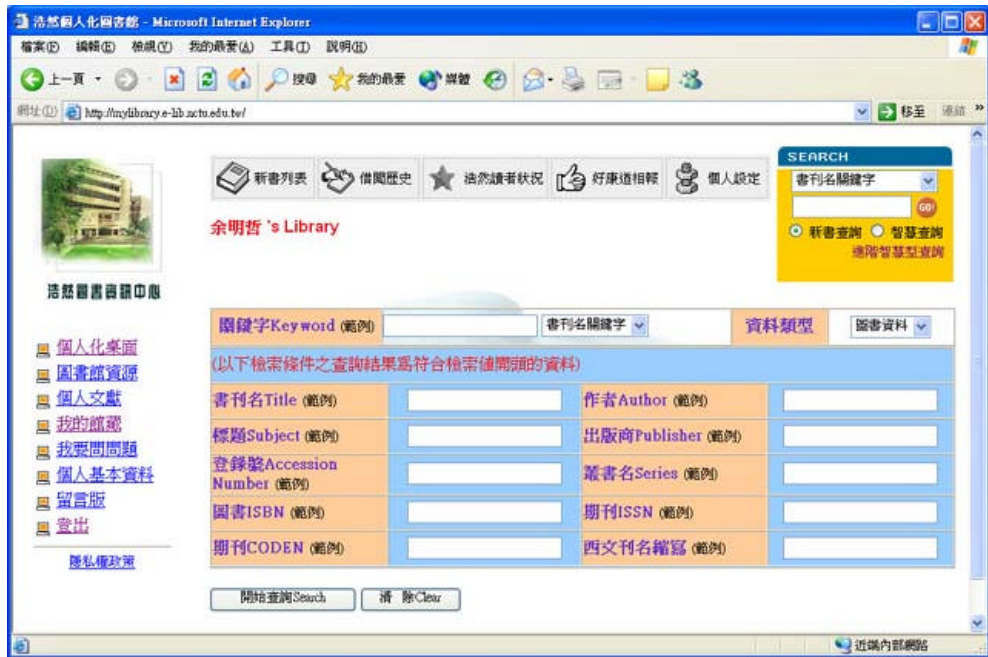


圖 10 個人化智慧型查詢系統介面

以上這幾種圖書館所提供的推薦服務，都能提供給讀者搜尋資料上莫大的幫助。尤其在現在許多圖書館都朝向電子圖書館（e-Library）或數位圖書館（Digital Library）的方向前進，提供給讀者即時且便捷的服務是圖書館的首務。所以必須運用科技進步帶來的新技術和新方法，藉以提升圖書館服務的品質並增加服務的項目，使圖書館跟上時代逐漸成長進化。

第三章 圖書館館藏推薦

本論文圖書館個人化館藏推薦系統中所採用的推薦方法，是結合協力式過濾和內容導向過濾這兩大方法。本章分別介紹這兩種方法如何應用在系統上，並說明如何結合這兩個方法。

第一節 協力式過濾的推薦方法

在進行協力式過濾時，我們首先用關聯規則探勘找出使用者間的關係，希望藉以找出具有相同興趣的使用者群。然後在具有相同興趣的使用者之間，推薦給彼此曾經借閱過的館藏。

我們將每一本書當作一筆交易 (Transaction)，而其中的項目 (Item) 為借過這本書的人，例如現在有 A, B, C 三本書，和 1, 2, 3, 4, 5 這五個讀者。從這五個讀者的借閱歷史檔中，可以建立一個交易集 (Transaction Set)，其中每筆交易表示每本書曾被哪些讀者借閱過。如表 1 中的第一筆資料代表 A 這本書曾被讀者 1, 2, 4, 5 這四個人借閱過，第二筆資料代表 B 這本書曾被讀者 1, 3, 5 這三個人借閱過，其他筆資料以此類推。

書	借閱讀者群
A	1, 2, 4, 5
B	1, 3, 5
C	2, 4, 5

表 1 以一本书為一筆交易的交易集

假設最小支持度 (Minimum Support) 為 2，則從表 1 的交易集經過關聯規則探勘後，可以得到一些項目集 (Itemset)，若項目集的支持度高於預設的最小支持度，即可稱為頻繁項目集。這些出現在同一個頻繁項目集中的讀者群，就是我們想找出來的相同興

趣讀者群。如表 2 所示，在表 1 的例子中可以得到 3 個頻繁項目集，或稱為三個讀者群，其中每一群裡的讀者彼此間具有相同興趣。

群編號	頻繁讀者群	支持度
S001	1, 5	2
S002	2, 4	2
S003	3, 5	2

表 2 代表具有相同興趣讀者群的頻繁項目集

雖然用一般關聯規則探勘可以找出相同興趣讀者群，可是在圖書館的環境下卻遇到兩個問題。第一個問題是在圖書館中，借閱同一本書的讀者在比例上不多，而利用一般關聯規則探勘所找出來的相同興趣讀者群都是要曾借閱過多本相同書的讀者，這造成能找出來的相同興趣讀者群不夠多，只佔全部讀者群中的一小部分。這種結果會嚴重影響到後來的推薦，因為協力式過濾推薦需要從跟讀者有相同興趣讀者群中，找出其他讀者常借閱的書來當推薦，萬一找不到跟讀者有相同興趣的讀者群，就無法？讀者找出適合的推薦。第二個問題是圖書館中大多數讀者借閱館藏的數量並不多，在[29]中曾指出在交通大學浩然圖書館中約有 1/2 的讀者其借閱數少於 10 本，這使得在關聯規則探勘時必須採用非常小的最小支持度，才能找出足夠的頻繁項目集。但是當最小支持度設太低時便需要耗費大量的時間進行探勘，而且也常因此找出不恰當的頻繁項目集。

對於第一個問題，改善的方法是放鬆讀者屬於同一筆交易的限制。之前是以一本書為一筆交易，所以讀者必須借閱過相同的書才會屬於同一個交易中，現在我們改以一個類別為一筆交易，只要讀者曾借閱過這個類別中的任一本書就屬於同一筆交易中，亦即，將原來的交易集中，屬於同一類書的交易聚集起來，形成一個以一個類別為一筆交易的交易集。因為在圖書館中，如果某本館藏被借閱走了，則其他讀者就無法借閱了，這時候通常讀者還是會借閱相同類別中其他相關的館藏。所以在圖書館這種環境下，要判斷讀者是否跟其他讀者有相同興趣，只依讀者間有沒有借閱過相同的館藏來判斷兩人

是否興趣相同，在現實上太過嚴苛了，要是依照彼此之間是否借閱過相同類別的館藏來判斷會顯得合理許多。基於上述理由，我們改變原本以一本書為一筆交易的形式，改成為一個類別為一筆交易，只要借過同一個類別下任一本書的讀者，都視為興趣類似，而可以放在同一筆交易中。

這樣的作法或許會降低最後所找出的相同興趣讀者群的準確性，因為只要是借過同類別書籍的讀者則他們可以放在同一個交易中，而不像之前必須借過同一本書才能放在同一筆交易中。這樣探勘出來的結果其意義是只要讀者常常借閱相同類別的書籍就會被認定興趣相同，而非像之前必須常常借閱同一本的書籍才會被認定興趣相同。這樣其實就是放鬆判斷讀者是否興趣相同的條件，所以可能會使最後出來的頻繁項目集(相同興趣的讀者群)結果不像之前那麼精確。

但，如此做可以使改變後的交易集裡，屬於同一筆交易的讀者數變多，也就是交易中的項目變多了，會有助於後續找出更多的頻繁項目集。因為這樣可以讓讀者更有機會同時存在同一個交易中，而被視為具有相同興趣；使讀者間更容易產生關聯性，探勘時可以找出更多相同興趣的讀者群，使之後系統可以替更多讀者做館藏推薦。

可是若以一個類別為一筆交易，將衍生另一個問題，聚集過後的交易中必須考慮其中項目出現的次數。因為有些讀者在同一個類別中借過多本書，當交易聚集時這些讀者會在新的同一筆交易中出現多次，這些出現的次數必須被保留，否則將造成支持度降低及資訊的流失，違背原先聚集交易的目的。以表 3 為例，若書名第一碼英文字母相同為同一類，聚集後可以分別得到考慮個數和不考慮個數的交易集如表 4，其中考慮個數的交易集中每個項目“:”之後的數字代表個數。

書	借閱讀者群	書	借閱讀者群
A.1	1, 2, 5	D.1	1, 3, 5
B.1	1, 3, 4, 5	D.2	1, 3, 4
B.2	2, 3, 4	E.1	4, 5
C.1	2, 3, 5	E.2	3, 5
C.2	1, 3, 4, 5	E.3	3, 4

表 3 以書籍為交易的交易集

類別	借閱讀者群(不考慮次數)	借閱讀者群(考慮次數)
A	1, 2, 5	1:1, 2:1, 5:1
B	1, 2, 3, 4, 5	1:1, 2:1, 3:2, 4:2, 5:1
C	1, 2, 3, 4, 5	1:1, 2:1, 3:2, 4:1, 5:1
D	1, 3, 4, 5	1:2, 3:2, 4:1, 5:1
E	3, 4, 5	3:2, 4:2, 5:2

表 4 聚集過後以類別為交易分別不考慮次數和考慮次數的交易集

在考慮次數的交易集中，計算項目集在每個交易中支持度的方法為找出項目集中所有項目分別在交易中的出現次數，取其中最少的次數為其支持度。以計算長度為 r 的項目集 s 在交易 D_i 的支持度 \sup_s^i 為例，公式如下：

$$\sup_s^i = \min_{0 < j \leq r} (\sup_{s_j}^i)$$

方程式 2 項目集在每個交易中支持度的計算公式

其中 s_j 代表項目集 s 中第 j 個項目， $\sup_{s_j}^i$ 代表 s_j 在交易 D^i 中的次數(支持度)。例如在考慮個數的交易集中，項目集 $\{3, 4\}$ 於類別 C 中的支持度為 1，其算式如下：

$$\sup_{\{3,4\}}^C = \min(\sup_3^C, \sup_4^C) = \min(2, 1) = 1$$

另外，表 5 將項目集{3, 4}分別在不考慮個數的交易集、考慮個數的交易集和交易聚集前的交易集內的支持度列出。

類別	{3, 4} 項目集出現次數		
	聚集後交易集 (不考慮次數)	聚集後交易集 (考慮次數)	聚集前交易集
A	0	0	0
B	1	2	2
C	1	1	1
D	1	1	1
E	1	2	1
支持度	4	6	5

表 5 項目集 {3, 4} 在各個交易集中的支持度

在表 5 中，可以發現不考慮次數的交易集中，項目集 {3, 4} 得到的支持度為 4，甚至比聚集前交易集中的支持度 5 少。因為不考慮次數交易集的類別 B 中，項目集 {3, 4} 會失去支持度 1，反而造成支持度下降。而項目集 {3, 4} 在考慮次數的交易集中可以得到 6 的支持度，比聚集前交易集中的支持度高。因為在考慮次數交易集的類別 E 中，項目集 {3, 4} 不但沒有因為聚集的結果在交易 B 中失去支持度，而且在類別 E 中因為 3 和 4 都分別出現 2 次，所以 {3, 4} 的支持度提升為 2，比原來聚集前交易集中的支持度增加 1。

再以另外一個例子來說明？何聚集交易集可以增加找出的頻繁項目集。表 6 中假設 A.1, A.2, A.3 這三本書同屬於類別 A，則 A.1, A.2 和 A.3 可以聚集成新的交易如表 7。其中類別 A 的借閱讀者群為取 A.1, A.2, A.3 這三本書的借閱讀者群聯集，並且保留原本讀者出現的次數，在讀者編號之後符號‘:’後面的數字代表出現過的次數，也就是讀者在這一個類別中借閱過的館藏數量。在表 6 和表 7 中，如果項目集後面有符號‘:’，則‘:’之後的數字代表那個項目集出現的次數，如果沒有則代表僅出現一次。

書	借閱讀者群	可產生的項目集
A.1	1, 2, 3	{1}, {2}, {3}, {1,2}, {1,3}, {2,3}, {1,2,3}
A.2	1, 2	{1}, {2}, {1,2}
A.3	3, 4	{3}, {4}, {3,4}
統計		{1}:2, {2}:2, {3}:2, {4}:1, {1,2}:2, {1,3}:1, {2,3}:1, {3,4}:1, {1,2,3}:1

表 6 屬於相同類別的交易集

類別	借閱讀者群	可產生的項目集
A	1:2, 2:2, 3:2, 4:1	{1}:2, {2}:2, {3}:2, {4}:1, {1,2}:2, {1,3}:2, {1,4}:1, {2,3}:2, {2,4}:1, {3,4}:1, {1,2,3}:2, {1,2,4}:1, {2,3,4}:1, {1,2,3,4}:1

表 7 將相同類別的書聚集起來的新交易集

從表 6 和表 7 中發現，聚集後的交易集可以產生的項目集增加了。比較表 6 中書 A.1, A.2, A.3 和表 7 中類別 A 所能產生的項目集，類別 A 增加了 {1, 4}, {2, 4}, {1, 2, 4}, {2, 3, 4}, {1, 2, 3, 4} 這五個項目集。

有以上兩個例子，可以得知在聚集過後的交易集中有必要額外考慮次數，也就是必需特別記錄讀者在這一個分類中曾借過的書籍數量，這樣才能確保不會造成支持度和資訊的流失，並且能真正提高項目集的支持度，使探勘結果能找出更多的頻繁項目集，解決原本因項目集的支持度不足而無法找出足夠頻繁項目集的問題。

而針對一般讀者借閱數偏少的問題，解決方法是將所有人的借閱數都經過正規化 (Normalize) 的處理，不再直接考慮讀者的借閱數，而是考慮讀者在每一分類中借閱的比例。例如一個讀者總共借過 16 本書，其中有 8 本 A 類的書、6 本是 B 類的書和 2 本是 C 類的書，則我們認為讀者的興趣 $8/16 = 50\%$ 在 A 類， $6/16 = 37.5\%$ 在 B 類， $2/16 = 12.5\%$ 在 C 類。將每個讀者的借書數都轉成比例來表示之後，讀者在每類或每筆交易中

出現的值都變成介於 0 與 1 之間，這樣一來不論讀者借書量的多寡，經過轉換後就都沒有差別了。假設表 8 為以一個類別為一筆交易的交易集，我們可以找出其中每個項目的支持度，亦即每個項目在交易集中的出現頻率，如表 9。轉換過後的交易集中，每個項目在每筆交易中的值(權重)，為該項目在每筆交易中出現的次數除以在交易集中總出現頻率。例如讀者 2 在交易中的總出現頻率為 5，其原本在類別 A 的出現次數為 1，所以讀者 2 在轉換後的類別 A 中的權重為 $1/5 = 0.2$ ，最後轉換完的結果如表 10 所示。表 10 為讀者 1, 2, 3, 4, 5 在 A, B, C, D, E 這五類中的興趣程度，而不論借閱數量多寡每個讀者在這五類的總和(支持度)皆為 1。

類別	借閱讀者群
A	2:1, 3:1
B	1:1, 2:1, 3:1, 4:2, 5:1
C	2:2, 3:1, 5:2
D	1:1, 2:1, 3:1, 4:2, 5:3
E	1:1, 3:1, 4:2

表 8 以一個類別為一筆交易的交易集

讀者	出現頻率
1	3
2	5
3	5
4	6
5	6

表 9 單一項目的出現頻率

項目 類別	1	2	3	4	5
A	0	0.2	0.2	0	0.17
B	0.33	0.2	0.2	0.33	0
C	0	0.4	0.2	0	0.33
D	0.33	0.2	0.2	0.33	0.5
E	0.33	0	0.2	0.33	0
支持度	1	1	1	1	1

表 10 經過正規化處理後的交易集

經過前述的聚集交易集並考慮項目數量及正規化項目數量這二種轉換之後，交易集資料中每筆交易不再只包含項目，還必須多考慮每個項目各自有不同的權重。在此我們採用[17]所提出的乏析式資料探勘，這種資料探勘的方法主要用來處理量化的資料 (Quantitative Data)。所謂量化的資料指的是一筆交易中，每個項目具有各自的權重值，且每個項目具有的權重不盡相同。

而乏析式資料探勘和非乏析式資料探勘最大不同的地方在於計算支持度的方式。非乏析式資料探勘在計算支持度時，如果計算的項目集中所有項目同時存在一筆交易中時，則這個項目集的支持度就加一，代表這個項目集的所有項目同時出現一次。之所以在非乏析式資料探勘中可以如此做的原因是，非乏析式資料探勘的交易中每個項目的權重都視為相同，一般假設為 1。但是在乏析式資料探勘中，交易中的每個項目權重不見得相同，所以必須用特殊的方法來處理。而在[17]提出的乏析式資料探勘中，計算支持度的方式分成兩個步驟：

- ◆ 步驟一：先計算具有 r 個項目的項目集 s 在每一個交易 $D^{(i)}$ 中的值 $f_s^{(i)}$ ，

$$f_s^{(i)} = f_{s_1}^i \wedge f_{s_2}^i \wedge \dots \wedge f_{s_r}^i$$

方程式 3 項目集於單筆交易內的支持度計算公式

其中 $f_{s_j}^{(i)}$ 代表 s 中第 j 個項目在交易 $D^{(i)}$ 中的值，在這裡交集 (?) 的運算代表

取最小值，所以 $f_s^{(i)}$ 可改寫成

$$f_s^{(i)} = \text{Min}_{j=1}^r f_{s_j}^{(i)}$$

方程式 4 以取最小值為交集運算的支持度計算公式

◆ 步驟二：計算項目集 s 在所有 n 筆交易中的支持度，

$$\text{support}_s = \sum_{i=1}^n f_s^{(i)}$$

方程式 5 乏析資料探勘中項目集的支持度計算公式

support_s 就代表項目集 s 的支持度。以表 10 為例，欲計算項目集 {1, 2} 的支持度，首先需要先算出 {1, 2} 在 A, B, C, D, E 每一個交易中的支持度。{1, 2} 在交易 A, B, C, D, E 中的支持度分別為

$$f_{\{1,2\}}^A = \text{Min}(f_1^A, f_2^A) = \text{Min}(0, 0.2) = 0$$

$$f_{\{1,2\}}^B = \text{Min}(f_1^B, f_2^B) = \text{Min}(0.33, 0.2) = 0.2$$

$$f_{\{1,2\}}^C = \text{Min}(0, 0.4) = 0$$

$$f_{\{1,2\}}^D = \text{Min}(0.33, 0.2) = 0.2$$

$$f_{\{1,2\}}^E = \text{Min}(0.33, 0) = 0$$

然後把 {1, 2} 在這些交易的支持度加總起來：

$$\text{support}_{\{1,2\}} = f_{\{1,2\}}^A + f_{\{1,2\}}^B + f_{\{1,2\}}^C + f_{\{1,2\}}^D + f_{\{1,2\}}^E = 0 + 0.2 + 0 + 0.2 + 0 = 0.4$$

最後 $\text{support}_{\{1,2\}} = 0.4$ 就是 {1, 2} 這個項目集的支持度。表 11 為將計算 $\text{support}_{\{1,2\}}$ 的整個過程列出，而表 12 就是所有兩個項目的項目集在各個交易中的支持度以及在所有交易的支持度。

類別	讀者 1	讀者 2	$\{1, 2\} = \text{Min}(1, 2)$
A	0	0.2	0
B	0.33	0.2	0.2
C	0	0.4	0
D	0.33	0.2	0.2
E	0.33	0	0
支持度	1	1	0.4

表 11 找出項目集 {1, 2} 的支持度

項目集 類別	{1,2}	{1,3}	{1,4}	{1,5}	{2,3}	{2,4}	{2,5}	{3,4}	{3,5}	{4,5}
A	0	0	0	0	0.2	0	0.17	0	0.17	0
B	0.2	0.2	0.33	0	0.2	0.2	0	0.2	0	0
C	0	0	0	0	0.2	0	0.33	0	0.2	0
D	0.2	0.2	0.33	0.33	0.2	0.2	0.2	0.2	0.2	0.33
E	0	0.2	0.33	0	0	0	0	0.2	0	0
支持度	0.4	0.6	0.99	0.33	0.8	0.4	0.7	0.6	0.57	0.33

表 12 所有長度為 2 的項目集的支持度

本論文用到的乏析式資料探勘跟 Apriori [2]演算法的步驟類似，都是先用已經找到的具有 n 個項目的頻繁項目集或稱為長度為 n 的頻繁項目集，來推測下一層哪些長度為 $(n+1)$ 的項目集可能是頻繁項目集，而這些尚未算出支持度的可能頻繁項目集也稱為候選 (Candidate) 項目集。然後再檢查整個交易集來算出每一個候選項目集的支持度，把低於最小支持度的候選項目集去掉，剩下來的就是這一層的頻繁項目集。之後再利用這些剛找出來的頻繁項目集繼續推測下一層長度加一的候選項目集，依此類推。當無法找出任何頻繁項目集或無法從這一層的頻繁項目集推出下一層的候選項目集時，探勘的動作就停止，而在過程中找到的頻繁項目集就是交易集中所有符合要求的頻繁項目集。

以表 10 的例子來說明，在本論文中所有長度為 1 僅具單一項目的項目集都是頻繁項目集，其支持度固定都是支持度的上限 1，不必再去檢查。這樣接下來如果全部有 k

個項目，也就是有 k 個讀者的話，則長度為 2 的候選項目集共有 $C(k,2) = k(k-1)/2$ 個。表 12 就是所有長度為 2 的候選項目集及其支持度。假設最小支持度設為 0.55，則可以得到的長度為 2 的頻繁項目集共有 6 個，如表 13 所示。

頻繁項目集	支持度
{1, 3}	0.6
{1, 4}	0.99
{2, 3}	0.8
{2, 5}	0.7
{3, 4}	0.6
{3, 5}	0.57

表 13 第二層中具有兩個項目的頻繁項目集

由於頻繁項目集的任一子集合必然也是頻繁項目集，所以從長度為 n 的頻繁項目集可推測出哪些長度為 $(n+1)$ 的項目集可能是頻繁項目集。在這個例子中長度為 3 的候選項目集有 {1, 3, 4} 和 {2, 3, 5}，因為 {1, 3, 4} 的子集合 {1, 3}, {1, 4}, {3, 4} 和 {2, 3, 5} 的子集合 {2, 3}, {2, 5}, {3, 5} 都屬於第二層的頻繁項目集。{1, 3, 4} 和 {2, 3, 5} 的支持度計算分別如表 14 和表 15。

類別	讀者 1	讀者 3	讀者 4	{1, 3, 4} = Min(1, 3, 4)
A	0	0.2	0	0
B	0.33	0.2	0.33	0.2
C	0	0.2	0	0
D	0.33	0.2	0.33	0.2
E	0.33	0.2	0.33	0.2
支持度	1	1	1	0.6

表 14 找出項目集 {1, 3, 4} 的支持度

類別	讀者 2	讀者 3	讀者 5	$\{2, 3, 5\} = \text{Min}(2, 3, 5)$
A	0.2	0.2	0.17	0.17
B	0.2	0.2	0	0
C	0.4	0.2	0.33	0.2
D	0.2	0.2	0.5	0.2
E	0	0.2	0	0
支持度	1	1	1	0.57

表 15 找出項目集 $\{2, 3, 5\}$ 的支持度

從表 14 和表 15 可知，項目集 $\{1, 3, 4\}$ 和 $\{2, 3, 5\}$ 的支持度分別為 0.6 和 0.57，仍然滿足預設的最小支持度 0.55。因為 $\{1, 3, 4\}$ 和 $\{2, 3, 5\}$ 無法再產生長度為 4 的候選項目集，所以整個探勘的工作到此告一段落，所找到的頻繁項目集共有 $\{1, 3\}$, $\{1, 4\}$, $\{2, 3\}$, $\{2, 5\}$, $\{3, 4\}$, $\{3, 5\}$, $\{1, 3, 4\}$, $\{2, 3, 5\}$ 這八個。

從關聯規則探勘所得到的頻繁項目集就是我們所需要的具有相同興趣的讀者群。而要產生讀者的推薦清單，主要包含三個步驟：第一步是先從之前找出的頻繁項目集中找出讀者的同好群，第二步把所有同好曾借閱過的館藏整理出來，第三步再把這些館藏依照推薦度高低排列，最後才推薦給讀者，詳細過程說明如下。

當讀者要求推薦時，必需先找出所有和這個讀者有相同興趣的其他讀者。舉例而言，當讀者 1 要求推薦時，假設有包含讀者 1 的頻繁項目集為 $\{1, 2, 3\}$, $\{1, 2, 5\}$, $\{1, 3, 8\}$, $\{1, 6, 7\}$ ，這些都是代表和讀者 1 有相同興趣的讀者群。從這些讀者群中可以得到與讀者 1 有相同興趣的有 2, 3, 5, 6, 7, 8 這 6 個讀者，這些讀者稱為讀者 1 的“同好”，而對讀者 1 的推薦就來自這些同好的歷史借閱紀錄。我們會統計這些同好的借閱資料，看這些同好經常借閱哪些書籍，愈常被借閱的書籍被視為愈有價值推薦給讀者 1。

在協力式過濾中，推薦度的計算是看一本書被同好中幾個人借閱過，以書籍 b 和讀者 u 為例，其推薦分數 Rec 的計算公式如下：

$$Rec(u,b) = \frac{fno_{u,b}}{fno_u}$$

方程式 6 協力式過濾的推薦分數計算公式

其中 fno_u 代表讀者 u 的總同好數， $fno_{u,b}$ 代表讀者 u 的同好中借過這本書的人數。例如讀者 1 有 6 個同好，有一本書 A 被這 6 個同好中的 5 個人借過，則書 A 對讀者 1 的推薦度就為 $Rec(1, A) = 5/6 = 83.33\%$ 。

綜合而言，本論文提出的協力式過濾推薦係先用關聯規則探勘找出跟讀者有相同興趣的同好，再利用這些同好來幫讀者過濾圖書館的館藏，找出適合讀者的館藏。

第二節 內容導向的推薦方法

本論文修正[8]中所提出用來建立超級市場中商品和顧客模型的方法，將其應用在建立圖書館中館藏與讀者的模型，因為超級市場和圖書館的情境有某種程度上的相似，所以只要稍微加以修改就很容易應用到圖書館，並且符合實際需求。

首先關於建立模型基礎的分類樹，因為圖書館已有很好的館藏分類系統，所以很容易就可以建立館藏的分類樹。在此我們以『中國圖書分類法』為基礎，取其前三層的分類表來建立系統分類樹。第一層分成十大類，而在每一個分類之下都有十個子分類，因此第二層有 $10 \times 10 = 100$ 個分類，第三層則有 $100 \times 10 = 1000$ 個分類。根據[8]的概念，本論文系統中的每一個物件都是以一個具有 1000 個屬性的向量來表達。

在？每一個物件找出他們的表示式之前，需要先找出每一層分類間的關聯規則，因為我們所採用的分類樹是三階層的架構，所以必須分別找出這三層間各自的關聯規則。

假設用來找出分類間關聯規則的交易集是以一個讀者在 120 天內的借書為一筆交易，交易中的項目為這段期間內該讀者借閱過的書，如表 16。

交易編號	項目群
T001	000 0034, 803.1 6000 v.20, 805.103 0725, 811.9 7531
T002	310.84 0257, 312.92 3421 c.2, 313 2341, 448.6 5423, 471.55 2876
T003	005.1 4444, 010.307 8700, 020.7 1080, 889 8769 v.2, 890 8728-1 v.1
T004	312.91695 5202-2, 313.98 4005-2, 861.57 6584 v.55, 862.57 0021 v.5
T005	030.8 4500 v.2, 080.25 1557 v.9, 312.33 5214

表 16 以書為交易項目的交易集

在表 16 中每筆交易中的書都是以索書號來表示，因為索書號中包含了分類上的意義。以中國圖書分類法為主體的索書號分成三個部份，第一部分為實際的分類號，第二部份是代表作者或叢書，最後是代表其他資訊。以索書號“030.8 4500 v.2”為例，“030.8”是該本書在中國圖書分類法中的分類號，“4500”指的是作者或叢書名的編號，最後“v.2”代表其他的資訊，例如在這裡“v.2”是代表第二冊。在交易集中我們將只取分類號的部份。

中國圖書分類法是一種十進位的分類，每一層可以有 0~9 的編號，而分類號的長度可以依照需要增加。但因為我們的分類樹只用到其最上面三層，所以最多只需取分類號中前三個數字就可以了。例如第三層也就是最底層的交易集中，每筆交易的項目是以每本書的索書號前三碼當代表，如表 17；而第二層和第一層交易集中的項目也分別取索書號的前兩碼和第一碼來表示，如表 18 和表 19。

交易編號	項目群	項目群投影
T001	000, 803, 805, 811	000, 803, 805, 811
T002	310, 312, 313, 448, 471	310, 312, 313, 448, 471
T003	005, 010, 020, 889, 890	005, 010, 020, 889, 890
T004	312, 313, 861, 862	312, 313, 861, 862
T005	030, 080, 312	030, 080, 312

表 17 探勘第三層分類所需的交易集

交易編號	項目群	項目群投影
T001	00, 80, 80, 81	00, 80, 81
T002	31, 31, 31, 44, 47	31, 44, 47
T003	00, 01, 02, 88, 89	00, 01, 02, 88, 89
T004	31, 31, 86, 86	31, 86
T005	03, 08, 31	03, 08, 31

表 18 探勘第二層分類所需的交易集

交易編號	項目群	項目群投影
T001	0, 8, 8, 8	0, 8
T002	3, 3, 3, 4, 4	3, 4
T003	0, 0, 0, 8, 8	0, 8
T004	3, 3, 8, 8	3, 8
T005	0, 0, 3	0, 3

表 19 探勘第一層分類所需的交易集

我們採用[20]所提出來的 H-mine 作為探勘關聯規則的演算法，其原因是這種演算法不需要產生候選項目集，可以有效率地找出關聯規則。

找出每一層的關聯性規則後，就可以開始建立每本書的表示式。書 i 的表示方法如下：

$$\text{Book}_i = (w(L_3C_1, i), w(L_3C_2, i), \dots, w(L_3C_k, i), \dots, w(L_3C_n, i))$$

方程式 7 書籍屬性向量表示式

其中 L 是代表階層 (Level)， C 代表分類 (Class)， L_3 在這裡代表第三層， L_3C_k 代表第三層中第 k 個分類，而 $w(L_3C_k, i)$ 代表是 Book_i 在第三層第 k 個分類的權重 (Weight)。根據 2.4 節中[8]提出的方程式 1 加以修改以符合圖書館的需要，我們提出計算 $w(L_3C_k, i)$ 的公式如下，其中 $L_p(a)$ 代表書或分類 a 在第 p 層所屬的分類：

$$w(L_3C_k, i) = \begin{cases} 1.0 & \text{if } L_3(i) = k & (\text{within same class of level 3}) \\ 0.8 & \text{if } L_3(i) \Rightarrow k & (\text{within associated class of level 3}) \\ 0.6 & \text{if } L_2(i) = L_2(k) & (\text{within same class of level 2}) \\ 0.5 & \text{if } L_2(i) \Rightarrow L_2(k) & (\text{within associated class of level 2}) \\ 0.3 & \text{if } L_1(i) = L_1(k) & (\text{within same class of level 1}) \\ 0.2 & \text{if } L_1(i) \Rightarrow L_1(k) & (\text{within associated class of level 1}) \\ 0 & \text{otherwise} \end{cases}$$

方程式 8 書籍屬性向量權重給定公式

用這個模型，可以？每本書找出其向量表示式。接下來可以再利用書的表示式來建立每個讀者的表示式代表其各自的興趣向量。由於讀者借閱過的書籍都隱含著讀者的興趣，所以可以利用讀者借閱過的書來建立讀者的表示式。在此我們將讀者所有借閱過的書的向量取平均代表讀者的興趣向量，讀者 j 的興趣向量 $Reader_j$ 如下所示，其中 n 代表讀者 j 共借過 n 本書：

$$Reader_j = \left(\sum_{i \in \text{all books reader } j \text{ had borrowed}} Book_i \right) / n$$

方程式 9 讀者興趣向量計算公式

將書和讀者的向量表示式都建好之後，便可利用這些表示式來計算書和讀者間的關係程度。本論文以計算兩個向量距離的公式來求得書與讀者間的符合度 $Like(Reader_j, Book_i)$ ，用以預測讀者 j 對書 i 的喜好程度，公式如下：

$$Like(Reader_j, Book_i) = \frac{Reader_j \cdot Book_i}{|Reader_j| \times |Book_i|}$$

方程式 10 內容導向過濾的符合度計算公式

$Like(Reader_j, Book_i)$ 也就是內容導向過濾推薦對讀者 j 關於書 i 的推薦度。利用這個方法內容導向的推薦系統可以？每個讀者判斷出哪些館藏跟讀者興趣的符合度較高。

第三節 結合協力式過濾和內容導向的推薦方法

利用第一節提到的協力式過濾方法，每位讀者可以得到一份推薦清單，然而因為這份清單是參考同好的借閱歷史所得到的，所以不一定都符合讀者的興趣。對第二節的內容導向過濾推薦方法而言，雖然有辦法幫每個讀者計算他的興趣與書籍間的符合度，但是如果要把所有圖書館館藏都計算過符合度再來做推薦，將相當沒有效率。所以如果能結合第一節所得到的推薦清單和第二節的書和讀者的表示式，剛好可以互補解決各自的問題。

首先用協力式過濾的方式得到讀者的推薦清單和每一件館藏的推薦度，然後再用內容導向過濾的方法計算這份清單內所包含的館藏和讀者間的符合度。結合協力式過濾所得到的推薦度 Rec 和內容導向過濾得到的符合度 $Like$ ，我們可以得到新的推薦度 New_Rec ，其計算公式如下：

$$New_Rec(j,i) = \mathbf{b} \times Rec(j,i) + (1 - \mathbf{b}) \times Like(j,i)$$

方程式 11 新的推薦分數計算公式

其中 j 代表某一讀者， i 代表某一館藏， \mathbf{b} 為可調整的參數，其值介於 0 - 1 之間，代表要取 Rec 和 $Like$ 的權重，可以視需要增減。如果 \mathbf{b} 的值取 0.5，代表協力式過濾和內容導向過濾的影響各半。

得到新的推薦分數後，再把從協力式過濾得到的推薦清單，依照新的推薦分數排序，最後把重新排序過的結果推薦給讀者，完成整個推薦流程。

第四章 系統實作

在這一章中，會介紹本論文提出的圖書館個人化館藏推薦系統於交通大學浩然圖書館中實作的細節，並說明系統運作的流程。

第一節為圖書館個人化館藏推薦系統的整體系統概述。第二節裡說明資料的來源和一些必要的資料前置處理，使其成為符合系統需要的資料型態。第三節為建立內容導向過濾需要的系統模型和內部物件屬性表示式的步驟。第四節介紹實際上產生推薦清單的流程。第五節中展示了推薦系統於交通大學浩然圖書館 myLibrary 上呈現的形式。

第一節 系統架構

圖 11 為圖書館個人化館藏推薦系統的架構圖，主要分成關聯規則探勘 (Association Mining)、協力式過濾推薦 (Collaborative Recommend) 和內容導向過濾推薦 (Content-Based Recommend) 這三個部分：

- ◆ 關聯規則探勘：這個部分主要將借閱歷史資料庫依照需求先經過前置處理以產生適合的交易集，再分別經過乏析式資料探勘和 H-mine 產生讀者與讀者間的關聯規則和書與書間的關聯規則，以供之後協力式過濾推薦和內容導向過濾之用。
- ◆ 協力式過濾推薦：這部分依照讀者與讀者間的關聯規則從借閱歷史資料庫中找出同好間推薦適合的館藏推薦清單。
- ◆ 內容導向過濾推薦：這部分包含建立書籍的屬性向量表示式資料庫和讀者的個人興趣向量表示式 (Profile) 資料庫，並依照讀者的個人興趣將之前從協力式過濾推薦得到的館藏推薦清單重新排序，最後產生具有個人化的館藏推薦清單。

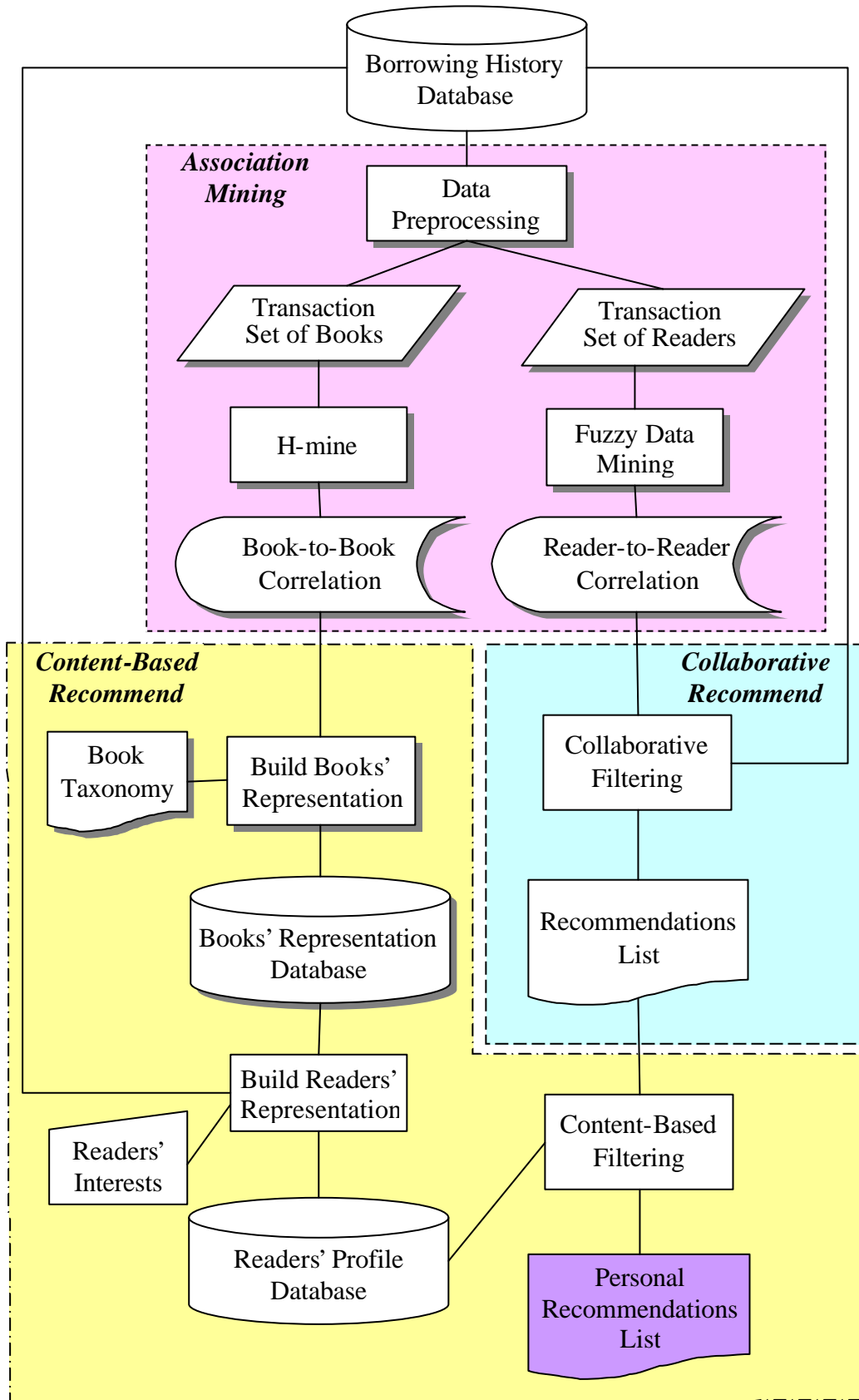


圖 11 圖書館個人化館藏推薦系統架構圖

在圖 11 中有陰影部分的圖示代表其為離線(Offline)處理程序,例如關聯規則探勘中探勘讀者間和書籍間的關聯規則及內容導向過濾推薦中建立書籍的屬性向量表示式,這些因為都需要大量時間和空間運算,所以皆採取離線處理的方式,在系統上線前先準備好。而系統中其他部分則皆為線上作業,例如得到讀者的興趣類別以建立讀者的個人興趣向量表示式和產生館藏推薦清單等,都是直接在網頁上和讀者互動。

第二節 前置處理

本系統目前實作所採用的資料來源是國立交通大學浩然圖書館 2001 年 5 月到 2003 年 4 月這兩年間的讀者借閱歷史檔,每一筆資料代表一個讀者借閱一本書的記錄。在 2001 年 5 月到 2003 年 4 月這段期間中有 643,557 筆借閱資料,7,682 個讀者曾借閱過館藏。在關聯規則探勘前,必須先將此借閱歷史檔經過資料前置處理(Data Preprocessing),以分別產生協力式過濾和內容導向過濾所需的交易集。

協力式過濾需要的交易集是以一個類別為一筆交易,在本論文中是以中國圖書分類法的類別為分類基準,每本書所在的類別以其索書號上的分類號為準。交易中的每一個項目代表一位曾借閱過該類別書籍的讀者和此讀者借閱過該類別書的次數。

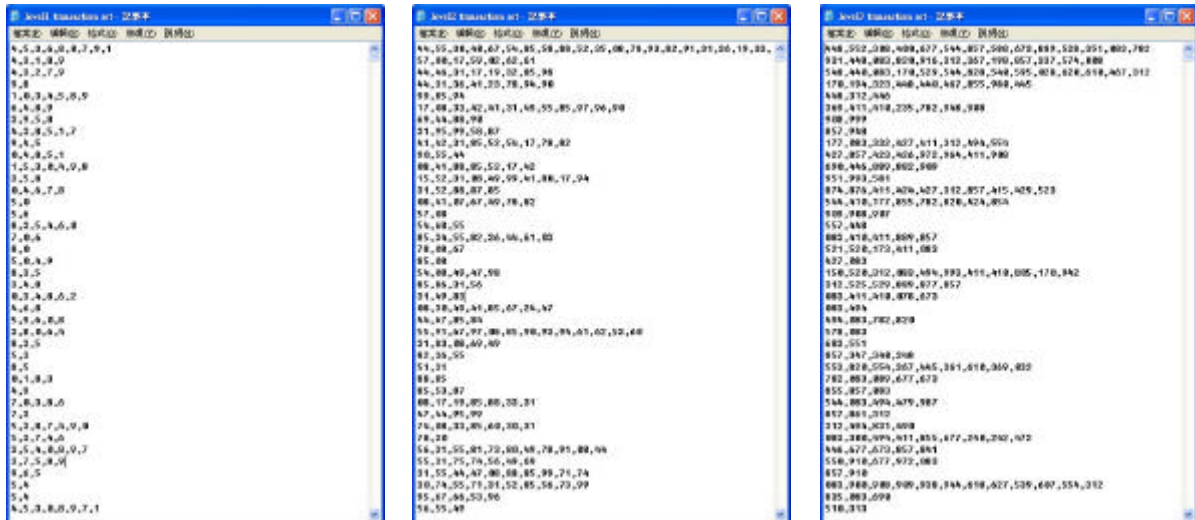
目前我們僅做總類(General)和語文類(Language and Literature)書籍的推薦,經過前置處理過後產生的交易集中有 836 筆交易。圖 12 是部分的交易內容,其中每一行為一筆交易,交易中的項目間以符號“;”隔開,每個項目在符號“:”之前為項目名稱(讀者編號),之後為該項目出現的次數(讀者於此類別借閱過的館藏數量)。

內容導向過濾需要的交易集是以一個讀者在某段時間內的借閱書籍為一筆交易,在實作上取某一讀者在 120 天內借閱過的書籍當做一筆交易。因為內容導向過濾需要對分類樹的每一層分類作相關規則探勘,所以每一層都需要自己的交易集。各層的交易集如

圖 13, 第一層、第二層和第三層的交易集中的數量分別有 24,079 筆、25,112 筆和 42,060 筆交易。



圖 12 協力式過濾需要的交易集



(a)

(b)

(c)

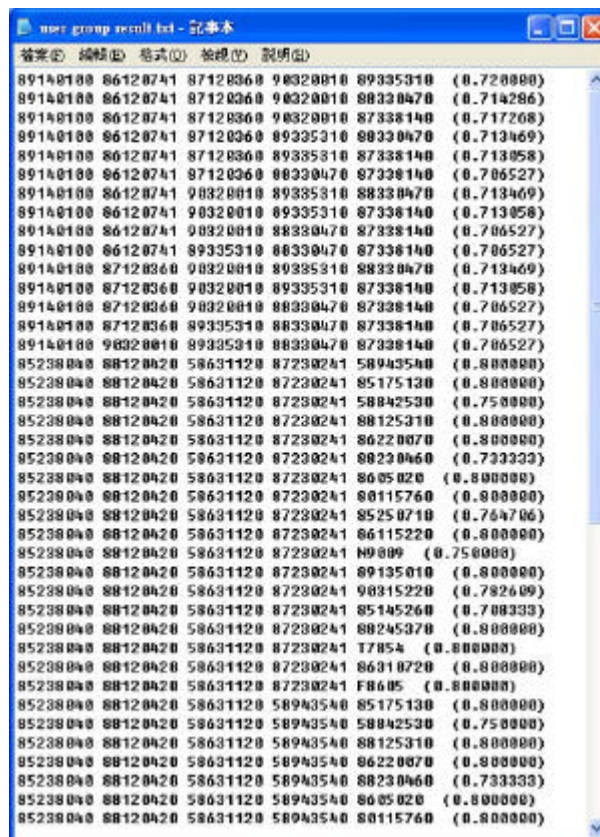
圖 13 內容導向過濾需要的交易集

(a) 為第一層的交易集；(b) 為第二層的交易集；(c) 為第三層的交易集

第三節 找出相關讀者群和建立館藏與讀者的表示法

在產生給讀者的推薦清單前，必須先找出協力式過濾所需要的相關讀者群及內容導向過濾需要的館藏與讀者的表示法。

相關讀者群是用乏析式資料探勘來尋找，而在實作乏析式資料探勘時則參考了[3]提出以先序樹實作 Apriori 的程式，加以改寫成可處理量化資料的乏析式資料探勘。當最小支持度為 0.7 時，可將將近三千名讀者找到所屬的相關讀者群，部份結果節錄如圖 14。在圖 14 中每行代表一個相關讀者群，裡面的項目代表讀者，在同一讀者群中的讀者會被視為有相同興趣，在每一行最後在括號“()”中間的數字代表這個讀者群的支持度。



讀者群	讀者	支持度
89140100	86120741 87120360 90320010 89335310	(0.720000)
89140100	86120741 87120360 90320010 88330470	(0.714286)
89140100	86120741 87120360 90320010 87338140	(0.717268)
89140100	86120741 87120360 89335310 88330470	(0.713469)
89140100	86120741 87120360 89335310 87338140	(0.713058)
89140100	86120741 87120360 88330470 87338140	(0.706527)
89140100	86120741 90320010 89335310 88330470	(0.713469)
89140100	86120741 90320010 89335310 87338140	(0.713058)
89140100	86120741 90320010 88330470 87338140	(0.706527)
89140100	86120741 89335310 88330470 87338140	(0.706527)
89140100	87120360 90320010 89335310 88330470	(0.713469)
89140100	87120360 90320010 89335310 87338140	(0.713058)
89140100	87120360 90320010 88330470 87338140	(0.706527)
89140100	87120360 89335310 88330470 87338140	(0.706527)
85238040	88120420 58631120 87230241 58943540	(0.800000)
85238040	88120420 58631120 87230241 85175130	(0.800000)
85238040	88120420 58631120 87230241 58842530	(0.750000)
85238040	88120420 58631120 87230241 88125310	(0.800000)
85238040	88120420 58631120 87230241 86220070	(0.800000)
85238040	88120420 58631120 87230241 88230460	(0.733333)
85238040	88120420 58631120 87230241 8605020	(0.800000)
85238040	88120420 58631120 87230241 88115760	(0.800000)
85238040	88120420 58631120 87230241 85250710	(0.764706)
85238040	88120420 58631120 87230241 86115220	(0.800000)
85238040	88120420 58631120 87230241 N9009	(0.750000)
85238040	88120420 58631120 87230241 89135010	(0.800000)
85238040	88120420 58631120 87230241 98815220	(0.782609)
85238040	88120420 58631120 87230241 85145260	(0.708333)
85238040	88120420 58631120 87230241 88245370	(0.800000)
85238040	88120420 58631120 87230241 T7854	(0.800000)
85238040	88120420 58631120 87230241 86310720	(0.800000)
85238040	88120420 58631120 87230241 F8605	(0.800000)
85238040	88120420 58631120 58943540 85175130	(0.800000)
85238040	88120420 58631120 58943540 58842530	(0.750000)
85238040	88120420 58631120 58943540 88125310	(0.800000)
85238040	88120420 58631120 58943540 86220070	(0.800000)
85238040	88120420 58631120 58943540 88230460	(0.733333)
85238040	88120420 58631120 58943540 8605020	(0.800000)
85238040	88120420 58631120 58943540 88115760	(0.800000)

圖 14 相關讀者群的頻繁項目集

在內容導向過濾中也需要用關聯規則探勘找出分類樹中各層的關聯規則，在此用的相關規則探勘演算法是 H-mine。分別對每一層的交易集執行一次 H-mine，在各層應不同的資料數量分別採取不同的最小支持度 (Min. Support)、最小確信值 (Min. Confidence) 和找出來的關聯規則，如表 20 所列。各階層的關聯規則格式和其確信值如表 21 所列舉。

階層	最小支持度 (Min. Support)	最小確信值 (Min. Confidence)	關聯規則數量 (Rules Number)
第一層	0.1	0.8	18
第二層	0.05	0.7	41
第三層	0.001	0.8	603

表 20 各階層分類探勘的最小支持度、最小確信值及最後關聯規則數量

階層	關聯規則	確信值 (Confidence)
第一層	4 5 8 ⇒ 3	0.823094
	4 5 0 ⇒ 3	0.826161
	5 3 1 ⇒ 4	0.805466

第二層	44 85 08 ⇒ 31	0.769517
	88 08 31 ⇒ 85	0.758092
	85 88 31 ⇒ 08	0.80084

第三層	448 308 471 ⇒ 312	0.8125
	448 820 861 ⇒ 857	0.818182
	448 889 312 855 ⇒ 857	0.824561

表 21 各階層所探勘出的關聯規則和確信值

得到這三層分類的關聯性規則後，搭配 3.2 節所提出利用方程式 8 計算書籍屬性向量的方法，可以每一本書找出其表示式。但由於系統只有分類樹中第三層的每一分類建立其表示式，因此屬於同一分類的書皆共用同一個表示式。例如索書號中分類號分別為 312.3 和 312.52 的這兩本書，實際上都是共用 312 這個分類的屬性向量表示式。因

為分類樹第三層中共有 1000 個分類，而這 1000 個分類各有各自的表示式，所以整個系統共有 1000 個不同的表示式來表示所有圖書館的館藏。圖 15 是分類 312 下館藏的屬性向量分布圖，其中圈起來的部份為類別 312 所在的地方，因為館藏屬於分類 312，所以其值為 1。

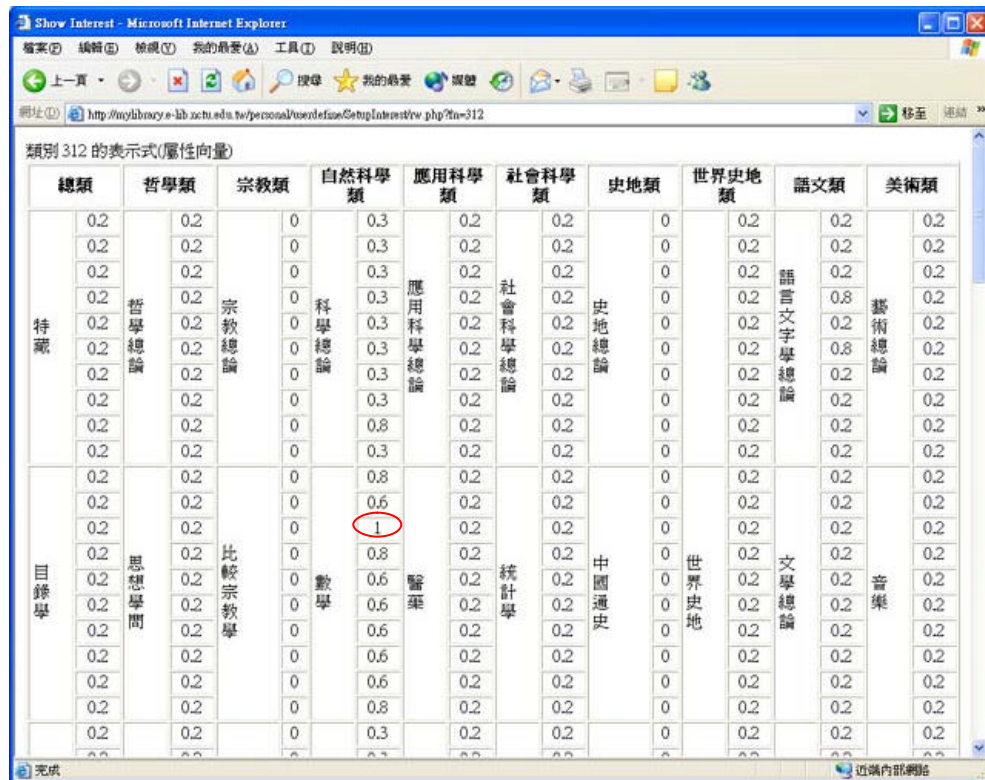


圖 15 類別 312 的屬性向量表示式

把書的表示式 $Book_i$ 建好後，根據 3.2 節中的方程式 9，可以利用讀者的借閱歷史檔來產生讀者本身的興趣表示式 $Reader_j$ 。但是在這裡除了利用借閱歷史檔來產生讀者的表示式外，讀者還可以藉由調整自己的興趣來改變其興趣表示式。系統中設計了一個介面讓讀者選擇自己真正喜歡的館藏類別，如圖 19。但因為一般讀者對中國圖書分類法不是很熟悉，所以系統只提供第一層和第二層涵蓋範圍比較大的分類給讀者選擇，而且在每一分類名稱之後也加注讀者在這一類中所借閱過的書數。如果讀者選擇了他有興趣的類別，系統便會在原本讀者的表示式 $Reader_j$ 中對這些讀者選定的類別增加一定比例的權重，而且這增加的比例會因為分類在分類樹上的層數而不同。如果讀者是選擇層數比較高，範圍較廣泛 (General) 的分類，則增加的權重比例較低，如果是層數比較低，範

圍較小 (Specific) 的分類，則增加的權重比例較高。其理由在於若讀者能指定比較底層，範圍比較明確的分類，代表讀者的興趣愈明確，所以可以在那個分類增加比較高的權重。修正之後的讀者興趣向量表示式 $Reader'_j$ 如下：

$$Reader'_j = \begin{cases} Reader_{jk} \times (1 + g) & \text{if class } k \text{ is subclass of class} \\ & \text{chosen by reader } j \text{ in level 1 only} \\ Reader_{jk} \times (1 + d) & \text{if class } k \text{ is subclass of class} \\ & \text{chosen by reader } j \text{ in level 2} \\ Reader_{jk} & \text{otherwise} \end{cases}$$

方程式 12 讀者興趣向量修正公式

其中 $Reader_{jk}$ 代表讀者 j 在類別 k 原來的權重， $Reader'_j$ 代表修正後的權重，在實作上，我們取 $g = 0.2$ 、 $d = 0.5$ 。而建好的讀者興趣向量跟書的屬性向量類似，代表讀者在各個分類的興趣權重，如圖 16。

The screenshot shows a web browser window displaying a library website. The main content is a table with 11 columns representing different subject categories and multiple rows of numerical interest weights. The categories are: 總類, 哲學類, 宗教類, 自然科學類, 應用科學類, 社會科學類, 史地類, 世界史地類, 語文類, and 美術類. The numerical values range from 0 to 0.6413. The table is organized into a grid with sub-headers for each category.

總類	哲學類	宗教類	自然科學類	應用科學類	社會科學類	史地類	世界史地類	語文類	美術類
0.1868	0.1188	0	0.2418	0.1706	0.1188	0	0.1422	0.1846	0.1422
0.1868	0.1188	0	0.2418	0.1706	0.1188	0	0.1422	0.1846	0.1422
0.1868	0.1188	0	0.2418	0.1706	0.1188	0	0.1422	0.1846	0.1422
0.1868	0.1188	0	0.2418	0.1706	0.1188	0	0.1422	0.6121	0.1422
0.1868	0.1188	0	0.2418	0.1706	0.1188	0	0.1422	0.1846	0.1422
0.1868	0.1188	0	0.2418	0.1706	0.1188	0	0.1422	0.6607	0.1422
0.1868	0.1188	0	0.2418	0.1706	0.1188	0	0.1422	0.1846	0.1422
0.1868	0.1188	0	0.285	0.1706	0.1188	0	0.1422	0.1846	0.1422
0.1868	0.1188	0	0.6413	0.1706	0.1188	0	0.1422	0.1846	0.1781
0.1868	0.1188	0	0.2418	0.1706	0.1188	0	0.1422	0.1846	0.1422
0.1868	0.1188	0	0.7895	0.2138	0.1188	0	0.1422	0.2278	0.1422
0.1868	0.1188	0	0.6113	0.2138	0.1188	0	0.1422	0.1846	0.1422
0.1868	0.1188	0	1	0.1706	0.1188	0	0.1422	0.2278	0.1422
0.1868	0.1188	0	0.7895	0.1706	0.1188	0	0.1422	0.2332	0.1422
0.1868	0.1188	0	0.6113	0.1706	0.1188	0	0.1422	0.1846	0.1422
0.1868	0.1188	0	0.6113	0.2138	0.1188	0	0.1422	0.1846	0.1422
0.1868	0.1188	0	0.6113	0.1706	0.1188	0	0.1422	0.1846	0.1422

圖 16 讀者的個人興趣向量

第四節 產生推薦清單

前述產生相同興趣讀者群和館藏與讀者的表示式都是離線 (Offline) 作業，在系統作業之前就必須準備好，其主因為關聯規則探勘的工作通常需要花費冗長的時間。之後產生推薦清單的工作就是直接在線上 (Online) 作業。

產生推薦清單的工作主要分成如下幾個步驟：

1. 找出跟讀者具有相同興趣的同好

首先從先前以乏析式資料探勘所得之相同興趣讀者群中，找出包含目標讀者的頻繁項目集，在這些頻繁項目集中的其他讀者就是和目標讀者具有相同借閱興趣的同好群。以表 22 為例，若想找出讀者 3 的同好，則必須先找出包含讀者 3 的頻繁項目集，將這些頻繁項目集聯集，則可以得到讀者 3 的同好群 $\{1, 2, 5, 7, 8, 9\} = (\{1, 2, 3, 5\} \cup \{2, 3, 8, 9\} \cup \{3, 7, 9\} \cup \{1, 3, 7, 8\} \cup \{2, 3, 7, 9\}) - \{3\}$ 。

#	頻繁項目集
S001	{1, 2, 3, 5}
S002	{2, 3, 8, 9}
S003	{3, 7, 8}
S004	{1, 3, 7, 8}
S005	{2, 3, 7, 9}

表 22 包含讀者 3 的頻繁項目集

2. 找出同好的借閱歷史

找出目標讀者的同好群之後，開始檢索這些同好的借閱歷史檔。假設讀者 3 和其同好群 $\{1, 2, 5, 7, 8, 9\}$ 曾借閱過的書如表 23，則從表 23 中可以找出讀者 3

的同好群借過但是讀者 3 沒借閱過的書籍，並且可以算出每本書的推薦度。協力式過濾推薦度的算法為同好借閱次數除以同好數，參考方程式 6，如書 G 曾被四個同好借過，而讀者 3 的同好共有六人，所以書 G 的推薦程度為 $4/6=0.66$ 。表 24 即為讀者 3 所得到的推薦書目和每本書的推薦度，表中的書目以推薦度排序。

讀者	借閱書籍
1	A, B, C, E, G, H, I
2	A, B, D, G
3	A, B, E, H, J
5	B, E, I, J
7	B, C, D, F, G, I
8	A, D, E, H
9	C, F, G, H, J

表 23 讀者借閱歷史

推薦書目	同好借閱次數	推薦度
G	4	$4/6 = 0.66$
C	3	$3/6 = 0.50$
D	3	$3/6 = 0.50$
I	3	$3/6 = 0.50$
F	2	$2/6 = 0.33$

表 24 讀者 3 在協力式過濾推薦裡得到的推薦數目

3. 計算推薦清單與讀者興趣的符合度

因為同好的推薦清單並不一定完全符合讀者的興趣，所以必須修正推薦清單的推薦度，將讀者興趣納入考量。假設每本書的屬性向量表示式如表 25，每

本書均以具 10 個屬性的向量表示(在系統實作時讀者興趣向量的長度為 1000) , 則可以得到讀者 3 的興趣向量為 $(A + B + E + H + J) / 5 = (0.6, 0.6, 0.28, 0.24, 0.64, 0.12, 0.36, 0.56, 0.24, 0.44)$ 。利用讀者 3 的興趣向量和表 24 中推薦清單中每本書的屬性向量, 用方程式 10 可以計算出讀者的興趣與每本書的符合度, 結果如表 26 所示。

書目	屬性向量
A	(1, 0.6, 0.2, 0.4, 0.6, 0, 0.4, 0.6, 0.2, 0.2)
B	(0.6, 1, 0.4, 0.4, 0.6, 0.2, 0.6, 0.4, 0.6, 0.4)
C	(0.2, 0.4, 1, 0.2, 0.2, 0.4, 0.6, 0.4, 0.4, 0.2)
D	(0.4, 0.4, 0.2, 1, 0.2, 0.2, 0.4, 0.2, 0.2, 0)
E	(0.6, 0.6, 0.2, 0.2, 1, 0, 0.2, 0.6, 0.4, 0.4)
F	(0, 0.2, 0.4, 0.2, 0, 1, 0.4, 0.2, 0.2, 0.2)
G	(0.4, 0.6, 0.6, 0.4, 0.2, 0.4, 1, 0.4, 0.4, 0.2)
H	(0.6, 0.4, 0.4, 0.2, 0.6, 0.2, 0.4, 1, 0.2, 0.2)
I	(0.2, 0.6, 0.4, 0.2, 0.4, 0.2, 0.4, 0.2, 1, 0.2)
J	(0.2, 0.4, 0.2, 0, 0.4, 0.2, 0.2, 0.2, 0.2, 1)

表 25 書的屬性向量表示式

推薦書目	推薦度	符合度	新的推薦度 ($b = 0.2$)	排序
G	0.66	0.80	0.772	1
I	0.50	0.75	0.70	2
C	0.50	0.72	0.676	3
D	0.50	0.68	0.644	4
F	0.33	0.46	0.434	5

表 26 推薦書目的新的推薦度和排序

4. 重新排序推薦清單

得到推薦清單中每本書和讀者興趣的符合度後，再搭配之前在協力式過濾中所得到的推薦度，可以將這兩個分數結合起來根據方程式 11 來計算新的推薦度，並且對推薦清單重新排序，使推薦清單能更符合讀者的個人興趣，這裡我們取 b 值為 0.2，結果如表 26 所示。

5. 以網頁形式把推薦清單呈現出來

得到最後排序好的推薦清單後，就可以將這些推薦書目呈現出來供讀者參考。呈現推薦清單的方式是先列出整份清單的書目和一些簡要資訊，例如書名、作者和出版社這些資料，當讀者有興趣時可以點選書名上的超連結進一步檢視詳細書目資料，或者也可以點選“書籍簡介”欄位上網路書店的超連結，以連結到網路書店上觀看書評。圖 18 就是最後推薦結果的呈現畫面。

第五節 系統在 myLibrary 上的呈現方式

個人化館藏推薦系統實作在交通大學浩然圖書館中的個人化數位圖書資訊環境 myLibrary 上 (<http://mylibrary.e-lib.nctu.edu.tw/>)。myLibrary 是一套個人化的圖書館系統，在 myLibrary 上使用者可以自訂興趣，建立一個屬於自己的圖書館網頁，幫助自己整理和收集資料。而個人化館藏推薦系統就是 myLibrary 其中一項服務。

在 myLibrary 中，讀者在登入之後點選“我的館藏”，會先看到類似圖 17 的畫面，裡面包含有圖書館的本月新書與百大熱門館藏 (Top 100)，此二功能皆屬於無個人化的推薦。

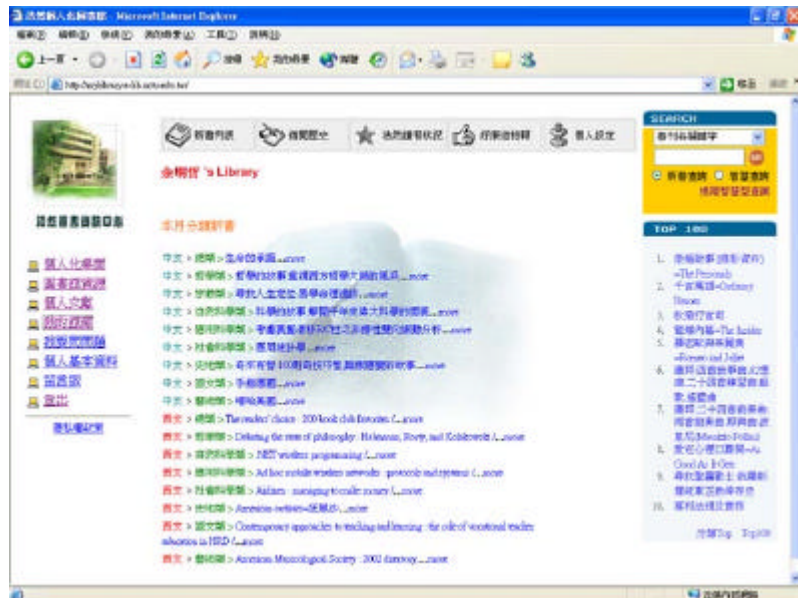


圖 17 myLibrary 中“我的館藏”畫面

接下來點選圖 17 上方“好康道相報”的連結，就可以看到系統所推薦的館藏，如圖 18。點選書名的超連結就可以看到詳細的書目資料。書籍簡介欄位可以直接讓讀者連到網路書店上看看書的短評或心得感想，目前提供博客來 (<http://www.books.com.tw>) 和金石堂 (<http://www.kingstone.com.tw>) 這兩家網路書店的連結。Score(分數)欄位便是系統對這本書的推薦分數，範圍為 0 - 1 之間。而推薦滿意度則是用來讓讀者提供回饋 (Feedback)，回應系統推薦的館藏是否符合其興趣。



圖 18 系統推薦館藏的呈現畫面

系統也提供給讀者自由調整興趣的功能，只要選擇“個人設定”中的“興趣類別設定”就可以設定自己的興趣。目前系統中提供選擇設定的是中國圖書分類法中第一層和第二層的分類，圖 19 是設定的畫面，每個分類名稱後面的括號內的數字代表讀者在該分類中曾借過的書數。此外，當讀者第一次要求館藏推薦時，系統會先要求讀者設定個人興趣，已取得讀者興趣給後來推薦之用。當讀者設定完個人興趣時，系統會自動更新讀者的個人興趣向量。另外，在“個人設定”中讀者也可以選擇“檢視個人興趣”來查看自己的興趣分佈，呈現畫面如圖 16 所示。



(a)

(b)

圖 19 設定個人興趣畫面 (a) 設定第一層十大分類的興趣；(b) 設定第二層各個分類的興趣

比對本論文所提出的個人化館藏推薦系統與由[23]所提出的一般推薦系統模型，表 27 列出了三部份元件(1)輸入與輸出；(2)推薦方法；(3)其他設計考量分別對應到的項目。

輸入與輸出		推薦方法	其他設計考量	
輸入	輸出		個人化	推薦送達方式
直接詢問讀者興趣 讀者借閱歷史 社群借閱歷史	預測，有排序的推薦清單	屬性導向推薦 館藏間的關聯性 讀者間的關聯性	長期個人化	等候選取

表 27 圖書館個人化館藏推薦系統各細部元件

第五章 館藏推薦滿意度評估

本章將評估圖書館個人化館藏推薦系統的執行效果，計算讀者對系統推薦書籍的滿意度。第一節說明如何收集讀者回應。第二節介紹用以評估推薦滿意度的幾個方法。第三節列出最後得到的結果數據並加以分析討論。

第一節 收集讀者對館藏推薦的回應

在館藏推薦的呈現畫面中提供了“書籍簡介”的連結讓讀者參考，並希望讀者依此回應推薦滿意度。圖 20 中圈選起來的部份就是讀者回應滿意度的地方，讀者選擇完後可以按下“送出滿意度”按鈕回報對館藏推薦的滿意度。在這裡的滿意度就是指讀者對這本書的喜好程度，分成五個等級讓讀者選擇，由好到壞分別是“非常滿意”、“滿意”、“普通”、“不滿意”和“非常不滿意”五種，實際上內部代表的分數為 0.9, 0.7, 0.5, 0.3, 0.1。讀者所送出的資料將會被存到系統資料庫中以供之後的分析，作為系統改進的參考。

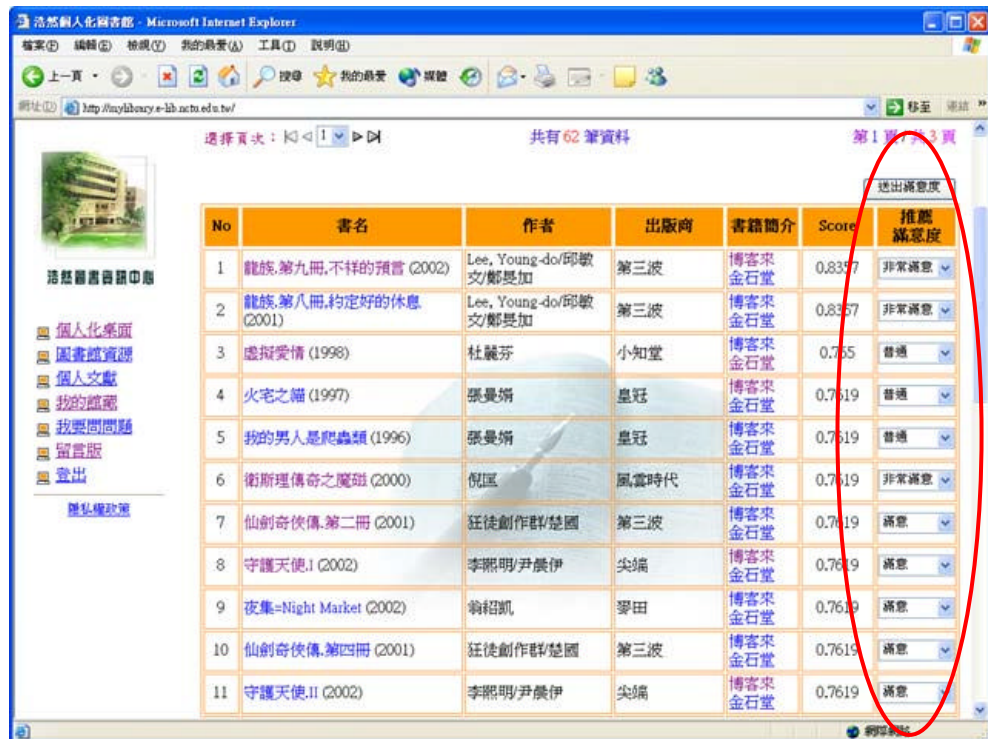


圖 20 讀者滿意度回應畫面

第二節 計算讀者對館藏推薦的滿意度

我們總共採用三種方法來計算讀者對館藏推薦的滿意度。前兩種是 Breese 和 Kadie 在[4]所提出的兩種測量推薦系統滿意度的公式，分別為平均誤差值 (Mean Absolute Error, MAE) 和排序分數度量值(Rank Score Measure, RSM)。第三種為計算讀者對所推薦的館藏的平均滿意度 (Average Satisfaction)。

平均誤差值主要用來計算使用者對系統推薦物品的整體滿意度，公式如下：

$$S_a = \frac{1}{m_a} \sum_{j \in p_a} |p_{a,j} - v_{a,j}|$$

方程式 13 平均誤差值計算公式

其中 S_a 代表讀者 a 的整體滿意度， m_a 代表讀者 a 得到的推薦數量， p_a 為系統給讀者 a 的所有館藏推薦， $p_{a,j}$ 為系統預測讀者 a 對館藏 j 的滿意度， $v_{a,j}$ 則是讀者 a 對館藏 j 的實際滿意度。

平均誤差值的主要意義是計算讀者實際滿意度和系統預測滿意度之間的差距，如果相差越少則 S_a 會愈小，代表系統的推薦跟讀者的興趣愈接近，推薦的效果愈好，反之則愈差。

排序分數度量值則是計算使用者對系統推薦的排序是否滿意，公式如下：

$$R_a = \sum_j \frac{\max(v_{a,j} - d, 0)}{2^{(j-1)/(a-1)}}$$

方程式 14 排序分數度量值計算公式

R_a 就是代表讀者 a 對館藏推薦排序的滿意度。其中 j 為在推薦清單中排序的位置， j 愈小代表排的愈前面，推薦程度越高。反之，如果 j 愈大代表排序愈後面，推薦程度

比較低。 $v_{a,j}$ 是讀者 a 對推薦館藏 j 的實際滿意度； d 為讀者滿意度的平均值； a 代表計算公式中的半衰期。

R_a 的意義在於如果讀者比較喜歡的館藏都在推薦清單排序比較前面的位置，而比較不喜歡的館藏都集中在排序比較後面的位置，則 R_a 可以得到最大的值。相反地，萬一讀者比較不喜歡的館藏都在推薦清單排序比較前面的位置，而比較喜歡的館藏卻都集中在排序比較後面的位置，那 R_a 所得到的值就會很低。所以 R_a 的值可以看出讀者對推薦清單排序的滿意度。通常排序分數度量值還需要將 R_a 除以 R_a^{\max} 正規化，使得到的值能介於某個範圍內。 R_a^{\max} 代表當推薦排序完全依照讀者喜好排序時，亦即讀者喜歡的都在排序的前面，不喜歡的都在後面，所能得到的最大排序滿意值。這樣得到的分數可以比較明確地看出系統排序的結果到底有多接近讀者理想的排序。正規劃的排序分數度量值 R 計算公式如方程式 15。

$$R = 100 \times \frac{\sum_a R_a}{\sum_a R_a^{\max}}$$

方程式 15 排序分數度量值正規化計算公式

平均滿意度的計算較為簡單，純粹統計讀者回應滿意度的平均值。先算出每個讀者的滿意度平均值為其個人滿意度，再計算所有讀者個人滿意度的平均值，最後計算出來的值即為我們需要的所有讀者對系統推薦館藏的平均滿意度。

第三節 結果分析

在交通大學浩然圖書館的個人化館藏推薦系統中，我們收集 415 位讀者對系統推薦館藏的滿意度回應後，利用平均誤差值、排序分數度量值和平均滿意度這三種方法來判斷讀者對這個個人化館藏推薦系統的滿意度，結果如表 28。

評估方式 Top N	平均誤差值	排序分數度量值	平均滿意度
Top 10	0.117571	0.827082	0.672096
Top 20	0.111500	0.764413	0.671700
Top 30	0.110378	0.734757	0.670155
Top 40	0.108454	0.706947	0.670730
Top 50	0.108790	0.685876	0.670002

表 28 系統滿意度調查結果

表 28 中的 Top N 代表取系統給讀者的前 N 名推薦的書籍滿意度，例如 Top 10 代表取系統給讀者的推薦中，推薦排名在前 10 名的書籍滿意度來計算。

圖 21 為平均誤差值結果的折線圖，誤差值大約落在 0.1 和 0.12 之間。因為系統中給予讀者選擇的滿意度，每個區間相差 0.2，所以誤差 0.1 代表系統的推薦分數和讀者選擇的滿意度差距仍在一個區間誤差內，代表誤差仍在可接受的合理範圍內。

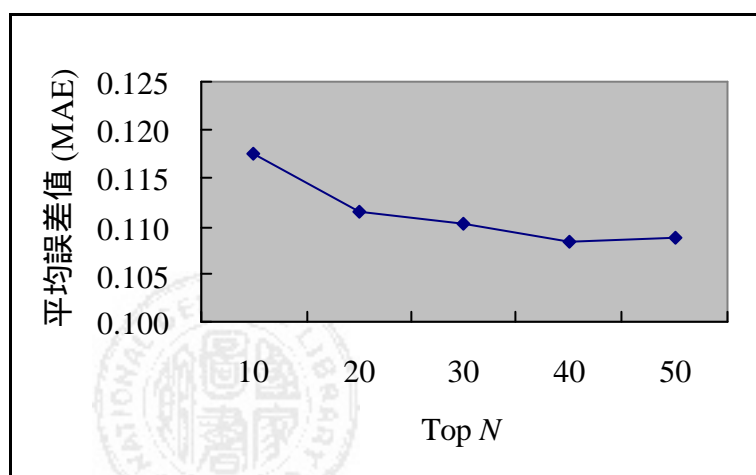


圖 21 Top N 的平均誤差值

排序分數度量值的結果如圖 22 所示，在 Top 10 得到最佳值 0.83，之後排序分數度量值隨推薦增多而降低，在 Top 50 得到最低值 0.69，變動的範圍約在 0.15 以內。

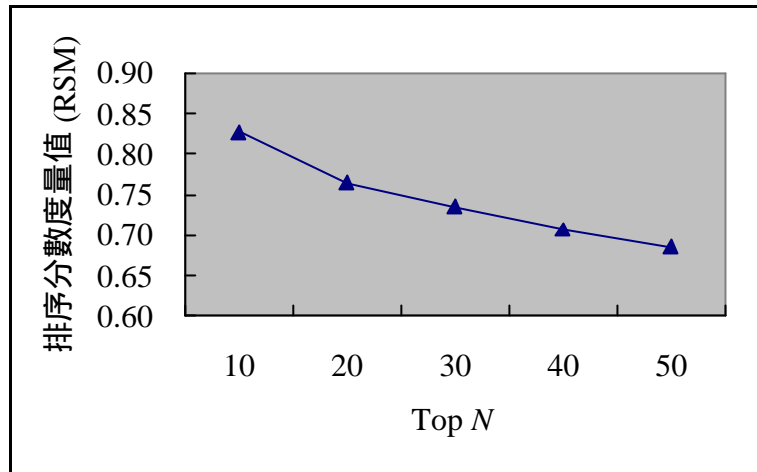


圖 22 Top N 的排序分數度量值

圖 23 中的結果顯示讀者平均滿意度大約都維持在 0.67 左右,介在滿意度普通 (0.5) 和滿意 (0.7) 之間。當推薦數多時有稍微降低的趨勢,可是變動並不劇烈,這代表我們的推薦中不管排序前後,讀者喜好程度都差不多,這個現象同時反應出優缺點。優點是推薦的館藏讀者普遍喜歡,而缺點是推薦排序較高的推薦並沒有得到比較高的滿意度。

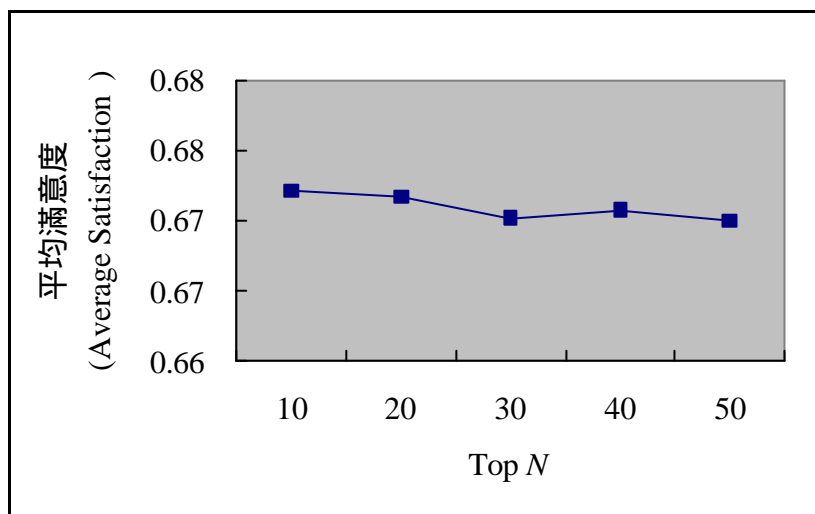


圖 23 Top N 的讀者平均滿意度

從以上的平均誤差值、排序分數度量值和平均滿意度這三種評估的結果,可看出我們所提出的圖書館個人化館藏推薦系統普遍符合讀者的興趣,達到預期的效果。

第六章 結論與未來研究方向

第一節 結論

本篇論文結合了協力式過濾和內容導向過濾，提出了一個適合圖書館環境的圖書館個人化館藏推薦系統，並實作在交通大學浩然圖書館的 myLibrary 個人資訊環境上。

在推薦方法的應用上，協力式過濾方面以乏析式資料探勘找出讀者間的關聯性，使興趣相近的讀者能彼此分享資訊，互相推薦館藏。而內容導向過濾方面則建立一個適合圖書館館藏環境的系統模型，分別找出館藏的屬性向量表示式及讀者的興趣向量表示式，使系統能有效地依照讀者興趣來推薦館藏。

圖書館個人化館藏推薦系統實作在 myLibrary 上，同時提供了以下幾個功能：

1. 根據讀者興趣提供推薦：先由協力式過濾找出推薦清單，再以內容導向過濾排序，能有效找出符合讀者興趣的館藏。
2. 讀者自訂興趣：讀者可依系統提供的興趣選擇功能改變自己的興趣，進而更新自己的個人興趣檔案，影響系統館藏推薦的結果。
3. 提供書評簡介：在推薦呈現的頁面中，有提供書評簡介的超連結，但因為圖書館內並沒有書評簡介的資料，所以我們以書籍的國際標準書碼 (ISBN) 或書名當關鍵字，到網路書店上查詢該推薦書籍的相關資料。當讀者對系統推薦的書籍沒有概念時，可以參考這些資料來判斷自己是否有興趣。

最後在系統評估的結果上，我們得到的平均誤差值、排序分數度量值和滿意度都達到不錯的效果，證明這個圖書館個人化館藏推薦系統確實可行，能帶給讀者一個新的管道來接觸圖書館的館藏。

在未來的數位圖書館中，圖書館將面臨到和網路商店類似的問題，例如網路商店缺乏店員來推銷產品，數位圖書館在網路上也沒有館員來協助讀者找尋資料。因此如何能將數位圖書館的功能做到完整，使讀者不需要實際館員協助就能順利使用圖書館的資源，或者如何能用新的技術在數位圖書館中建立虛擬館員模仿實際館員給予讀者協助，都是將來數位圖書館必須克服的挑戰。這裡我們提出的圖書館個人化館藏推薦系統，就是在這方面做努力，希望能幫助讀者找出圖書館中符合興趣的館藏，節省讀者搜尋的時間，增進圖書館與讀者的互動，使讀者能更方便使用圖書館資源。

第二節 未來研究方向

這個圖書館個人化館藏推薦系統，雖然在效能評估的結果上有不錯的結果，但是這其實只是推薦系統在圖書館上的初探而已，事實上仍有很大進步的空間。根據系統實作的心得及讀者回應的意見，我們認為還有以下幾點值得當作圖書館個人化館藏推薦系統以後改善的方向：

1. 增加探勘相同興趣讀者群的深度：在協力式過濾中，屬於同一群的讀者因為借的館藏類似，所以推薦的館藏會比較集中，有助於提高在協力式過濾的推薦分數。利用關聯規則探勘時，如果能盡量找出較長的頻繁項目集，使同一群的讀者數量能增加，則推薦出來的館藏種類會比較集中，更符合讀者興趣。
2. 讀者借閱書籍數量：當探勘讀者間的關聯規則時，我們聚集交易改以類別為交易基礎，並正規化項目出現次數為比率。但在事實上這種轉換可能會將一些原本在數量上可呈現的資訊遺失了。例如在類別 A、B 和 C 中，讀者 1 分別借過 2、3、5 本書，讀者 2 分別借過 20、30、50 本書，讀者 3 分別借過 20、20、20 本書，若從比例上來看讀者 1 和讀者 2 具有相當高的興趣相關，但從數量上來看則讀者 1 和讀者 3 的興趣可能比較相似。所以如果能將這些數量所代表的意義也能納入關聯規則探勘時的考量，將可使探勘出來的結果更具意義。

3. 同好的興趣相似強度：目前我們在協力式過濾中，會把每個同好在推薦中所佔的份量當作相同，沒有考慮到讀者與每個同好間的興趣相似強度應該不盡相同。當我們在找出相同興趣的讀者群時，其實可以發現每個讀者群的支持度不同，代表其興趣相似度是不同的。如果能將這些資訊加入協力式過濾中，應該可以得到更準確的推薦結果。
4. 增加內容導向過濾的分類：目前在內容導向過濾的分類有 1000 類，但由於圖書館館藏並不是平均分配在每個分類中，所以有些分類館藏特別多，造成目前使用的分類不足以區分該分類下的館藏。如果可以增加分類的數量，甚至只針對某些館藏較多的分類下增加分類數量，使內容導向過濾能有效區分該分類下館藏的不同，將有助於提高內容導向過濾推薦的準確度。
5. 西文圖書推薦：現在因為內容導向過濾中採用的分類是中國圖書分類法，所以無法將以美國國會圖書分類法分類的西文圖書納入推薦。如果可以整合中國圖書分類法和美國國會圖書分類法，找出彼此分類互相對應的方法，則系統就可以增加西文圖書的推薦服務。
6. 已閱讀過或擁有的圖書：給讀者的推薦基本上都是不在讀者借閱歷史中的館藏，可是事實上有些推薦雖然不在讀者的借閱歷史中，但讀者卻早已閱讀過或擁有了。所以系統應該提供選項使讀者可以消除這些推薦，甚至我們可以收集讀者對這些已閱讀過或擁有的圖書的喜好程度，這些資訊可以更有效反應讀者的實際興趣，將有助於系統判斷讀者的興趣。
7. 個人興趣選擇：系統目前供讀者選取個人興趣的分類是以中國圖書分類法為標準，可是事實上許多讀者根本不了解這種分類法，因為這種分類法不夠直覺。將來應該提供給讀者選擇的分類能盡量一般化、直覺化，然後再由系統將讀者所選擇的選項對應到中國圖書分類法中或者實際需要的分類表上。
8. 滿意度的選擇：在系統滿意度評估上，滿意度的預設值應設為空值，當讀者有親自選擇滿意度時才記錄讀者給予的值，避免影響評估的客觀性。另外，滿意度的選擇區間可以縮小，使讀者可以選擇更詳細的滿意度，例如現在有五種滿

意度可以選擇，每個滿意度間相差 0.2，將來可以改為有十個滿意度供選擇，每個選項之間差距 0.1。但提供給讀者選擇的滿意度間距仍不宜太小，避免造成讀者選擇上的困擾。提供恰當的滿意度選擇格式，能有助於得到更準確的系統評估數據，以當作之後系統改進的參考。

9. 考慮讀者興趣與時間之間的關係：有時候讀者興趣會隨時間改變而變更，這點可以從讀者的借閱歷史中察知。所以當我們建立讀者的個人興趣檔時，可以考慮借閱書籍與時間的關係，在不同時期借的書其重要性應該給予適當分配，例如最近借的書可能比較能反應讀者現在的興趣，可以給予較高的重要性比重。
10. 由回應滿意度調整興趣：讀者所回應推薦的滿意度都含有讀者的興趣在內，這些資訊可用來調整讀者的個人興趣檔，使個人興趣檔能更切合讀者興趣。

以上幾點都是目前的推薦系統中還未完成的工作，有些可以增加推薦的準確度，有些可以更準確判斷出讀者的興趣，如果能夠一一實作完成必定可以再次提升推薦效能，
? 讀者提供更適合的館藏推薦。

參考文獻

- [1] R. Agrawal, T. Imielinski, and A. Swami, "Mining association rules between sets of items in large databases," *Proceedings of the 1993 ACM SIGMOD international conference on Management of data*, Washington, D.C., United States, 1993.
- [2] R. Agrawal and R. Srikant, "Fast Algorithms for Mining Association Rules," *Proceedings of the 20th VLDB Conference*, Santiago, Chile, 1994.
- [3] C. Borgelt and R. Kruse, "Induction of Association Rules: Apriori Implementation," *14th Conference on Computational Statistics*, Germany, Berlin, Compstat, 2002.
- [4] J.S. Breese and C. Kadie, "Empirical Analysis of Predictive Algorithms for Collaborative Filtering," *Proceedings of the Conference on Uncertainty in Artificial Intelligence*, Madison, WI, 1998.
- [5] D. Cosley, S.K. Lam, I. Albert, J.A. Konstan, and J. Riedl, "Is Seeing Believing? How Recommender Interfaces Affect Users' Opinions," *Proceedings of the conference on Human factors in computing systems*, Ft. Lauderdale, Florida, United States, 2003.
- [6] J. Delgado and N. Ishii, "Multi-agent learning in recommender systems for information filtering on the Internet," *International Journal of Cooperative Information Systems*, vol. 10, no. 1-2, pp. 81-100, MAR-JUN 2001.
- [7] A. Kent et al., "Use of Library Materials: the University of Pittsburgh Study," PA.: Pittsburgh University, 1979.
- [8] R.D. Lawrence, G.S. Almasi, V. Kotlyar, M.S. Viveros, and S.S. Duri., "Personalization of Supermarket Product Recommendations," *Data Mining and Knowledge Discovery*, vol. 5, no. 1-2, pp. 11-32, JAN-APR 2001.
- [9] W.Y. Lin, S.A. Alvarez, and C. Ruiz, "Efficient Adaptive-Support Association Rule Mining for Recommender Systems," *Data Mining and Knowledge Discovery*, vol. 6, no. 1, pp. 83-105, 2002.
- [10] D.I Lin and Z.M. Kedem, "Pincer-search: an efficient algorithm for discovering the maximum frequent set," *IEEE Transactions on Knowledge and Data Engineering*, vol. 14, no. 3, pp. 553-566, May/Jun 2002.
- [11] W.P. Lee, C.H. Liu, and C.C. Lu, "Intelligent agent-based systems for personalized recommendations in Internet commerce," *Expert Systems with Applications*, vol. 22, no. 4, pp. 275-284, May 2002.
- [12] B. Mobasher, H. Dai, T. Luo, and M. Nakagawa, "Discovery and Evaluation of Aggregate Usage Profiles for Web Personalization," *Data Mining and Knowledge Discovery*, vol. 6, no. 1, pp. 61-82, 2002.

- [13] B.J. Mirza, B.J. Keller, and N. Ramakrishnan, "Studying Recommendation Algorithms by Graph Analysis," *Journal of Intelligent Information System*, vol. 20, no. 2, pp. 131-160, 2003.
- [14] D. Goldberg, D. Nichols, B.M. Oki, and D. Terry, "Using collaborative filtering to weave an information tapestry," *Communications of the ACM*, vol. 35, no. 12, December 1992.
- [15] T.P. Hong, "A fuzzy AprioriTid mining algorithm with reduced computational time," *The 10th IEEE International Conference on Fuzzy Systems*, vol. 1, Melbourne: Australia, December 2001, pp. 360-363.
- [16] J. Han and M. Kamber, *Data Mining: Concepts and Techniques*. Morgan Kaufmann, 2000.
- [17] T.P. Hong, C.S. Kuo, and S.C. Chi, "Mining association rules from quantitative data," *Intelligent Data Analysis*, vol. 3, no. 5, 1999.
- [18] T.P. Hong, C.S. Kuo, and S.C. Chi, "A fuzzy data mining algorithm for quantitative values," *Third International Conference on Knowledge-Based Intelligent Information Engineering Systems*, Adelaide, SA, Australia, December 1999, pp. 480-483.
- [19] J. Han, J. Pei, and Y. Yin, "Mining Frequent Patterns without Candidate Generation," *Proceeding of the 2000 ACM SIGMOD International Conference on Management of Data*, Dallas, Texas, United States, 2000.
- [20] S. Ko and J. Lee, "Discovery of User Preference through Genetic Algorithm and Bayesian Categorization for Recommendation," *Lecture Notes in Computer Science*, vol. 2465, pp. 471-484, 2002.
- [21] S. Ko and J. Lee, "User Preference Mining through Collaborative Filtering and Content Based Filtering in Recommender System," *Lecture Notes in Computer Science*, vol. 2455, pp. 244-253, 2002.
- [22] J. Pei, J. Han, H. Lu, S. Nishio, S. Tang, and D. Yang, "H-Mine: Hyper-Structure Mining of Frequent Patterns in Large Databases," *Proceeding of the 2001 IEEE International Conference on Data Mining*, San Jose, California, United States, 2001, pp. 441-448.
- [23] P. Resnick and R. Varian, "Recommender Systems," *Communications of the ACM*, vol. 40, no. 3, March 1997.
- [24] J.B. Schafer, J.A. Konstan, and J. Riedl, "E-Commerce Recommendation Applications," *Data Mining and Knowledge Discovery*, vol. 5, no. 1-2, pp. 11-32, JAN-APR 2001.
- [25] A.I. Schein, A. Popescul, L.H. Ungar, and D.M. Pennock, "Methods and Metrics for Cold-Start Recommendations," *Proceedings of the 25th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, Tampere, Finland, August 2002, pp. 253-260
- [26] M.J. Zaki, "Scalable Algorithms for Association Mining," *IEEE Transactions on Knowledge and Data Engineering*, vol. 12, no. 3, May/June 2000.

- [27] 陳莉君，「線上個人化參考文獻管理系統」，國立交通大學，碩士論文，民國九十二年。
- [28] 戴玉旻，「圖書館借閱記錄探勘系統」，國立交通大學，碩士論文，民國九十一年。
- [29] 吳安琪，「利用資料探勘的技術及統計的方法增強圖書館的經營與服務」，國立交通大學，碩士論文，民國九十年。
- [30] 楊雅雯，「個人化數位圖書資訊環境 - 以 PIE@NCTU 為例」，國立交通大學，碩士論文，民國九十年。