

書目探勘讀者使用圖書館之行為

謝建成

副教授

台灣師範大學圖書資訊學研究所

E-mail: jcshieh@ntnu.edu.tw

林湧順

台灣師範大學附屬中學

E-mail: zaire@ms49.hinet.net

摘要

在數位化的時代裡，資料彙整分析與資料探勘所探索發掘出的資訊，對單位政策制訂與決策建立是具有高度參考價值的。圖書館經營的目的是要能夠更符合讀者的需求，因此主動發掘讀者的需求，主動提供讀者所需要的資訊，是現今圖書館重要的工作項目。資料探勘技術於圖書館之應用—被稱之為書目探勘(bibliomining)，確實能有效協助管理者進一步了解讀者對圖書館之需求。書目探勘之資料資源主要以圖書館自動化系統之借閱記錄與館藏資料為軸心，不同於以往的是，本研究嘗試整合讀者個人多方面特性資料(非圖書館自動化系統之讀者資料)，並依據研究主題需求，萃取、彙整與轉換資料，建立相關資料倉儲，以「圖書」、「讀者」與「時間」三個面向透過資料彙整交叉分析與資料探勘技術之群集分析、分類分析與關聯規則分析等，探討讀者使用圖書館之行為。此結果可提供圖書館管理者在經營決策於館藏政策、圖書推薦、預算分配、圖書館管理等之重要的且客觀的參考依據。

關鍵詞：書目探勘，資料探勘，讀者行為

前言

面對現今多元服務的特質，傳統圖書館自動化系統(library automation system)技術或圖書館整合系統(library integrated system)技術，對支援圖書館的有效管理、正確決策、優質服務等，均不足以勝任，諸如流通櫃檯管理者

2006/10/11投稿; 2006/11/18修訂; 2006/12/05接受

對於工讀生人力支援的動態需求，能否及時且確切的掌握，以提升服務水準？圖書採購者對於讀者借閱圖書類別的行為模式，能否充分解析影響因素，以滿足閱讀需求？因此，為解決上述類似的困境，圖書館需要有效率的資料彙整分析技術與工具。Nicholson於2003年提出書目探勘(bibliomining-data mining for libraries)一詞定義利用資料探勘(data mining)及書目計量(bibliometric)工具於圖書館服務所產生之資料的應用，是為開創提供圖書館管理與服務資訊一新的研究領域。資料倉儲(data warehousing)與資料探勘技術，近幾年來已成功的運用在許多不同領域，諸如客戶消費行為分析、病患疾病關聯分析、網路行銷決策支援、異常統計資料探索等，但應用於圖書館的卻非常有限且零散。目前主要應用有圖書館使用者行為分析、自動化館藏發展機制、經費使用現況分析與預測、圖書館採購政策分析等。同時對於圖書館資料探勘資料的來源，因受限於圖書館相關服務作業資料的收集不夠完備周全，故仍以圖書館自動化系統資料庫及圖書館網頁紀錄檔為主，然而這對資料探勘技術的應用有相當大的缺失。因此在運用資料探勘技術於圖書館相關議題時，必須考慮引進涵蓋範圍更廣泛的可能資料，以期求得更完善的結果。

過去研究探討圖書館讀者之行為，無論以書目計量法或透過資料倉儲與資料探勘技術，無外乎都針對圖書館所能掌握的資料，像是圖書館自動化系統相關資料庫、網頁記錄檔等。然而，讀者行為的探討必須與其本身某些特性或特質息息相關，因此在探討圖書館讀者行為模式的同時，也必須考量涵括讀者個人特性資料，以求更準確了解讀者使用行為，進一步協助圖書館提升讀者服務品質。

本研究嘗試整合讀者個人多方面特性資料(非只有圖書館自動化系統之讀者資料)，並依據研究主題需求，萃取、彙整與轉換資料，建立所相關資料倉儲，透過資料彙整交叉分析與資料探勘技術之群集分析、分類分析與關聯規則分析等，探討讀者使用圖書館之行為模式。為順利驗證此嘗試，本研究的資料以國立台灣師範大學附屬高級中學之圖書館自動化系統資料庫為主，並加入了學生學籍資料(來自學務處)、學生成績資料與教師授課資料(來自教務處)等非圖書館記錄資料等。

二、資料探勘技術

資料探勘技術是近來相當熱門的研究領域，資料探勘被定義為：發掘潛藏在大量資料中可用或未知的資訊，以提供管理者決策參考的過程；此過程(Han & Kamber, 2006)包括如圖1所示：

(一)資料清理與整合(data cleaning and integration)

資料清理主要是處理資料中錯誤、遺失、不完整或不一致的情形；資料來

源若是多元的，則必須經過整合，通常我們會建立一資料倉儲完成此資料前置處理（data pre-processing）工作。

（二）資料選取與轉換（data selection and transformation）

了解領域知識，從資料倉儲中萃取分析工作所需的資料，建立目標資料集。資料轉換則是針對所萃取的整併資料進行資料格式的轉換、資料維度的彙整、資料概念的一般化、資料數值的標準化等。

（三）資料探勘（data mining）

利用不同的智慧型演算法，如類神經網路、基因演算法、決策樹、統計回歸、群集方法、關聯規則、線性分析等，探索關聯規則、廣義關連規則、分類規則、群集規則、分析趨勢、偵測離群值/例外值，以挖掘出有用或有興趣的資料特徵型樣。

（四）評估與展示（evaluation and presentation）

透過視覺化的工具或知識表現的技術，展現探勘的特徵型樣或模式，藉由領域專家分析詮釋資訊內容，以提供決策支援之用。

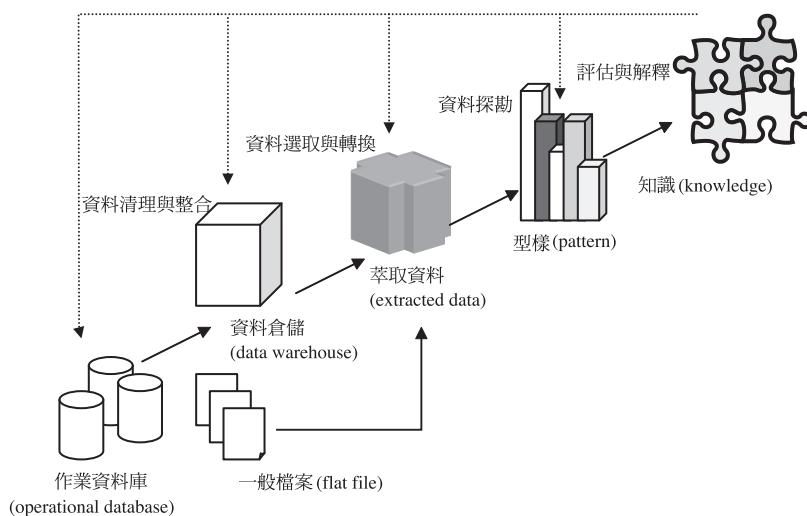


圖1 資料探勘步驟

三、資料探勘技術的應用

資料探勘技術的應用領域相當廣泛，舉凡科學、犯罪防治（犯罪行為與特質的相關研究）、命理（面相）、農業（經緯度與樹種之關係）、行銷、工業、商業、體育、財務、銀行、製造廠、通訊、電信業（顧客上線維持率與交叉銷

售)、網路相關行業(網頁上瀏覽路徑分析,提供網頁或網站瀏覽建議)、零售商、製造業、醫療保健、製藥業等都可以利用相關的技術支援分析決策。其主要研究探討議題包括:客戶行為分析、目標行銷、購物籃分析、客戶關係管理(customer relationship management)、銷售預測、產品行銷策略分析、時間與店區的多維分析、促銷活動的策略分析、顧客忠誠度分析、貸款償還預測和顧客信用政策分析、目標市場的顧客分類與群聚、犯罪防制等(廖虹雲, 2004; Jiao & Zhang, 2004; Chiang, Wang, Lee & Lin, 2004; 楊雅淨, 2003; 歐陽良全, 2003; 邱崇兼, 2003; 阮士峰, 2003)。應用的相關演算法則包括關聯規則分析、循序型樣分析、分類分析、群集分析、時間序列分析等。

四、資料探勘於圖書館應用

(一)書目探勘學

書目探勘學一詞是Nicholson於2003年首先提出,在此之前,有關資料探勘應用於圖書館領域的研究雖已有一些成果,但研究人員是以圖書館的資料探勘(data mining in library)稱之。書目探勘學的定義是指利用資料探勘及書目計量工具於圖書館服務所產生之資料的應用,以支援圖書館管理決策或調整圖書館服務。更進一步來說,書目探勘學是一種結合資料探勘技術、書目計量學、統計學的工具,為的是要在圖書館運作系統環境中,萃取出包括讀者的基本行為模式與需求、圖書館運作模式與資源運用等資訊,以提供圖書館管理者對於圖書館經營的決策參考。

Schulman (1998)研究提出,當圖書館已具有資料大或具特定用途之資料庫運作時,圖書館管理者應會考慮到建置一個決策支援系統。此時絕不可能利用傳統人工的方式來掌握不斷變動的使用者行為模式及趨勢,因此利用資料探勘技術來了解圖書館使用者的行為,就可讓圖書館重新規劃館藏發展方向與政策的制定,並設計圖書館相關活動。Guenther (2000)指出,為了要確定圖書館的服務與館藏是否合乎需要,就要先了解使用者的需求是什麼,為了要做出正確的館藏發展政策,就要先了解使用者的資訊需求。因此對於圖書館的管理者而言,是可以利用書目探勘來取得所需要的資訊。Nicholson (2003)提到,圖書館人員可以利用敘述統計方法或資料探勘技術,發掘出使用者的行為模式,了解這些行為模式將有助於決策較佳的館藏政策,更合理化圖書館的服務機制,制訂適切的網際網路時代圖書館服務項目,了解使用者使用圖書館的模式等。Papatheodorou 等人(2003)提到,藉由資料探勘的技術分析圖書館數位化資料,可以找出使用者的共同行為,以建立一套有意義的群組關係提升資訊獲取。而此分析的結果亦可以找出圖書管理重要的工作項目以及建議可行的工作方式,對於圖書館及其使用者將有助益於幫助管理者重新制訂符合各類使用者

需要的館藏政策、權威資料以及使用者介面，以達到最佳化服務，提供管理者決策的參考資訊，幫助使用者易於找出他有興趣的資訊提供個人化服務。

經以上敘述可以了解，藉由資料探勘技術的協助，圖書館可以獲得決策的支援，同時更加了解圖書館使用者的行為，以便制訂更符合使用者需要的服務項目與制度，而且資料探勘亦能讓圖書館重新規劃館藏的發展方向，並根據資料探勘所了解的使用者行為及所得到的資訊，合理分配圖書館的資源。

圖書館所能使用的資料探勘技術相當多，依據圖書館所需的管理目標及決策項目而有不同的使用方式，不過無論技術如何，原始資料才是資料探勘主要的基礎，選擇適當的資料作分析，才可以獲得正確且合於需求的資訊。而就目前資料探勘於圖書館的應用，其主要資料來源有：

1. 圖書館自動化系統資料庫

在Larsen(1996)、Mancini(1996)、Atkins(1996)及Peters(1996)等論文，均提到利用圖書館自動化系統資料庫的資料，可以獲取有利於圖書館管理的資訊，並得以提升管理績效以及制定管理決策項目。圖書館的自動化系統中，最值得進行資料探勘的就是圖書流通資料，因為圖書借閱記錄向來是讀者實際使用圖書館資源的「證據」，也是讀者滿足個人資訊需求的行為結果，其中潛藏大量圖書與讀者間互動的歷史紀錄，以及有意義的關係或規則。因此，圖書借閱記錄能反映讀者的實際資訊需求，對於掌握讀者興趣，進而作為加強圖書館資源利用之基礎，具有一定的參考價值。同時，為了解讀者特性與借閱間的關係，可以將讀者的興趣項目（如讀者興趣記錄檔）及特性項目（如讀者記錄檔中的性別、教育程度）涵蓋其中，以了解不同社群間的資訊需求是否有不同的趨向。

2. 電子資源使用紀錄

現今網路資訊發達，資訊量激增，同時電子資源也不斷的被開發出來，因此圖書館的經營方式不可能再將服務項目只集中於有形的書本之上，對此Banerjee(1998)提到，即使紙本資料未來仍是圖書館重要的館藏形式，但是電子形式出版品的資訊量將大大的增加，尤其電子資源是沒有時間與空間限制的，於是透過網路使用圖書館資料庫檢索、電子期刊等電子資源的使用者，將會是有增無減的。Peters(1996)指出電子資源使用記錄可以讓圖書館進行電子資源採購的效益評估，以作為圖書館經營者管理資訊系統的基礎，而以往圖書館員憑藉直覺與經驗來執行館藏發展工作，現在可以利用此一科學的方法協助進行。雖然圖書館的借閱記錄取得容易且資料量大，有利於資料探勘的進行，但是圖書館為了要提升全面性的服務品質，則不可忽略電子資源的紀錄統計資訊，讓電子資源的服務也能正確符合讀者的需要。而藉由圖書館網站上登錄記錄檔的探勘分析，圖書館決策者可以了解哪些電子資源使用率較高，哪些服務在網站上常被使用，以發現圖書館服務的主要項目，並提供足夠的資源以滿足

使用者的資訊需求。

3. 其他資料資源

除了上述的紀錄資料之外，Lawrence 等人 (1999) 以及 Ronald 等人 (2001) 均指出，圖書館可以利用研究論文中的參考文獻 (citation) 作資料探勘分析，以了解目前研究趨勢，以及確認新興的研究領域，同時也可以了解到哪些參考文獻頻繁的被使用。藉由以上的分析，除了可對研究的趨向有所了解之外，也可以利用所得的結果來作為研究領域的分類，將所得的結果作為讀者書目推薦的參考，因為這樣的資料是直接由研究者的研究資料所作的客觀分析，具有極大的參考意義。

(二) 書目探勘的步驟

Nicholson (2003) 定義書目探勘的步驟：

1. 決定焦點的範圍 (determining areas of focus)

此範圍可能是圖書館某一特定問題或一般決策所需。直接資料探勘是以某一問題為焦點，如「不同的系所或讀者對電子期刊的使用有何差別？」；非直接資料探勘即是反覆利用許多工具以找出可用的型樣 (patterns)，找出的型樣必須對圖書館員來說是有意義的才算是正確的。

2. 確定內部與外部資料來源 (identifying internal and external data sources)

書目探勘所需的是異動、非整合、低階層的資料。內部資料來源是指已存在於圖書館系統者，如讀者借閱資料；外部資料來源是指非存於現有圖書館系統者，如教師研究專長。因此如果為維護讀者隱私權而刪除日常作業資料，那將使書目探勘更加困難，而資料倉儲將可同時保護讀者隱私權與圖書館系統歷史資料。

3. 蒐集、淨化與匿名化資料形成資料倉儲 (collecting, cleaning and anonymizing the data into a data warehouse)

資料倉儲是源自於不同作業層系統，經過淨化與匿名化後，為分析格式化的一作業層資料庫，資料倉儲建立的過程中牽涉到要如何保有決策所需要的重要資料又要能保障讀者的隱私。圖書館員透過 SQL 語言，從原始作業層資料庫萃取所需的資料，以產生資料倉儲資料庫，此步驟是耗時的且循環不斷的。

4. 選擇適當分析工具 (selecting appropriate analysis tools)

一旦資料倉儲建立後，接下來便是分析工作。我們可以透過傳統的統計聚集工具，輕易的計算並產生許多不同的報表；或利用線上分析處理 (on-line analytical processing) 工具，透過交談方式呈現支援決策所需的資訊；或是利用人工智慧工具探勘龐大的資料庫找出有用的型樣。

5. 藉由資料探勘發現型樣 (discovery of patterns through data mining)

資料探勘工作主要有兩類：描述性 (description) 與預測性 (prediction)。描述性資料探勘的目標就是解讀過去至今的資料，發掘不同型樣的特徵，如不同讀者群的借書行為等。預測性資料探勘就是利用挖掘出來的已知解釋未知，預測未來或是說明現有的。

6. 分析和實作結果 (analyzing and implementing the results)

結果產出，無論是統計報表或決策模式，都必須經由熟悉領域的圖書館館員，以未經使用過的資料測試驗證結果來強化所得模式。圖書館員對所發現的結果要有一致的認知；型樣或許是以資料驗證從圖書館工作中所獲得的默慧知識。最後則是實作探勘的結果並觀察其隨時間的變化以調整原有模式。

(三)資料探勘技術於圖書館之應用

1. 關聯規則分析的應用

賴雨廷 (2002) 將圖書館中仍未被讀者借閱過的新書做一個推薦系統推薦。首先，要先了解什麼類型的讀者喜歡哪一種類的書籍，這部分的資訊稱為正規化關聯式規則。興趣性較低或多餘的規則將會刪除，以形成有效規則集合，而這些集合就是之後要作為推薦新書用的。針對新書推薦的規則，也定義了一種新的逼近方式來計算其感興趣的程度，以及能有效地建立推薦的計劃。王毓菁 (2002) 期望協助館藏資料能有效率地分類收藏，進一步的因地因時因人不同，提供館藏發展計畫一個有力的參考，以增進圖書資源的使用率及圖書經費預算的有效分配。因此由客戶關係管理的概念出發，根據不同館藏資料的分類，閱覽者的基本資料，及閱覽者借閱館藏資源的借閱記錄，使用關聯規則分析探勘圖書館讀者資訊需求的特徵，做為圖書館館藏發展時的參考依據。余明哲 (2002) 旨在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。因此利用關聯規則探勘技術，從讀者的借閱紀錄檔找出關聯規則。從這些探勘的結果，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傾向。接著利用推薦系統找出給讀者的推薦書目清單，再將推薦清單依照讀者興趣做個人化的排序，最後找出合適的館藏推薦給讀者。曹健華 (2002) 利用讀者記錄檔與館藏歷史借閱記錄檔，來產生關聯規則，以作為館藏推薦的基礎。藉由讀者借閱記錄，依個人館藏借閱習慣、興趣，分析讀者借閱的關聯規則，進而主動推薦相關館藏給讀者，除滿足讀者借閱需求外，更進一步提供讀者潛在興趣的館藏可供借閱資訊，以提升圖書館館藏的使用率，並增加讀者的滿意度。吳安琪 (2001) 在讀者記錄檔、館藏記錄檔與借閱記錄檔中找出讀者借閱館藏的共同性。利用讀者借閱館藏的共同性及讀者借閱館藏的順序，來達到以下的目標：吸引讀者到館借閱、提升館藏借閱率、提升讀者忠誠度、協助館藏複本採訪政策、促進館藏流通率。洪志

淵(2001)從圖書借閱資料庫中挖掘出讀者與圖書間的關聯規則，並交由圖書館專家詮釋規則上的知識，以運用於新書推薦。因此研究中不需要讀者在圖書館裡留下個人的喜好檔案，首先確認出讀者與圖書借閱行為有關且相互獨立的讀者屬性，再來提出三個演算法來找出最大項目集合(large item sets)，並做實驗來評量效率，除此之外，更訂出一套有興趣規則(interesting rules)的評量方法，以判斷讀者的興趣趨向，並根據讀者族群特性，應用在圖書館的新書推薦上。陳建銘(2001)利用關聯規則分析借閱記錄檔，找出讀者借閱書籍的偏好與借閱情況，以做為館藏發展計畫的參考，滿足學生需求，同時可以提高整體圖書館借閱率，也可提供圖書館作為其他決策工作時的之參考。戴玉旻(2001)藉由資料倉儲的技術處理圖書館借閱記錄的前置作業，以及運用關聯規則探勘演算法H-Mine的技術找出借閱館藏的關聯性，並以H-Mine為基礎發展廣義相關規則探勘演算法及多重最小支持度廣義相關規則探勘演算法，以便找出借閱類別的關聯性。發掘出讀者社群關係後，運用這些成果達到以下目標：提供讀者借閱館藏的建議、推薦讀者新進館藏。

Kao, Hang & Lin (2003)利用圖書館圖書流通記錄檔，並配合資料探勘技術可以發掘過去年度裡，讀者的主要需求為何，以做為未來年度預算分配參考依據。Wu (2003)認為圖書館流通記錄檔中可以切合需要，以資料探勘技術可以發掘讀者實質需求，依此建立一應用的模式(data mining based model，簡稱DMBA)，讓圖書館在預算分配上有一個可茲依循的參考。卜小蝶(2002)以分類號第三層000~999作關聯規則分析，利用相似性比對方法，推估相似借閱行為所反映出的圖書類號，以做為圖書推介的依據。柯皓仁、楊雅雯、吳安琪、戴玉旻(2002)將圖書館的所有讀者視為一個大社群，從中了解成員在館藏借閱的共同性與順序性。然而，讀者背景與學科領域可能會影響到其借閱行為，因此先將讀者分群(如根據系所、學院、性別、年級等分群)，再針對每一群讀者探索其借閱的共同性和循序型，使資料探勘的結果更能切合讀者的需求。以館藏借閱的共同性來說，興趣相同的讀者往往會借閱類似的館藏，若能發掘出館藏借閱的共同性，當有某位讀者借閱某館藏時，便可推薦給他借過此館藏的讀者亦曾借閱的其他館藏。

2. 分類分析的應用

Neumann等人(2003)提到對於圖書館而言，讀者借閱行為推薦服務是非常大有可為的，以資料探勘技術對自動化系統中的圖書借閱記錄檔與讀者搜尋記錄檔進行資料的分析，如此一來便可輕的建立一個像亞馬遜(amazon.com)一樣的顧客導向式的入口網站。同時，讀者也可減少搜尋和評估資訊的時間，不僅是增進讀者服務效能，更可以對圖書館員在管理圖書館上有很大的助益。

3. 群集分析的應用

曾勇森(2002)將資料探勘技術應用在圖書館書籍與讀者之間，主要目的是為書籍找尋適性化之讀者，以及為讀者找尋適性化之書籍。首先將書籍，或讀者，透過相似度的計算，將所有的書籍或讀者分成多個群組，群組內的書籍或讀者具有高度相似性，而群組間具有高度的差異性，再利用群組內成員的借閱型態趨勢，找尋該群組最適性之讀者或書籍並加以推薦之。卜小蝶(2001)探索借閱記錄中隱藏的重要規則，包括圖書與讀者、讀者與讀者、圖書與圖書間的關係。圖書館欲主動推薦相關新書或特定讀者輸入關鍵詞查詢出相關圖書時，可根據借閱記錄，分析有興趣的類號，並透過分類號群集、相似系所分類號群集，以及重要分類號與系所關聯等，進一步將圖書重新排列加以推薦。

4. 次序相關分析的應用

曾勇森(2002)利用時間順序的觀念，將所有書籍的借閱次序先行予以排列，求出滿足最小支持度之高頻項目集合，再將所有高頻項目集合進行最大化次序之確認，完成最大化確認後之高頻項目集合，即是所有書籍之借閱次序集合。此時只須找出各書籍適合之讀者或讀者適合之書籍，並予以推薦，即完成利用次序相關分析中找尋書籍適性化之讀者，或找尋讀者適性化之書籍並加以推薦之動作。吳安琪(2001)改進次序相關分析演算法，找出讀者借閱館藏的順序。利用讀者借閱館藏的共同性及讀者借閱館藏的順序，達到以下的目標：吸引讀者到館借閱、提升館藏借閱率、提升讀者忠誠度、協助館藏複本採訪政策、促進館藏流通率。

Neumann等人(2003)利用亞馬遜網站推薦系統來探討應用在科學圖書館的使用者分析上。藉由所發展的「推測性可能購買模式」，並運用再次購買的理論推測模式，作為科學圖書館使用者行為模式的分析，將讀者讀書借閱順序做一分析，以得到使用者行為基礎模式。柯皓仁等人(2002)則針對讀者可能會依據一定的順序來借閱(例如先借入門，再借進階)，因此若發現許多讀者都按照一定的順序來閱讀某些館藏，那麼當有某位讀者借閱這些館藏中的某一本時，便可建議他按照順序來閱讀相關書籍。

就以上對於書目探勘的應用與研究分析，我們初略將資料探勘技術應用於圖書館領域整理歸納如表1。

表1 資料探勘技術應用於圖書館領域之彙整表

資料探勘技術	應	用	範	疇
關聯規則分析	* 找出讀者個人特性與圖書之間的關聯性			
	* 利用讀者相似的特性推薦圖書			
	* 將同質性圖書推薦給適當的讀者			
	* 探求讀者資訊需求特徵，做為圖書館館藏發展參考依據			
	* 圖書館預算分配			

- | | |
|--------|---------------------------|
| 分類分析 | * 建立讀者圖書間之特性模式以做為推薦圖書參考 |
| | * 建立顧客導向之服務 |
| 群集分析 | * 探討使用者的集群特性及其借閱行為的傾向。 |
| 次序相關分析 | * 依據讀者圖書借閱的順序，推薦給其他未借閱之讀者 |
| | * 尋找書籍適性化之讀者或讀者適性化之書籍 |
| | * 圖書館業務人力資源的安排 |
-

五、以書目探勘探討讀者使用圖書館之行為

就專業分類來看，圖書館運用資料探勘技術主要領域包括：圖書館讀者行為分析、館藏發展分析、經費使用分配分析與預測、圖書館採購政策分析、讀者個人化服務、讀者社群需求分析等。以圖書館業務範疇與資源服務，資料探勘技術的應用可以更為廣泛且深入地探究如何將資料探勘技術運用於圖書館，以提昇圖書館作業效率與服務品質。而資料探勘的基礎來自於原始作業資料（operational data），圖書館自動化系統資料庫圖書館網頁紀錄檔，可說是目前圖書館相關運作所產生的大量資料主要根源。然而圖書館無論是關於服務讀者或處理業務所衍生的相關資料，其間之關聯性甚強且牽涉範圍甚廣，有異於其他領域的應用，因此在應用資料探勘技術於圖書館之同時，確定有意義的相關資料資源是非常重要的。以下本研究將以高中生使用圖書館行為模式為例，說明相關資料資源的重要性，並驗證藉由考量更多相關資料資源於書目探勘，對使用者在行為上的分析與探討，將會獲致更多可供參考的資訊。

(一)研究對象

本研究的對象為國立台灣師範大學附屬高級中學（簡稱附中）。附中在北區屬於明星高中，為一完全中學（包括高中部跟國中部），歷來就以開放的校風聞名，學校並不全然只為升學，或埋首於書本，強調的是五育均衡發展，學生有相當的自主性來發展自己的興趣與專長，因此學生的特性是相當多元的。而且就地域來說也是相當的多元，並不會過於侷限某些地區或某些程度的學生。這樣的學生素質，使得本研究擁有多樣的空間，不至過於單調。針對附中圖書館92年8月至93年7月的借閱記錄檔，可以看到借閱記錄為22,000餘筆，且往前推一年也有24,000餘筆，平均每天借書量為80餘冊，這對一般高中圖書館而言，借閱人次是相當的多。因此，針對資料探勘而言，資料量已經足夠，不至於因為缺少資料數量而使得探勘結果過於稀少而失去意義。

雖然在學校圖書館裡，借閱排行榜中前20名的書籍往往都是由文學類書集所霸佔，但是以基本的敘述統計方法所得的資料顯示，在附中圖書館的借閱記錄檔中，前20名不只出現文學類書籍，還包含有地理類、歷史類與自然科學

類。這就充分表現出附中學生的多樣性與不受拘束性。這將會使探勘的結果更加豐富與具有參考價值。

(二)書目探勘之資料資源

為探討中學生對圖書館使用行為，本研究所要探勘的資料是以附中圖書館自動化系統的借閱記錄檔、讀者記錄檔及館藏記錄檔為主，另外為了要了解學生是否會因個人特性造成借閱習慣有所差異，本研究將讀者特性合併於資料資源中。這裡的所謂的讀者特性包含性別、學習組別、學業表現、教師授課表、學籍資料等等，由圖2可以了解本研究所使用的資料資源範疇。

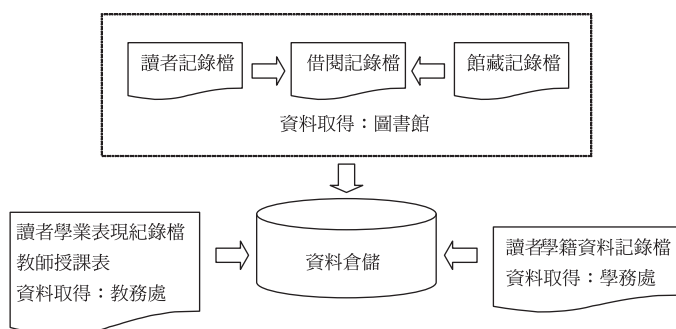


圖2 資料資源整合說明

(三)書目探勘主題

本研究應用資料探勘技術，針對借閱記錄檔、讀者記錄檔及館藏記錄檔為基礎，探索圖書館讀者的社群特性，並期望能運用資料探勘的成果來提昇圖書館的經營與服務，使圖書館扮演更積極的角色，不僅是被動的提供讀者資訊，而且能夠主動發現讀者的需求，並推估讀者潛在的需要，滿足讀者的資訊需求。因此本研究預計探討主題與使用方法如下：

1. 讀者社群與借閱行為模式分析

(1)敘述統計

以敘述統計方式對資料倉儲的資料進行分析，找出借閱率較高的書籍，或借閱率較高的讀者，並分析所代表的意義，以了解讀者行為模式的大趨向，並作為圖書館館藏發展與推薦圖書時的重要參考。

(2)分類分析

利用借閱記錄檔及讀者特性檔，分析讀者不同的社群間的借閱行為的差異性，以了解讀者的行為模式，這模式不僅可做為圖書館的營運參考基礎，同時對於教學計畫以及輔導學習而言也是重要的參考依據，著實為重要的資訊，尤其在相當有限的圖書館預算中，採購符合各社群間所需要的館藏，將可使有限

的資源，作最充分的利用。

(3)孤立點分析

附中圖書館服務全體學生，因為師大附中學生的多元性，故對於特殊需要的學生也必須要有所照顧，或許有些學生，借閱行為不同於一般學生，會有比較屬於自己的借閱行為，研究對於這樣的學生發掘其特性，是相當有趣的，若能找到一些規則或是模式，對於特殊學生的輔導與推薦將會有所幫助。

2. 增進個人化服務工作

(1)關聯規則分析

最主要目的就是要作圖書推薦，以借閱記錄檔，找出讀者借閱圖書的共通性，藉此推薦相關圖書給讀者，例如借閱甲書的人通常都會再借閱乙書，則可依此推薦給借閱甲書的讀者，主動的提供服務，增加圖書的流通性。

(2)時間序列分析

讀者在借閱館藏時，可能會先借入門書籍再借深入探討的書籍，如果把讀者借閱館藏的順序特性找出，則下次有某讀者借入門書籍時，即可推薦他借閱深入探討的書籍，讓讀者很容易地知道這本館藏的進階書籍有哪些。

但就中學圖書館而言，無法像大學圖書館一般，可以有充裕的資源製作個人化的網路服務系統，因此這裡的個人化服務，所強調的是當讀者作參考諮詢時，可以適切的提供讀者所適合的資訊，或以個別的方式對某些特定的社群或人員不定期提供相關的參考資訊。

3. 館藏政策維護

利用群集分析，找出不同群集之間不同的借閱行為，發掘出每個群集間普遍出現的書籍類型，藉此來作為採購圖書的依據。

(1)圖書館政策制定參考

對於圖書館的經營時間與人力的分配，一直是大家所討論的事項，對於讀者而言，當然提供越多人力、開館時間越長，將會越滿意，但是對圖書館本身而言，人力是極其有限的，因此將人力與時間發揮到最適當的程度，是圖書館管理者最需要思考的方向。

(2)敘述統計

以敘述統計方法對於館藏記錄檔作分析，找出每天、每週，甚至每年，讀者最常借閱圖書的時段，藉此可以作為人力上的安排與調配。

(3)時間序列分析

對於開館時間是利用時間序列分析，找出是否於每週、每月，甚至於每季、每年中讀者使用圖書館的時間規律性，一旦規則可以找出，將可作為圖書館開館時間延長或縮短的參考，這樣的資訊尤其在寒、暑假，將更顯重要，以資料探勘技術所得的資訊，將可說服學校決策單位提供適當的人力，同時對於圖書館館員也更可以接受開館的時間。

(四)書目探勘工具與使用平台

1. 資料探勘工具

本研究選定 SPSS Clementine 8.1 為書目探勘工具。根據 KDnuggets 2003 年 5 月的調查 (www.kdnuggets.com)：SPSS Clementine 是市場上最流行的資料探勘工具。Clementine 8.1 視覺化介面，可以讓資料探勘過程的互動式流程更為簡單。其存取資料包含純文字檔 (flat file) 及關聯式資料庫 (經由 ODBC)，所支援的資料探勘方法計有關聯分析模型 (association analysis)、序列分析模型 (sequence pattern analysis)、迴歸分析模型 (regression analysis)、群集分析模型 (cluster analysis)、決策樹分析模型 (decision tree analysis)、類神經網路分析模型 (neural network analysis)、資料縮減分析模型 (data reduction analysis) 等。

2. 資料資源

目前附中圖書館所使用的自動化系統稱為 TOTALS (Technology Opulent TRANSTECH Automated Library System)。這是台灣傳技公司自行開發的系統，目前已發展至第 2 版，因此簡稱為 T2 系統。目前 T2 系統可以在一般常見的作業環境下運作，不管是 Unix 環境或微軟 NT 系列均可運作。T2 系統共包含六大模組：管理、編目、流通、WebPAC、期刊、採訪等。本研究選用資料探勘的資料內容包括借閱記錄檔、讀者記錄檔，編目模組裡的書籍記錄檔等。同時為發掘讀者不同社群間不同借閱行為與差異，因此加入讀者的特性檔。這些讀者特性檔則源自教務處的學生成績系統，以及學務處的學生學籍系統，另外也包括了教務處裡的機密資料—教師授課表。

3. 資料探勘系統平台

資料探勘系統主要設備為一部筆記型電腦，硬體配備 Intel Pentium 4 1.6MHz CPU、512MB 記憶體、120GB 硬碟，軟體配備 Microsoft Windows NT 4.0 作業系統、Microsoft SQL Server 7.0 資料庫。

(五)資料探勘實施

1. 確定主題

第一步就是要確定書目探勘所要探究的主題是什麼。之前已經說明，本研究主要主題包含：讀者社群與借閱行為模式、增進個人化服務工作、館藏政策維護及圖書館政策制定參考。藉由更科學的方式探究這些主題，期望能更了解附中學生的資訊需求，除了可以讓附中圖書館的經營更符合附中學生的需要外，並且可以針對學生的需求，給予更適切的資訊或輔導學生學習的方向。

2. 選擇資料來源

本研究所使用的資料是以圖書館自動化系統的借閱記錄檔為主要資料來源，結合讀者記錄檔及館藏記錄檔，並將教務處成績資料庫的成績記錄檔及學

務處的學生學籍資料記錄檔結合於資料倉儲之中。但這些資料並不是儲存於同一個資料庫，因此必須建立一個合於書目探勘時使用的資料倉儲，以利探勘工作的進行。

3. 建立資料倉儲

這步驟預計花費整個書目探勘80%的時間，且成功與否的關鍵亦在此步驟。因此對於資料倉儲的評估與建置，將會在實作資料探勘時，不斷的進行調整與改進，以期達到最適當的探勘結果。本研究是以附中圖書館的借閱記錄為資料來源，整合讀者記錄檔及館藏記錄檔，並於探勘工作需要時，適當的加入讀者學業表現記錄檔或讀者學籍資料記錄檔，作為社群分析的主要分群項目。這些資料必須先經過預備資料的過程，包含資料清理、資料轉換，再經由適當的資料探勘工具分析，發掘有意義的型樣。附中圖書館全館藏書十萬餘冊，借閱記錄檔每年超過兩萬筆，因此為了有效執行資料探勘，以及獲取最新的讀者資訊，本研究使用92年8月至93年7月的借閱記錄檔，作為分析資料的來源檔。資料的前置處理與資料轉換是根據自動化系統的借閱記錄檔轉檔出來所需要的欄位，包括流水號、書本號、讀者代號、借閱日期、借閱時間等。基於前述書目探勘主題，本研究建立如下雪花模式資料綱要 (snowflake schema)，並將原有來源資料轉換成相對應之資料倉儲，以利後續探勘之用。

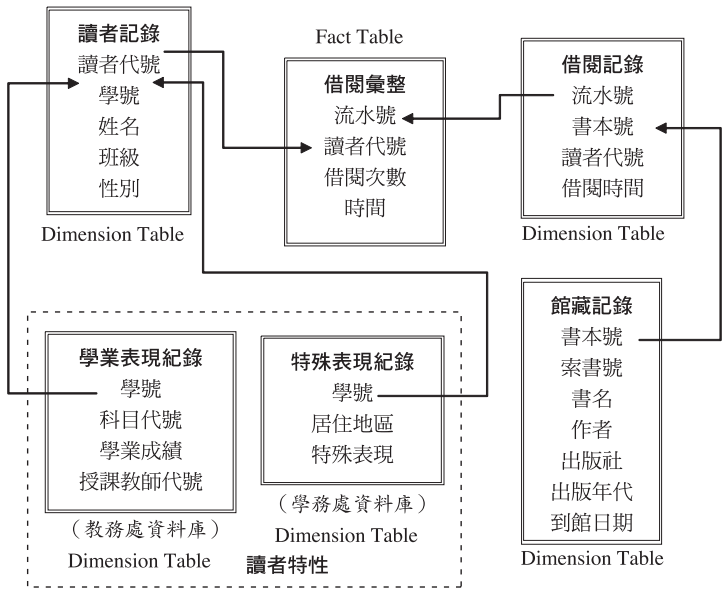


圖3 本研究資料倉儲資料模式

(六) 書目探勘結果與分析

1. 資料倉儲的統計分析

(1) 圖書大類

學生最常借閱是八類(文學類)的圖書，再來是三類(自然科學類)，其他則是七類(史地類)及九類(美術類)。寒暑假期間，八類圖書借閱的比例更高過其他月份，因此圖書館可針對此一趨勢，於寒、暑假期間針對八類圖書做圖書推薦的服務。

(2) 時間特性

在時間特性中，將時間分為「月份」、「星期」與「小時」三個項目，依據時間特性的統計分析，整理如表2。

表2 資料倉儲時間特性統計結果

分類項目	最多借閱	最少借閱
月份	12月、10月及3月	7月、8月及6月
星期	星期五	星期一
時段	12時、16時	9時、13時

就月份而言，在一年中以12月份的借閱次數最高，其次是10月、3月與4月，因此在這幾個月份，圖書館應該加強人力的支援。5月、6月及9月借閱次數也都相對較低，除了6月是因學年即將結束，提早圖書清點外，其他各月份應加強圖書館的行銷或舉辦圖書館利用等相關活動，讓讀者在這幾個月份裡，能多多利用圖書館的資源。而最少的月份是寒假的元月與暑假的7月、8月，是因沒有上課，借閱次數當然就明顯的少很多，這是無法避免的趨勢。不過學生在暑假空閒時間較多，可以多多鼓勵學生到圖書館借閱好書來閱讀。

就星期而言，星期五是整個星期裡借閱次數最高的一天。這應該是即將放假，學生準備要週末好好休閒或準備功課之用。因此圖書館應該加強人力的支援。星期一是整個星期裡借閱次數最低的一天。因為放假剛過，書籍還未看完，或尚無借閱圖書需要所致。

就時段而言，一天之中，12點是借閱次數最多的，而且比例高很多。因12點是午休時間，學生空閒時間較長，可以有比較多的時間到圖書館來借書，因此在人力編配上，要多加安排。另一個借閱時段比較高的是16點，剛好是放學時間，也符合了較長的休息時間，學生借閱的比例也較高。至於在9點及13點，因前個時段是早上剛到學校，後個時段是剛午睡起來，都還未進入狀況，於是借閱比例較低。

(3) 學生特性

依據學生特性的統計分析，整理如表3：

表3 資料倉儲學生特性統計結果

分 類 項 目		偏好借閱類別	
性別	男	八類、三類及七類	
	女	八類、七類及九類	
組別	一類組	男	八類、七類及五類
		女	八類、七類及九類
	二類組	男	八類、三類及七類
		女	八類、三類及七類
	三類組	男	八類、三類及九類
		女	八類、三類及九類
特殊班級	數理班	八類、三類及九類	
	音樂班	八類、七類及九類	
	美術班	八類、七類及九類	

就性別而言，在統計數據上，男生借閱的比例高於女生，比例大約是4：3，但附中學生男女生的比例是2：1，因此實際上女生借閱圖書的比例是比較高的。

就年級而言，二年級借閱的比例最高，其次是一年級，再次是三年級。這是因為二年級已經熟悉學校的環境，課業壓力也不大，加上老師對學生做作業的要求比較偏向報告方式，因此比例自然較高，三年級則因為課業壓力，所以借閱比例就少了很多。

在特殊班級，理班偏好八類、三類及九類，但是八類與三類的比例相差不多，是一個比較特殊的群組。是故以此結果做為圖書館推薦圖書及採購圖書的參考依據，來顧及少數的學生數群組。根據上述結果亦可以做為圖書館推薦圖書及採購圖書的參考依據。

(4)逾期分析

逾期還書對圖書館經營來說，是一個令人棘手的工作，因此針對學生逾期的狀況來分析，找出常逾期還書的特殊群組，可以在事前多做預防。

- a.附中學生逾期還書的比例為13%。
- b.男生女生逾期還書的比例也都一樣為13%。
- c.就類組而言，一、二、三類組的逾期還書比例分別為13%、14%、13%，相差不大。
- d.就月份、星期與時段而言，逾期還書比例都在11%～14%之間，比也都不是很明顯。
- e.就借閱圖書大類而言，0類（總類）及三類（自然科學類）逾期的比例較高，分別為16%及15%，四類（應用科學類）、五類（社會科學類）逾期的比例較低，分別為9%及11%。

根據以上結果可以得知，學生特性與逾期還書之間，並沒有太大的差異

性，這表示逾期還書是沒有特性之分的，屬於不預期事件。至於圖書類別方面也相差不多。

2. 資料探勘的結果分析

(1)時間特性與圖書大類間關係

a. 藉由群集分析可得知

12月份是各月借閱次數最頻繁的月份，所呈現的群集最多，而且多以八類（文學類）為主。每一天八類借閱的比例都相當高，尤其在每天12點的時段。星期五是每星期借閱次數最頻繁的一天，尤其每星期五的12點的八類圖書。

另外，星期一，10點的九類（美術類）圖書與星期三，9點的三類（自然科學類）的圖書都自成一個群集，表示這個時段所產生的借閱模式相當明顯。

根據以上所呈現的結果，可以作為圖書館推薦圖書時的一個參考依據，而且在借閱頻率比較高的時間點上，可以提高人員的配置，並針對特定時間作特定書籍的宣傳。

b. 藉由分類分析可得知

就各月份而言，分析中扣除八類（文學類）是最熱門的書籍，因為若不扣除八類圖書，則探勘結果將都會是八類圖書，相關分析結果如表4所示：

表4 分類分析時間特性探勘結果

分析類別	探 勘 結 果	
月份	6月	九類
	10月	七類
	其他月份	三類
星期	星期二	九類、三類
	其他日期	三類
一天	上午	七類及九類
	下午	三類

根據以上所呈現的結果，亦可做為圖書館圖書推薦時的參考依據。

c. 藉由關聯規則分析可得知

八類、三類、七類及九類是借閱行為中最常出現的圖書。就圖書借閱規則來看，借過九類再借過三類圖書之後，出現七類的機率很高。

就八類圖書而言，857（小說類）、855（散文類）、874（美國文學類）、861（日本文學類）、873（英國文學類）及876（法國文學類）最容易出現。

d. 藉由廣義關聯規則分析可得知

星期三，16時及15時，借閱三類（自然科學類）比例較高。星期四出現三類（自然科學類）的比例頗高。

另外，早上8點出現九類（美術類）的比例頗高。早上9點出現七類（史地

類)的比例頗高。星期一及星期四的16點出現七類(史地類)的比例頗高。這點結果跟分類分析所得的結果是一致的。

e. 藉由次序相關分析可得知

次序相關分析是以八類圖書來做探勘，因此可以得到以下結果。857、855、873、874、861都是交互出現的圖書。借閱857(小說類)與874(美國文學類)或855(散文類)與857(小說類)後，最常借閱的圖書還是857(小說類)。借閱861(日本文學類)之後，還會再借閱861(日本文學類)的圖書。873(英國文學類)與874(美國文學類)圖書會交互出現。根據以上的分析可以作為圖書館個人化服務的參考依據，同時也可以作為圖書推薦的一個依據。

(2)學生特性與圖書大類間的關係

a. 藉由群集分析可得知

所有學生均偏好八類圖書，二、三類組男生偏好三類圖書，數理班男生、女生均偏好八類、三類及四類，美術班及音樂班女生對於閱讀的興趣範圍較廣泛，男生則偏好八類。根據以上分析可以作為圖書館個人化服務的一個參考依據，同時也可以作為圖書推薦的一個依據。

b. 藉由分類分析可得知

就學生外在特性而言，所有學生均偏好八類，去除八類後，二、三類組學生偏好三類。而就八類圖書來看，857(中國文學類)是學生最喜愛的圖書。數理班學生對三類圖書喜好的程度相當於八類圖書。美術班及音樂班學生偏好八類及九類圖書。其他相關探勘結果，如表5、表6所示：

依類組探勘：

表5 分類分析學生特性類組探勘結果				
類組	探 勘		結 果	
一類組	一年級		六類(史地類)	
	二年級	男	一類(哲學類)	
		女	九類(美術類)	
	三年級		五類(社會科學類)	
二、三類組	一年級	男	330類(物理類)	
		女	308類(科學叢書類)	
	二年級		308類(科學叢書類)	
	三年級		330類(物理類)	

依學業成績探勘：

表6 分類分析學生特性學業成績的探勘結果

圖書類別	科目	探勘結果
八類圖書	國文成績	857類(中國文學類)，不因成績而有區別
	英文成績 高	874類(美國文學類)、876類(法國文學類)、873(英國文學類)
	一般	857類(中國文學類)
三類圖書	數學成績 高	一、二類組 312(算數類)
		三類組 330類(物理類)
	一般	312(算數類) 308類(科學叢書類)

上述分析結果是一個學生學習成績對於借閱行為的影響，國文類或許影響不大，因為八類圖書最吸引人的還是857(中國文學類)，但就英文成績而言，成績較好的學生對於外國文學類圖書還是有高度的興趣。另外，在數學成績表現上，成績較高的學生對於其他理科(物理、化學)的興趣也較高。

c.藉由廣義關聯規則分析可得知

表7 廣義關聯規則分析學生特性與各類圖書關係的探勘結果

圖書分類	探勘結果
大類分析 (扣除八類)	一類組女生 七類、九類
	三類組男生 三類
	所有學生 857(中國文學類圖書)
八類圖書	三類組女生 874(美國文學類)
	一類組男生 861(日本文學類)
	855(散文類)
三類圖書	數理班男生 330類(物理類)、340類(化學類)、310類(數學類)
	數理班女生 340類(化學類)、330類(物理類)、308類(科學叢書類)
九類圖書	美術班、音樂班 997類(智力遊戲類)
	995類(戶內遊戲類)

根據上述分析結果，可以了解到哪些類別的圖書適合哪些特性群組的學生，尤其特殊班級的學生，因為特殊班級學生對於自己的興趣需求有相當的認知，當圖書館有新書推出時，可以依循這樣的規則，讓讀者知道圖書館有新的資源，而這些資源或許正是他需要的資訊。

而就學生成績表現而言，以三類圖書來看，數學成績較高的學生偏好330類(物理類)、310類(數學類)及340類(化學類)圖書，成績中等的學生則偏好312類(算數類)圖書。英文及國文成績則不明顯。

另外，可針對任課教師來作進一步分析，其中編號為2及3的國文老師，班級學生借閱八類圖書最多，而八類圖書以855類（散文類）、857類（中國文學類）為主。編號為56、42、68及54的英文老師，班級學生借閱八類圖書偏好857類（中國文學類）、874類（美國文學類）及861類（日本文學類）圖書。針對編號為56、42、68及54英文老師的授課班學生來分析，英文成績較高的學生，除偏好857類外，對於874類（美國文學類）、876類（法國文學類）、873類（英國文學類）也有很高的興趣。依據上述的分析結果，可以發現，老師也會是影響學生借閱行為的一個變因。因此當老師推薦這些類別圖書時，是確有其需要的，圖書館應盡可能的配合採購，以滿足老師與學生的需求。

（3）時間特性與學生特性間的關係

男生借閱圖書的時間以12時為最多。女生則有類組之分，一類組是15時，二類組16時，而三類組則是12時。一年級學生星期二借書比例最高。數理班、美術班星期四借書比例最高，音樂班則是星期一借書比例最高。根據上述分析結果，可以知道特定學生在特定時間理借書的比例較高。根據這樣的規則，圖書館可以配合這些特定學生較常借閱的圖書來圖書推薦，像是看板宣傳或新書展示，可以讓學生獲得新的資訊。

六、圖書館管理決策建議

依據上述結果分析，對圖書館經營管理提出下列建議。

（一）圖書館經營

在借閱率較高的時間裡，須調配較多的人力資源，以應付學生的需要。另外，圖書館在做好書介紹時，不用再像以前一樣採用亂槍打鳥或擷取網路訊息的方式來進行，針對群集分析、分類分析以及廣義關聯規則分析所得的資訊，來做相關類別的好書介紹，並可針對特定群組做不同的好書介紹。這樣可以增加圖書館的服務品質，又不會浪費太多的人力與物力。

對於逾期還書的預防上，可以利用資料倉儲的分析與分類分析的預測模式，來做探勘。但是因本研究的數據，顯示逾期還書與學生特性間沒有顯著的關係，因此附中學生對於逾期還書的行為屬於不預期的事件，沒有特定的對象或特性可以預防，只能加強宣導，防止學生逾期還書。

圖書館在每年的12月、每週的星期五，及每天的12點及16點，需提高人力的調度，以維持正常的圖書館運作。逾期還書的比例相當高，因此要多加宣傳，以減少逾期還書的現象。針對一類（哲學類）、四類（應用科學類）、五類（社會科學類）及六類（史地類）的圖書多做宣傳，以提高圖書的利用率。

學生在借閱圖書時，除了八類之外，上午偏好三類圖書。這是屬於理工類

型的圖書，而下午偏好七類及九類圖書，這是屬於文史藝術類型的圖書。因此圖書館可以針對這樣狀況來作相關的新知提供或相關活動的宣導。

(二)館藏發展

首先，圖書館可以藉由分類分析的結果來開列採購書單，對大眾化的圖書做一個列表。另外，也可針對特定的學生採購符合他們需求的圖書，讓圖書採購經費花在最需要的圖書類別上。例如，八類圖書是學生最喜歡的圖書，而八類圖書中，又以857(中國文學類)、855(散文類)、874(美國文學類)、873類(英國文學類)等幾類的圖書更加有興趣，所以採購時就可針對這幾類更多著墨。另外，像美術班及音樂班的學生對於九類圖書有很高的興趣，因此在採購圖書上也應符合這些特定學生的需要。

在做館藏政策發展時群集分析、分類分析及關聯規則探勘，都可以演算出符合需要的結果，但是根據本研究的執行結果，分類分析是比較有多元性的一個演算方式，他可以針對各式的特性群組來做探勘，不像關聯規則探勘須要不斷的修正「支持度」與「信心水準」，才能有令人滿意的答案，或是像群集分析一樣，分析出來的群集結果，有時資料筆數會相差很多。

另外，分類分析可以針對某個特定族群的學生，再對各大類做進一步的分類與預測，讓研究可以更深入探討小群組間的差異性。

八類圖書中的874類(美國文學類)、873類(英國文學類)等類外國語文類圖書，具有提升英語學習成績的功效，更可多加採購。數理班學生對於三類圖書的需求大於八類的需求，這是一個特殊的群組，因此對於三類圖書要多加的充實其深度與廣度。為滿足美術班及音樂班學生的需求，應加強九類圖書的深度與廣度。三類圖書對於理科成績較好的學生有相輔相成的作用，因此對於三類圖書，可多加採購，以提升學生對於科學的興趣與認知。利用探勘結果中，各科任課教師班級學生借閱率較高之老師，商請他推薦相關圖書，以滿足該位老師的教學需求。

(三)個人化服務

根據關聯規則探勘、廣義關聯規則探勘，及次序相關分析，可以找出圖書與圖書之間的順序關係，或圖書與讀者間的關聯性，藉由這樣的關聯性探勘，圖書館可以了解到哪些資訊是相互有關係的，而這關聯性可以提供學生一個正確的資訊，讓學生了解到合於自己需求，例如三類組男生對於330類(物理類)、340類(化學類)及310類(數學類)圖書有特別的興趣，當有新書是屬於這些類圖書時，便可依循這樣的規則推薦給這類的讀者。這樣可以針對重點作宣傳，達到節省人力、物力，同時提升服務品質。

高級中學的圖書館礙於人力與資源的關係，無法從事一對一的個人化服務工作，但是可以對某些特定學生或特定對群組來做圖書推薦與宣傳，尤其當有學生來圖書館做館藏諮詢時，可以有一個即時且正確的資訊提供給需要的學生，讓學生了解到，自己的需求在哪裡，或合於自己需要的資訊是什麼樣的類型。主動出擊提供服務，這樣圖書館在校園裡才能夠確立更好的地位。

除八類外，一類組學生偏好七類及九類，二、三類組學生偏好三類及九類，因此在做圖書館推薦時，可朝這幾類對特定的對象來做推薦。數理班學生對於三類圖書的需求大於八類的需求，因此可以針對三類圖書多加推銷。美術班及音樂班可以多加推薦八類及九類的圖書。透過英文老師，可以讓英文成績較好的學生多多閱讀外國文學類圖書。根據關聯規則所探勘出來的模式，可以作為個人圖書推薦的參考，做成一個順序表格，公布於圖書館的相關位置，讓讀者可以了解到哪些圖書是有相關性的。例如857、855、874、873間都是相關聯的，可以依此為推薦圖書的參考。

七、結論與建議

依據研究結果發現，資料倉儲與資料探勘技術確實可運用於圖書館，探究不同圖書館相關之管理及服務議題，以協助圖書館增進讀者服務、改善管理績效、提升決策品質。但以目前圖書館現況而論，若要實施資料探勘技術於圖書館，將會遭遇到以下幾個問題。

(一)書目探勘的目標定義

運用書目探勘的目的是什麼？探討的範圍在哪裡？預期效益是什麼？所支援的資料是否足夠？這些都是在運用資料探勘技術於圖書館之前必須通盤考量的。圖書館資料探勘運用的目標，是管理者對規劃或管理圖書館所必須明確訂定的。分階段進行是為可行的方案：以相關讀者服務議題為優先目標，先就該圖書館讀者服務特性或使用行為有所了解，針對其內容提出各項業務措施，積極改善服務內容與品質，以提讀者滿意度；然後對圖書館作業範疇及其成本績效，運用資料探勘技術可以發掘出諸多管理盲點，如人力資源的有效運用等，確實增進作業流程效率與經費最佳使用。

(二)圖書館員使用的方便性

以目前資料探勘技術的運用，雖然有許多協助工具，諸如ETL(Extraction、Transfer、Loading)資料轉換工具彙整來自不同資料來源的各種形式資料，但因涉及過多資訊技術問題像是資料庫分析等，對圖書館員使用確實不容易，甚至於有障礙的。因此如何改善現有資料探勘運用的方便性，讓圖書館員或管理者

只專注於分析議題，而減少資訊技術的干擾，是為圖書館資料探勘運用重要課題之一。

(三)書目探勘的資料來源

圖書館資料探勘資料的來源，就目前來說，因為受限於圖書館相關服務作業資料的收集不夠完整，主要還是以圖書館自動化系統及圖書館網頁紀錄檔為主，然而這是非常不夠且欠缺的，諸如工讀生的資料、館際合作的資料、採訪廠商的資料等，圖書館無論關於讀者行為或作業方式，均異於其他領域的應用，不同資料間之關聯性甚強，因此建議圖書館必須重新檢討對讀者服務或作業管理相關資料的收集內容，為因應而後有效運用資料探勘於圖書館的目的。

參考文獻

- 卜小蝶(2001)。以圖書借閱記錄探勘加強圖書資源利用之探討。中國圖書館學會會報，66，59-72。
- 卜小蝶(2002)。使用者導向之圖書分類關聯分析研究。圖書資訊學刊，17，81-94。
- 王毓菁(2002)。圖書館閱覽者群組潛在特性探勘資訊系統。未出版之碩士論文，華梵大學工業管理研究所，台北縣。
- 余明哲(2002)。圖書館個人化館藏推薦系統。未出版之碩士論文，國立交通大學資訊科學研究所，新竹市。
- 吳安琪(2001)。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。未出版之碩士論文，國立交通大學資訊科學研究所，新竹市。
- 阮士峰(2003)。資料探勘在股市序列型樣的應用。未出版之碩士論文，國立東華大學資訊科學研究所，花蓮縣。
- 邱崇兼(2003)。以分群化技術探勘信用卡消費之異常交易。未出版之碩士論文，南台科技大學資訊管理研究所，台南縣。
- 柯皓仁、楊雅雯、吳安琪、戴玉旻(2002)。個人化及群體化圖書館資訊服務初探。國家圖書館館刊，91(1)，161-195。
- 洪志淵(2001)。圖書流通紀錄之一般化相關規則找尋之研究。未出版之碩士論文，國立中山大學資訊管理研究所，高雄市。
- 曹健華(2002)。應用資料探勘技術於數位圖書館之個人化服務及管理。未出版之碩士論文，南華大學資訊管理研究所，嘉義縣。
- 陳建銘(2001)。類神經網路於Web Mining之應用。未出版之碩士論文，國立台北科技大學商業自動化與管理研究所，台北市。
- 彭于萍(2005)。資料探勘應用於圖書館之探討。圖書與資訊學刊，50，76-91。
- 曾勇森(2002)。利用資料探勘技術增進圖書館之服務效益。未出版之碩士論文，南台科技大學資訊管理研究所，台南縣。
- 楊雅淨(2003)。資料探勘技術於綜合所得稅逃漏稅選查之應用。未出版之碩士論文，國立中正大學會計與資訊科技研究所，嘉義縣。
- 廖虹雲(2004)。利用資料探勘來預測顧客對不同產品類別之偏好程度。未出版之碩士論

- 文，朝陽科技大學資訊管理研究所，台中縣。
- 歐陽良全 (2003)。應用資料探勘技術於目標行銷之研究：以國內某醫院健康檢查中心為例。未出版之碩士論文，國立中正大學資訊管理研究所，嘉義縣。
- 賴雨廷 (2002)。利用資料探勘技術應用於圖書館新書推薦之研究。未出版之碩士論文，國立中山大學資訊管理研究所，高雄市。
- 戴玉旻 (2001)。圖書館借閱記錄探勘系統。未出版之碩士論文，國立交通大學資訊科學研究所，新竹市。
- 顏嘉惠 (2003)。資料探勘於圖書館行銷及顧客關係管理之應用。圖書與資訊學刊，42，58-68。
- Atkins, S. (1996). Mining automated systems for collection management. *Library Administration & Management*, 10(1), 16-19.
- Banerjee, K. (1998). Is data mining right for your library? *Computers in Libraries*, 18(10), 28-31.
- Chiang, D. A., Wang, Y. F., Lee, S. L., & Lin, C. J. (2003). Goal-oriented sequential pattern for network banking churn analysis. *Expert Systems with Applications*, 25(3), 293-302.
- Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). From data mining to knowledge discovery: An overview. In *Advances in Knowledge Discovery and Data Mining* (pp. 471-493). Massachusetts: Mit Press.
- Guenther, K. (2000). Applying data mining principles to library data collection. *Computers in Libraries*, 20(4), 60-63.
- Han, J., & Kamber, M. (2006). *Data Mining: Concepts and Techniques Reviews*. (2nd ed.). San Francisco: Morgan Kaufmann.
- He, J., Liu, X. T., Shi, Y., Xu, W. X., & Yan, N. (2004). Classifications of credit cardholder behavior by using fuzzy linear programming. *International Journal of Information Technology & Decision Making*, 3(4), 633-650.
- Jiao, J. X., & Zhang, Y. Y. (2005). Product portfolio identification based on association rule mining. *Computer-Aided Design*, 37(2), 149-172.
- Kao, S.C., Hang, H.C., & Lin, C.H. (2003). Decision support for the academic library acquisition budget allocation via circulation database mining. *Information Processing and Management: An International Journal*, 39(1), 133-147.
- Larsen, P. (1996). Mining your automated system for better management. *Library Administration & Management*, 10(1), 10.
- Lawrence, S., Giles, C. L., & Bollacker, K. (1999). Digital libraries and autonomous citation indexing. *IEEE Computer*, 32(6), 67-71.
- Mancini, D. D. (1996). Mining your automated system for system wide decision making. *Library Administration & Management*, 10(1), 11-15.
- Neumann, A., Geyer-Schulz, A., Hahsler, M., & Thede, A. (2003). An architecture for behavior based library recommender systems. *Information Technology and Libraries*, 22(4), 433-454.
- Nicholson, Scott. (2003). The bibliomining process: Data warehousing and data mining for library decision-making. *Information Technology and Libraries*, 22(4), 146-151.
- Papatheodorou, C., Kapidakis, S., Sfakakis, M., & Vassiliou, A. (2003). Mining user

- communities in digital libraries. *Information Technology and Libraries*, 22(4), 152-157.
- Peters, T. (1996). Using transaction log analysis for library management information. *Library Administration & Management*, 10(1), 20-25.
- Ronald, N. K., J. Antonio del Rio, James, Esther, A. H., Garcia, O., & Ana, M.R. (2001). Citation mining: Integration text mining and bibliometrics for research user profiling. *Journal of the Society for Information Science and Technology*, 52(13), 1148-1156.
- Schulman, S. (1998). Data mining: Life after report generators-libraries use this decision-support technique to chart a future course. *Information Today*, 15(3), 52.
- Wu. C. H. (2003). Data mining applied to material acquisition budget allocation for libraries: Design and development. *Expert Systems with Applications*, 25(3), 401-411.

Bibliomining User Behaviors in the Library

Jiann-Cherng Shieh

Associate Professor
Graduate Institute of Library & Information Studies
National Taiwan Normal University
Taipei, R.O.C
E-mail: jcsieh@ntnu.edu.tw

Yung-Shun Lin

The Affiliated Senior High School of NTNU
Taipei, R.O.C
E-mail: zaire@ms49.hinet.net

Abstract

The information discovered through data analysis and data mining can be great helpful for decision makings in organizations. For servicing users' complacence, libraries should actively explore the user needs and then provide them with required information. It is the critical task for libraries in this age. Bibliomining, data mining applied to library operations, can really assist in gasping patrons' requirements. Based on circulation data, the previous works have provided many suggestions about library management. In this research, we try to comprehend more data, not only circulation data but also various patrons-related data, to bibliomining their behavior in libraries. The results can be used as crucial and clinical information to aid libraries in collection policy making, material recommendation, budget allocation and other library management related issues.

Keywords: *Bibliomining; Data mining; User behavior*