

# 利用關聯規則發掘圖書館個人化之書籍推薦

## Using Association Rules to Find Personalized Book Recommendation for Library

陳垂呈\*

Chui-Cheng Chen

### 摘要

在本篇論文中，我們以讀者之借閱資料為探勘的資料來源，每一筆借閱資料包含有讀者曾經借閱過的書籍與其興趣度，並以某一讀者為探勘的目標，利用關聯規則 (association rules) 分別從下兩方面來發掘讀者個人化的書籍推薦：首先，我們只考量書籍是否出現在借閱資料中，設計一個探勘關聯規則的方法，且探勘出之關聯規則的前置項目組，必須被包含於此一讀者的借閱資料中，根據關聯規則所顯示出的特徵，我們可發掘此一讀者適性化的書籍推薦。再者，我們考量在借閱資料中包含有讀者對曾經借閱過之書籍的興趣度，設計一個探勘包含有興趣度之關聯規則的方法，且探勘出包含有興趣度之關聯規則的前置項目組，必須被包含於此一讀者的借閱資料中，根據包含有興趣度之關聯規則所顯示出的特徵，我們可發掘包含有興趣度之此一讀者適性化的書籍推薦。此探勘結果，對於圖書館在發掘個人化的書籍推薦時，可以提供非常有用的參考資訊。

關鍵字：資料探勘、關聯規則、借閱資料、書籍推薦

### Abstract

In this paper, we use readers' borrowing history records as the source data of mining. Each borrowing history record contains a reader ever borrowed books and the reader's the degrees of interest for those books. We let one reader as the target of mining, and use association rules to find the personalized book recommendations for the reader from two aspects, respectively. First, we only consider the books whether they are contained in borrowing history records or not, and propose a method to mine association rules whose antecedents are contained in the reader's borrowing history record. According to the characteristics of the association rules, we can find the adaptive book recommendations for the reader. Moreover, we consider the books with readers' interests in the borrowing history records, and propose another method to mine association rules with interesting degrees whose antecedents are contained in the reader's borrowing history record. According to the characteristics of the association rules with interesting degrees, we can find the reader's the adaptive book recommendations for considering his interests. The results of the mining can provide very useful information to find personalized book recommendation for library.

Keywords: data mining; association rules; borrowing history records; book recommendations

---

\* 南台科技大學資訊管理系副教授

(Associate Professor, Department of Information Management Southern Taiwan University of Technology)

## 壹、簡介

隨著資訊技術的發展，圖書館所提供的服務也愈來愈多樣化，包含電子圖書、網際網路內容擷取、儲存光碟、多媒體等服務，帶動了資訊儲存及檢索的新紀元。在資訊技術軟、硬體平台的支援下，圖書館已改變以往傳統搜尋資料的介面方式，配合網際網路的連結，使得圖書館有限的空間，卻擁有全世界豐富的館藏資源。但隨著資訊技術在軟硬體快速的發展，圖書館在管理與經營上是否能有效善用這些資訊技術，以創造出新的管理或服務讀者的方式，進而提昇圖書館有效的利用率，是圖書館管理者必須面對的課題之一。

在圖書館服務行銷的理念中，辜曼蓉(1999)曾指出圖書館的核心價值是顧客(讀者)和服務人員之間的互動，而館藏及資訊的傳播則扮演提昇圖書館之服務品質的輔助角色。因此，針對讀者在需求資訊的差異性下，善用個人化服務技術來調整圖書館的資訊服務，把最貼切的館藏資訊主動傳達給讀者個人，進而提昇館藏資料的利用率與圖書館的經營績效，是圖書館行銷重要的服務方式之一(卜小蝶，1998；Ou, Lin and Li, 2001)。

目前，在國內各大專院校有提供圖書館個人化資訊服務中，交通大學個人化數位圖書館資訊服務是最具有代表性的系統之一(湯春枝，2002; <http://mylibrary.e-lib.nctu.edu.tw/>)，其提供的服務有「智慧型圖書館查詢」、「個人化檢索」、「新書查詢」、

「新書列表」、「個人化推薦」、「個人借閱狀況」、「個人書籤」、「個人化桌面」及「使用者設定」等功能。在「個人化推薦」的功能中，其分別利用資料探勘技術(data mining techniques)中的關聯規則及次序相關(sequences)做為個人化書籍推薦的方法依據。因此，如何將最貼切的書籍推薦給讀者個人，是圖書館個人化資訊服務重要的功能之一。

每天在圖書館中均有相當大量的書籍被借閱，在這些被借閱的書籍資料中，往往隱藏著書籍之間的關聯性，例如讀者借閱了一本「電子商務」的書籍，我們會發現其中也會有借閱「網路行銷」書籍的傾向，或是一些與「電子商務」相關的書籍。因此，如何從累積數量龐大的借閱資料中，找出對讀者有用的資訊或其他知識，是圖書館管理者必須思考的問題之一。

資料探勘是從大量資料中找出潛在有用的資訊與知識，目前資料探勘技術已普遍地應用在各領域中(Han and Kamber, 2000)。在本篇論文中，我們以讀者之借閱資料為探勘的資料來源，每一筆借閱資料包含有讀者曾經借閱過的書籍項目與其興趣度值，並以某一讀者之借閱資料為探勘的目標，利用資料探勘技術中的關聯規則，從下兩方面來發掘此一讀者個人適性化的書籍推薦：

1. 在借閱資料中只考量書籍項目是否被借閱過：我們以某一讀者之借閱資料 $X$ 為探勘的目標， $X$ 為包含一個或以上之書籍項

目所形成的項目組，從借閱資料中探勘前置項目組包含於 $X$ 的關聯規則 $Y \rightarrow Z$ ,  $Y \subseteq X$ 、 $Y \cap Z = \emptyset$ ,  $Y$ 、 $Z$ 為包含一個或以上之書籍項目所形成的項目組。由於 $Y \subseteq X$ ，顯示關聯規則可表現出此一讀者的借閱傾向，我們從關聯規則所顯示出的傾向特徵，做為發掘此一讀者適性化之書籍推薦的依據。

2. 在借閱資料中考量書籍項目是否被借閱過及其興趣度：我們量化在借閱資料中包含有讀者對曾經借閱過之書籍的興趣度，並以某一讀者之借閱資料 $X$ 為探勘的目標， $X$ 為包含有興趣度之一個或以上的書籍項目所形成的項目組，從借閱資料中探勘前置項目組包含於 $X$ 的關聯規則： $Y \rightarrow Z$ ,  $Y \subseteq X$ 、 $Y \cap Z = \emptyset$ ,  $Y$ 、 $Z$ 為包含有興趣度之一個或以上的書籍項目所形成的項目組。由於 $Y \subseteq X$ ，顯示關聯規則可表現出此一讀者的借閱傾向，我們從關聯規則所顯示出的傾向特徵，作為發掘包含有興趣度之此一讀者適性化的書籍推薦。

我們根據所提出的方法，設計與建置一個讀者個人化之書籍推薦系統。此探勘結果，對圖書館在擬訂讀者個人化的書籍推薦，進而以主動積極的方式來吸引讀者到館借閱，可以提供非常有用的參考資訊。

## 貳、相關研究

資料探勘是從大量資料中找出潛在有用的資訊與知識，其可完成以下工作或更

多：分類(classification)、關聯規則、分群(clustering)、次序相關分析(sequential pattern analysis)、及預測(forecasting)等(Chen, Han and Yu, 1996)，其探勘結果對企業在從事行銷決策及市場預測等活動時，可以提供非常有價值的參考資訊(Berry and Linoff, 1997)。對於圖書館的書籍借閱而言，讀者往往必須在龐大的書籍資料中，找尋有興趣或想要借閱的書籍資料，而圖書館卻只能被動地等待讀者來借閱書籍。如此結果，不僅造成讀者搜尋書籍資料的困擾與不便，也造成書籍的借閱率不佳。

目前已有許多利用資料探勘技術於圖書館經營服務之應用的相關研究，其包含有陳慶瑄(2000)利用k-means的方法來找出學習社群，以支援電子圖書館之個人化服務。孫冠華(2003)利用關聯規則來提昇數位圖書館的個人化服務及管理。吳安琪(2001)提出利用資料探勘技術來發掘讀者的社群關係，進而達到吸引讀者借閱書籍，以提昇圖書館的借閱率與讀者忠誠度等目的。洪志淵(2001)利用資料探勘技術來找出讀者與書籍之間的一般化關聯規則，做為讀者之新書推薦的依據。張苑菁(2001)以模糊理論(fuzzy)與資料探勘技術來分析讀者的借閱資料，進而提供相關的書籍推薦給讀者參考。因此，如何利用資料探勘技術來提昇圖書館的經營與服務，已成圖書館管理及資料探勘技術應用重要的研究課題之一。

Agrawal, Imielinski and Swami (1993) 首先提出從交易資料中擷取關聯規則來顯示項目之間的關聯性，關聯規則的定義說明如下：假設 $I$ 是所有項目的集合， $T$ 是全部交易資料的集合，一筆交易資料 $T_j$ ， $T_j \in T$ ，是由一個或以上項目所組成的集合，稱之為項目組(itemsets)，若一個項目組包含有 $k$ 個項目，稱之為 $k$ -項目組( $k$ -itemsets)，以 $itemset_k$ 表示之， $k \geq 1$ 。在項目組 $X$ 與 $Y$ 之間有一關聯規則被表示成 $X \rightarrow Y$ ， $X、Y \subseteq I$ 且 $X \cap Y = \emptyset$ ，其中 $X$ 稱之為前置項目組(antecedent)，而 $Y$ 稱之為後置項目組(consequent)。有兩個參數 $s$ 與 $c$ 分別為支持度(support)與信賴度(confidence)，用來決定關聯規則是否成立。支持度 $s$ 的定義為：在所有的交易資料中，同時包含有 $XUY$ 的比率值，即 $s = (\text{包含有}XUY\text{之交易資料的數量}) / (\text{總交易數量})$ ；信賴度 $c$ 的定義為：在包含有 $X$ 的交易資料中，也同時包括有 $Y$ 的比率值，即 $c = (\text{包含有}XUY\text{之交易資料的數量}) / (\text{包含有}X\text{之交易資料的數量})$ 。擷取出來的關聯規則，其支持度與信賴度必須大於或等於所指定的最小支持度與最小信賴度，這樣的關聯規則才成立。

在擷取關聯規則的方法中，Apriori演算法(Agrawal and Srikant, 1994)是最具代表性的方法之一，其探勘過程主要分成以下兩個階段：首先，找出滿足最小支持度的所有項目組，這些滿足最小支持度的項目組稱之為高頻項目組(frequent itemsets)，若某 $k$ -項目組滿足最小支持度，

即稱之為高頻 $k$ -項目組(frequent $_k$ -itemsets)， $k \geq 1$ ，以 $frequent_k$ 表示之；然後，根據前階段所找出的高頻項目組及以最小信賴度為條件，計算出所有符合的關聯規則。例如ABC為高頻3-項目組，假如關聯規則 $AB \rightarrow C$ 滿足最小信賴度，則此關聯規則成立。探勘關聯規則的相關研究可參考(Park, Chen and Yu, 1997; Srikant and Agrawal, 1995)。接下來，我們說明Apriori演算法擷取關聯規則的步驟如下：

- (1) 找出 $frequent_{k-1}$ ， $k > 1$ ，若為 $\emptyset$ ，則停止執行。
- (2) 由(1)中組合任兩個有 $k-2$ 項目相同的 $frequent_{k-1}$ ，形成 $itemset_k$ 。
- (3) 判斷由(2)所找出的 $itemset_k$ 其所有包括的 $itemset_{k-1}$ 之子集合是否都出現在步驟(1)中，若成立就保留此 $itemset_k$ ，否則就刪除。
- (4) 再檢查由步驟(3)所擷取的 $itemset_k$ 是否滿足最小支持度，假如符合就成為 $frequent_k$ ，否則就刪除。
- (5) 計算 $frequent_k$ 所形成的關聯規則，若滿足最小信賴度，則關聯規則成立。
- (6) 跳至步驟(1)找 $frequent_{k+1}$ ，直到無法產生高頻項目組為止。

傳統關聯規則只考量其項目是否出現，即關聯規則中的各項目數量都為1，但考量現實的交易情形往往是包含了數量關係，因此包含有項目數量之關聯規則的相關研究就被提出(陳彥良、凌俊青和許秉瑜，2001；Srikant and Agrawal, 1996)，

在簡單關聯規則中的各項目包含有數量的關係，以 $m A \rightarrow n B$ 的形式表示之， $m$ 、 $n \geq 1$ ，其表示購買 $m$ 單位的A，也會有傾向購買 $n$ 單位的B，關聯規則中的項目其數量不一定為1，如此的表示方式將更符合現實中的交易情形。我們說明擷取簡單關聯規則的步驟如下(陳彥良、凌俊青和許秉瑜，2001)：

- (1) 找出 $frequent_{k-1}$ ， $k > 1$ 。
- (2) 由步驟(1)中找出任兩個有 $k-2$ 個項目相同的 $frequent_{k-1}$ ，包括 $frequent_{k-1}$ 與本身，組成 $itemset_k$ 。
- (3) 判斷由步驟(2)所找出的 $itemset_k$ ，其包含的 $itemset_{k-1}$ 之子集合，是否都出現在步驟(1)中，若成立就保留，否則就刪除。
- (4) 再檢查由步驟(3)所找出的 $itemset_k$ ，若滿足最小支持度，即成為 $frequent_k$ ，否則就刪除。
- (5) 對 $frequent_k$ 計算所有可能形成的關聯規則，若滿足最小信賴度者，即成為有效的簡單關聯規則。
- (6) 跳至步驟(1)找 $frequent_{k+1}$ ，直到無法產生高頻項目組為止。

在本篇論文中，我們以讀者之借閱資料為探勘的資料來源，並以某一讀者之借閱資料探勘的目標，利用關聯規則分別從下兩方面來發掘讀者個人化的書籍推薦：一是只考量書籍項目是否被借閱過；二是考量書籍項目是否被借閱過及其興趣度。藉由從關聯規則所顯示出的傾向特徵，以

分別做為發掘此一讀者個人適性化之書籍推薦、及發掘包含有興趣度此一讀者個人適性化之書籍推薦的依據。

## 參、發掘讀者個人適性化之書籍推薦

我們以讀者之借閱資料為探勘的資料來源，只考量書籍項目是否出現在讀者的借閱資料中，以某一讀者為探勘的目標，探勘前置項目組包含於此一讀者之借閱資料的關聯規則，做為發掘此一讀者適性化之書籍推薦的依據。接下來，我們將說明如何發掘某一讀者適性化之書籍推薦的探勘方法，並以一實例做說明。

### (一) 探勘個人化之書籍推薦的關聯規則

假設目前欲探勘之某一讀者的借閱資料為 $X$ ， $X$ 為包含一個或以上之書籍項目所形成的項目組，我們必須找出以下形式的關聯規則：

$Y \rightarrow Z$ ,  $Y \subseteq X$ 、 $Y \cap Z = \emptyset$ ， $Y$ 、 $Z$ 為包含一個或以上之書籍項目所形成的項目組， $Y \cup Z$ 是高頻項目組。

其顯示出的借閱傾向為：若讀者曾經借閱過 $Y$ 書籍，則也會有借閱 $Z$ 書籍的傾向。由於 $Y \subseteq X$ ，表示關聯規則可表現出此一讀者的借閱傾向特徵，且若 $Y$ 愈相似於 $X$ ，則關聯規則愈能反應出此一讀者的借閱傾向特徵，其借閱 $Z$ 書籍的傾向性也愈強。因此，我們藉由關聯規則的傾向特徵，可做為發掘此一讀者適性化之書籍推薦的依據。

爲了配合探勘的需要及避免計算與X無關的項目組，我們修改Apriori演算法，直接組合X中的項目與非X中的項目而形成項目組，並判斷這些項目組是否爲高頻項目組，探勘的過程說明如下：

- (1) 從X中及非X中找出 $frequent_1$ ，而且必須至少各包含有一項。
- (2) 由步驟(1)中，組合包含於X中之任一 $frequent_1$ 與非包含於X中之任一 $frequent_1$ 而形成 $itemset_2$ ，檢查 $itemset_2$ 是否滿足最小支持度，假如符合就成爲 $frequent_2$ ，否則就刪除。
- (3) 找出所有的 $frequent_{k-1}$ ， $k > 2$ 。
- (4) 由步驟(3)中，組合任兩個有 $k-2$ 項目相同的 $frequent_{k-1}$ ，形成 $itemset_k$ 。
- (5) 判斷由步驟(4)所找出的 $itemset_k$ ，其所包括的所有子集合 $itemset_{k-1}$ 是否都有出現在步驟(3)中(若 $itemset_{k-1} \cap X = \emptyset$ 、或 $itemset_{k-1} \cap \text{非}X = \emptyset$ ，則不列入考慮)，假如成立就保留此 $itemset_k$ ，否則就刪除。
- (6) 檢查由步驟(5)所找出的 $itemset_k$ 是否滿足最小支持度，假如符合就成爲 $frequent_k$ ，否則就刪除。
- (7) 計算 $frequent_k$ 所形成的關聯規則，其形式爲：  
 $Y \rightarrow Z, Y \subseteq X, Y \cap Z = \emptyset, \{Y \cup Z\} \in frequent_k$ 。
- (8) 跳至步驟(3)繼續找出 $frequent_{k+1}$ ，直到無法產生高頻項目組爲止。  
 從以上演算法的步驟(2)開始，我們所擷

取出的 $frequent_k$ ，必定爲 $frequent_k \cap X \neq \emptyset$ ，計算高頻項目組所形成的關聯規則 $Y \rightarrow Z$ ，且 $Y \subseteq X$ ，若滿足最小信賴度，則關聯規則成立。我們藉由關聯規則 $Y \rightarrow Z$ 所顯示出的借閱傾向，做成以下定義：

關聯規則 $Y \rightarrow Z$ 成立，且 $Y \subseteq X$ ，則Z爲此一讀者適性化的書籍推薦，且Y愈相似於X，則借閱的傾向性也愈強。

## (二) 實例說明

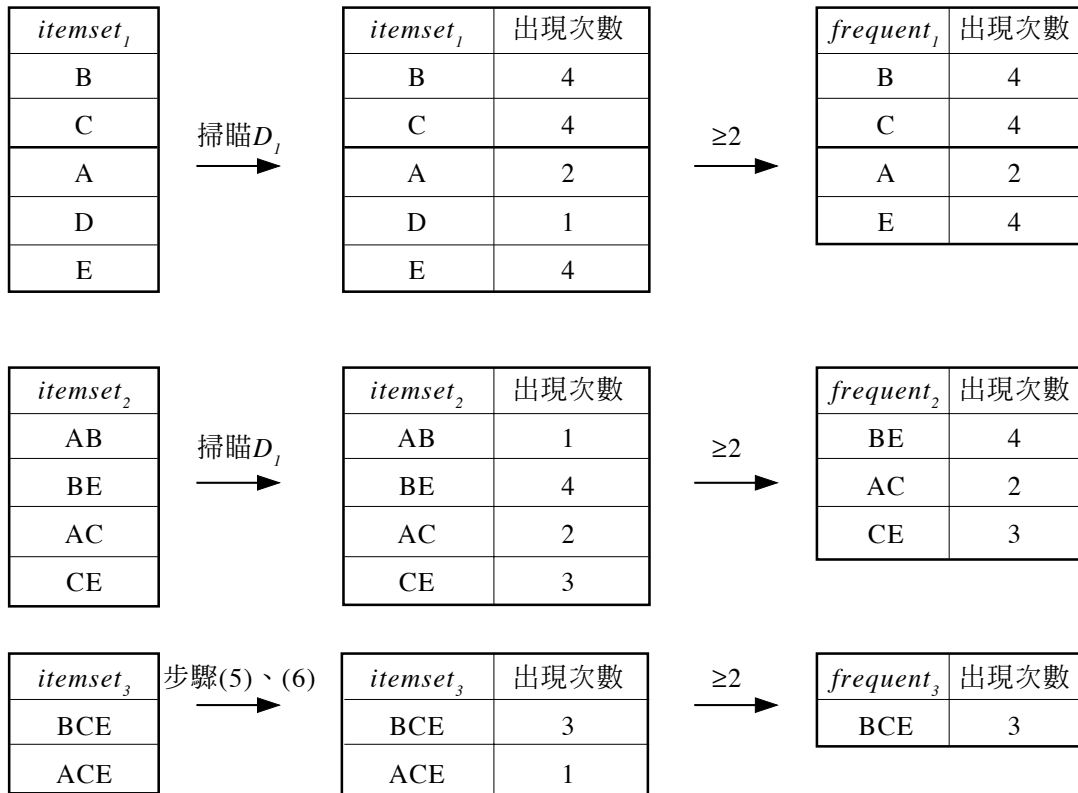
我們以表一之借閱資料庫 $D_1$ 爲例，來說明發掘某一讀者個人適性化之書籍推薦的探勘過程。 $I = \{A, B, C, D, E\}$ 爲所有書籍項目的集合， $R = \{R_1, R_2, R_3, R_4, R_5\}$ 爲5筆讀者之借閱資料的集合，最小支持度爲40% (即最小支持數量爲2)，最小信賴度設爲50%。假設目前欲探勘之某一讀者的借閱資料爲BC，以下我們說明發掘此一讀者適性化之書籍推薦的探勘過程。

我們以高頻3-項目組BCE爲例，計算形成的關聯規則 $BC \rightarrow E$ ，其信賴度爲 $3/3 = 100\%$ ，滿足最小信賴度，關聯規則成立。因此，藉由關聯規則 $BC \rightarrow E$ 所顯示出的借閱傾向特徵，可發掘E爲此一讀者適性化的書籍推薦。

表一 借閱資料庫 $D_1$

借閱資料	書籍項目
$R_1$	ACD
$R_2$	BCE
$R_3$	ABCE
$R_4$	BE
$R_5$	BCE

擷取包含有BC中任何項目之高頻項目組的過程如下：



無4-項目組。

### 肆、考量包含有興趣度之讀者 個人適性化的書籍推薦

讀者借閱過書籍後，經由閱讀的過程，必能更客觀的顯示出讀者對書籍的需求性與興趣度。因此，若能在讀者歸還借閱書籍時或是在具有個人化圖書館的功能中，藉由以詢問、填寫或問卷等方式來記錄其對借閱書籍的興趣度值，將曾經借閱過之書籍的興趣度也列入考量，則可探勘

出更精確之書籍項目彼此的關聯性。

在目前提供有個人化資訊服務的圖書館系統中，都將讀者個人借閱狀況列為基本的服務功能之一，例如交通大學個人化數位圖書資訊服務(PIE@NCTU)，此功能可列出讀者曾經借閱過的書籍項目。因此，只要對此服務功能稍做修改，讓讀者可以填寫其已借閱且歸還之書籍的興趣度值，即可蒐集、記錄到讀者對其曾經借閱

過之書籍的興趣度值。例如，我們將興趣度值共分為1到5等五個等級，若一讀者 $R_1$ 之借閱資料為A-書籍且興趣度值為4，而有另兩位讀者 $R_2$ 及 $R_3$ ，其借閱資料也都為A-書籍且興趣度值分別為1及3，雖然讀者 $R_1$ 、 $R_2$ 及 $R_3$ 都曾經借閱過A-書籍，但其對書籍的興趣度值卻不同，相較計算結果，顯示讀者 $R_1$ 與 $R_3$ 之間的興趣度是較 $R_1$ 與 $R_2$ 之間來得相似。

我們以讀者之借閱資料為探勘的資料來源，每一筆借閱資料包含有曾經借閱過之書籍項目與其興趣度的量化值，例如，假設{A, B, C, D, E}為5本書籍項目，而興趣度的量化值為1, 2, 3, 4, 5等5個等級，若一讀者之借閱資料為{1A, 2B, 3C}，則分別表示此一讀者對A-書籍的興趣度值為1、B-書籍的興趣度值為2、及C-書籍的興趣度值為3。我們仍以某一讀者之借閱資料為探勘的目標，在考量書籍興趣度下，探勘前置項目組包含於此一讀者之借閱資料的關聯規則，以發掘具有興趣度之此一讀者適性化的書籍推薦。接下來，我們說明如何發掘包含有興趣度之某一讀者適性化的書籍推薦的探勘方法，並以一實例做說明。

#### (一) 探勘包含有興趣度之個人化書籍推薦的關聯規則

假設目前欲探勘之某一讀者的借閱資料為 $X$ ， $X$ 為包含有興趣度之一個或以上的書籍項目所形成的項目組，對於發掘包含有興趣度之讀者個人適性化的書籍推薦，

我們必須找出以下形式的關聯規則：

$Y \rightarrow Z, Y \subseteq X, Y \cap Z = \emptyset, Y, Z$ 為包含有興趣度之一個或以上的書籍項目所形成的項目組， $Y \cup Z$ 是高頻項目組。

其顯示出的借閱傾向為：若讀者曾經借閱過 $Y$ 書籍，則也會有借閱 $Z$ 書籍的傾向。由於 $Y \subseteq X$ ，表示在考量書籍興趣度下，關聯規則可表現出此一讀者的借閱傾向特徵，且若 $Y$ 愈相似於 $X$ 及興趣度值愈大，則關聯規則愈能反應出此一讀者的借閱傾向特徵，其借閱 $Z$ 書籍的傾向性及興趣度也愈強。因此，藉由關聯規則顯示出的傾向特徵，可發掘包含有興趣度之此一讀者適性化的書籍推薦。

在探勘的過程中，我們將借閱資料中的書籍項目其數量若大於1，則以重複此項目來表示，例如借閱資料為2A3B，則以AABBB的形式做為探勘計算之用。以下是我們為了配合探勘的需要及避免計算與 $X$ 無關的項目組，而修改陳彥良、凌俊青和許秉瑜(2001)所提出之擷取簡單關聯規則的演算法，只探勘前置項目組必須包含於此一讀者之借閱資料的關聯規則。探勘的過程說明如下：

- (1) 從 $X$ 中及非 $X$ 中找出 $frequent_1$ ，而且必須至少各包含有一項。
- (2) 由步驟(1)中，組合包含於 $X$ 中之任一 $frequent_1$ 與非包含於 $X$ 中之任一 $frequent_1$ 而形成 $itemset_2$ ，檢查 $itemset_2$ 是否滿足最小支持度，假如符合就成為 $frequent_2$ ，否則就刪除。



- (3) 找出所有的  $frequent_{k-1}$ ， $k > 2$ 。
- (4) 由步驟(3)中，組合任兩個有  $k-2$  項目相同的  $frequent_{k-1}$ ，包括  $frequent_{k-1}$  與本身，形成  $itemset_k$ 。
- (5) 判斷由步驟(4)所找出的  $itemset_k$ ，其所包括的所有子集  $itemset_{k-1}$  是否都有出現在步驟(3)中(若  $(itemset_{k-1} \cap X) = \emptyset$ 、或  $(itemset_{k-1} \cap \text{非}X) = \emptyset$ ，則不列入考慮)，假如成立就保留此  $itemset_k$ ，否則就刪除。
- (6) 檢查由步驟(5)所找出的  $itemset_k$  是否滿足最小支持度，假如符合就成為  $frequent_k$ ，否則就刪除。
- (7) 計算  $frequent_k$  所形成的關聯規則，其形式為：  
 $Y \rightarrow Z, Y \subseteq X, Y \cap Z = \emptyset, \{Y \cup Z\} \in frequent_k$ 。
- (8) 跳至步驟(3)繼續找出  $frequent_{k+1}$ ，直到無法產生高頻項目組為止。

從以上演算法的步驟(2)開始，我們所擷取出的  $frequent_k$ ，必定為  $frequent_k \cap X \neq \emptyset$ ，且在步驟(4)中組合  $frequent_{k-1}$  與本身所形成的項目組，可以讓相同的項目重複化，藉此產生書籍項目的興趣度值。我們計算高頻項目組所形成的關聯規則  $Y \rightarrow Z$ ，且  $Y \subseteq X$ ，若滿足最小信賴度，則關聯規則成立。我們藉由關聯規則  $Y \rightarrow Z$  所顯示出的借閱傾向，做成以下定義：

關聯規則  $Y \rightarrow Z$  成立，且  $Y \subseteq X$ ，則  $Z$  為包含有興趣度之此一讀者適性化的書籍推薦。當  $Y$  愈相似於  $X$ ，則借閱的傾向性也

愈強，且  $Z$  包含之興趣度值愈大的項目，是更能貼近讀者興趣度的書籍推薦。

在書籍推薦的過程，也可將興趣度做為篩選書籍推薦的條件值，以找出符合興趣度條件值的書籍推薦。例如關聯規則  $2A \rightarrow 1B$  及  $2A \rightarrow 3C$  都成立， $\{A, B, C\}$  為書籍項目，由興趣度值的大小，可發掘  $C$  較  $B$  是更能貼近讀者興趣度的書籍推薦，若我們將興趣度條件值設為 2，則可發掘只有  $C$  是符合興趣度條件值之適性化的書籍推薦。

## (二) 實例說明

我們以表二之借閱資料庫  $D_2$  為例，來說明發掘包含有興趣度之某一讀者適性化的書籍推薦的探勘過程。 $I = \{A, B, C, D, E\}$  為所有書籍項目的集合， $R = \{R_1, R_2, R_3, R_4, R_5\}$  為 5 筆讀者之借閱資料的集合，興趣度值共分為 1 到 5 的五個級數，設定最小支持度為 50% (即最小支持數量為 2.5)，最小信賴度設為 70%。假設目前欲探勘之某一讀者的借閱資料為  $2B1C$ ，以下我們說明發掘包含有興趣度之此一讀者適性化的書籍推薦的探勘過程。

表二 借閱資料庫  $D_2$

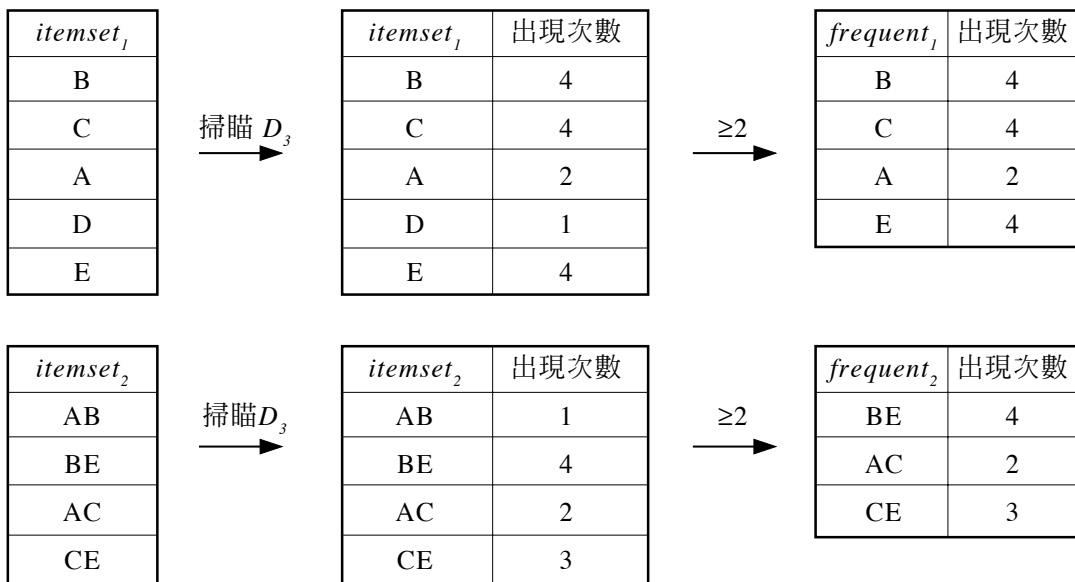
借閱資料	書籍項目
$R_1$	2A1C3D
$R_2$	2B2C1E
$R_3$	1A3B1C2E
$R_4$	2B1E
$R_5$	2B1CE

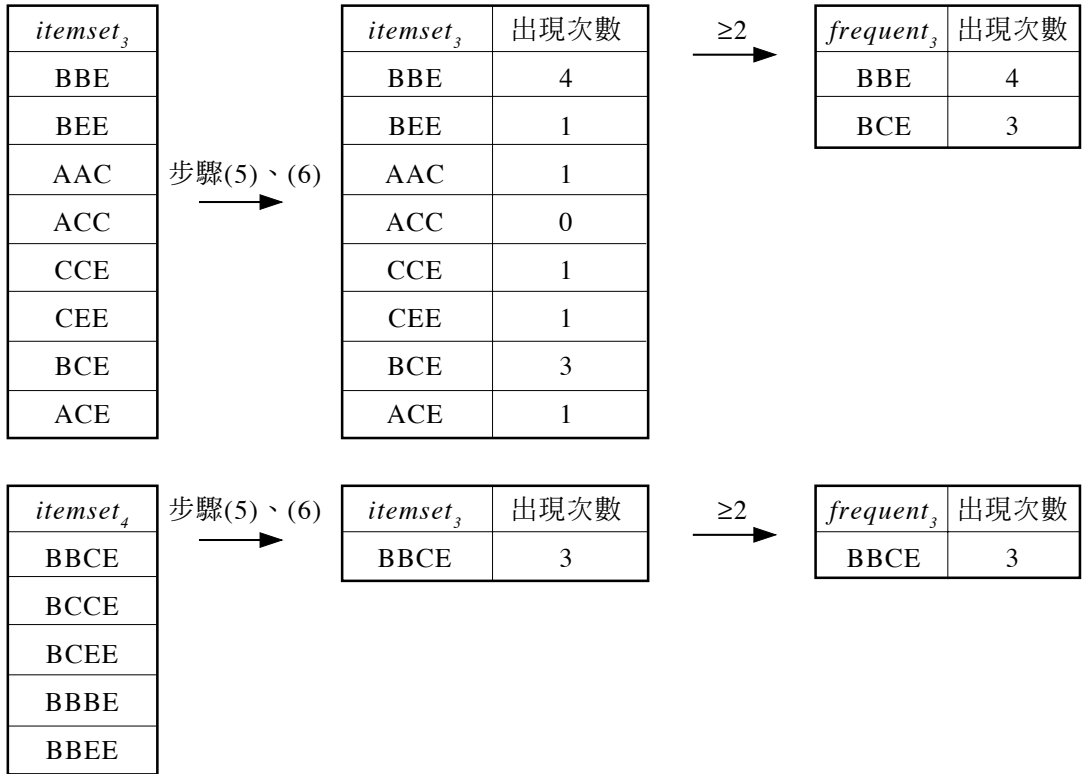
首先，我們將各借閱資料轉換成表三的借閱資料庫 $D_3$ 。

表三 借閱資料庫 $D_3$

借閱資料	書籍項目
$R_1$	AACDDD
$R_2$	BBCCE
$R_3$	ABBBCEE
$R_4$	BBE
$R_5$	BBCE

擷取包含有2B1C中任何項目之高頻項目組的過程如下：





無5-項目組。

我們以高頻4-項目組BBCE為例，計算形成的關聯規則2B1C→E，其信賴度為3/3 = 100%，滿足最小信賴度，關聯規則成立。因此，藉由關聯規則2B1C→E所顯示出的借閱傾向特徵，可發掘E為包含有興趣度之此一讀者適性化的書籍推薦。

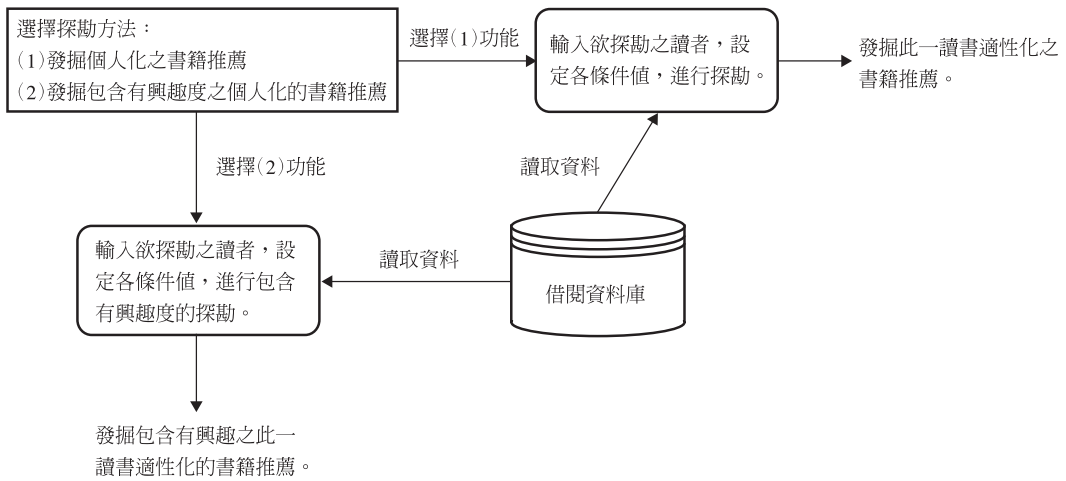
## 伍、讀者個人化之書籍推薦系統與效能評估

我們利用文中所提出的方法，應用到探勘讀者個人化之書籍推薦的系統實作上，系統探勘過程的模型如圖一。

我們以C#為撰寫的程式語言，在不失

一般性的條件下，假設書籍項目全部有26項，分別以A, B, C, ..., Z來表示之，在借閱資料中的每一書籍項目，分別以1, 2, 3, 4, 5等五個級數來量化讀者對書籍的興趣度，以亂數隨機產生每一讀者的借閱資料，每一筆借閱資料最多包含有10個書籍項目，共產生500筆借閱資料。例如借閱資料{1A,2B,3C}，表示曾經借閱過A, B, C等三本書籍且興趣度值分別為1, 2, 3。圖二為系統探勘的借閱資料，包含有「讀者編號」與「曾經借閱過的書籍」等欄位資料。

圖三為系統探勘的執行畫面，其中包含有兩項功能選項：「發掘個人化之書籍推



圖一 系統探勘過程模型

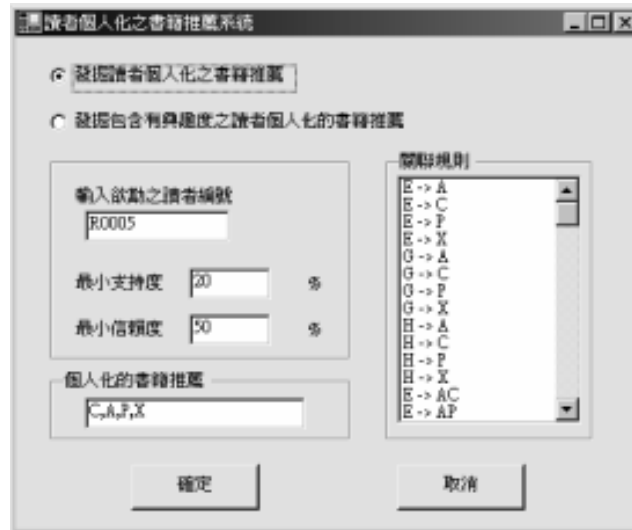
讀者編號	曾經借閱過的書籍
R0001	1A,3C,2E,1F,4K,2P,4Q,1Y
R0002	2G,1I,2K,3M,2U,1Z
R0003	1C,2E,1I,2Q,4U
R0004	2C,3D,2E,1F,4H,5M,3N
R0005	1E,2G,2H
R0006	2B,1C,3D,1E,2F,2H,1I,1K
R0007	2B,3C,1D,4G,2H,1L,3P,4R
R0008	2B,4G,2H,1I,2N,1Q,2V
R0009	1C,2D,4E
R0010	2B,3D,1J

圖二 借閱資料

薦」及「發掘包含有興趣度之個人化的書籍推薦」。假設目前點選「發掘讀者個人化之書籍推薦」功能，在「輸入欲探勘之讀者編號」欄位中填入欲探勘的讀者編號，並分別輸入「最小支持度」及「最小信賴度」等數值。經由文中所描述之發掘讀者個人適性化的書籍推薦的探勘方法，可在「關聯規

則」欄位中顯示出找到的關聯規則，並在「個人化的書籍推薦」欄位中顯示書籍推薦的結果，欄中位於愈前面的書籍項目，表示借閱的傾向性愈強，如圖三。

若點選「發掘包含有興趣度之個人化的書籍推薦」功能，經由文中所描述之包含興趣度的讀者個人適性化書籍推薦的探勘方



圖三 發掘讀者個人化之書籍推薦的執行畫面

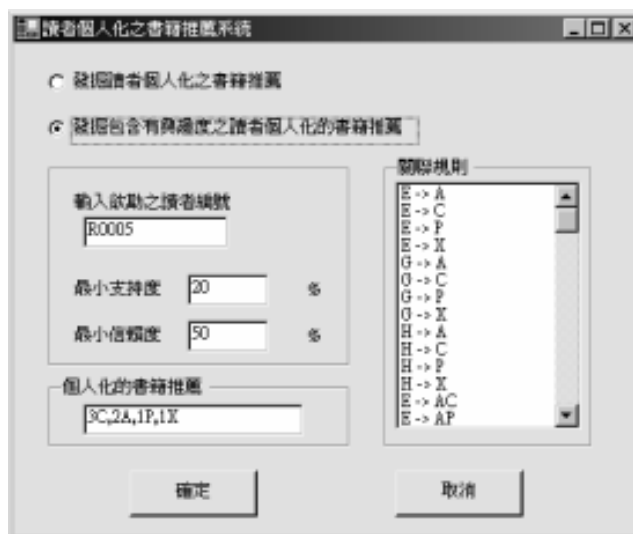
法，則可在「個人化的書籍推薦」欄位中顯示出包含有興趣度值之書籍推薦的結果，欄中位於愈前面的書籍項目，表示借閱的傾向性愈強且興趣度值愈大，如圖四。

接下來，我們實驗評估前面章節所描述之演算法的執行效能，其實驗平台說明如表四。

首先，我們以圖二之借閱資料做為效能評估的資料來源，並以包含有10個書籍項目之某一讀者的借閱資料為探勘的目標，因為在此條件下，其可能產生之項目組的數量最多，所需要的執行時間也可能較大。例如，在此情況下  $itemset_2$  的數量最多可能有  $10 \times 16 = 160$  種組合，但對Apriori演算法而言，其可能產生  $itemset_2$  的數量最多可能為  $26 \times 25 = 650$  種組合。我們設定最小信賴度設為50%，在不同最小支持度的條

件下，分別評估使用Apriori演算法的執行步驟、及利用文中所設計之發掘讀者個人適性化的書籍推薦的探勘演算法，稱之為PBR (Personalized Book Recommendation) 演算法，來找出前置項目組包含於此一讀者之借閱資料的關聯規則的執行時間，其執行結果如圖五。從圖五中顯示出當最小支持度設定愈低時，我們所設計之演算法愈優於Apriori演算法的執行效率。

再者，我們仍以圖二之借閱資料做為效能評估的資料來源，並以包含有10個書籍項目、且興趣度總合最大之某一讀者的借閱資料為探勘的目標，來評估包含有興趣度之讀者個人性化的書籍推薦的執行效能。因為在此條件下，其可能產生之項目組的數量較多，所需要的執行時間也可能較大。我們設定最小信賴度設為50%，在不



圖四 發掘具有興趣度之讀者個人化的書籍推薦的執行畫面

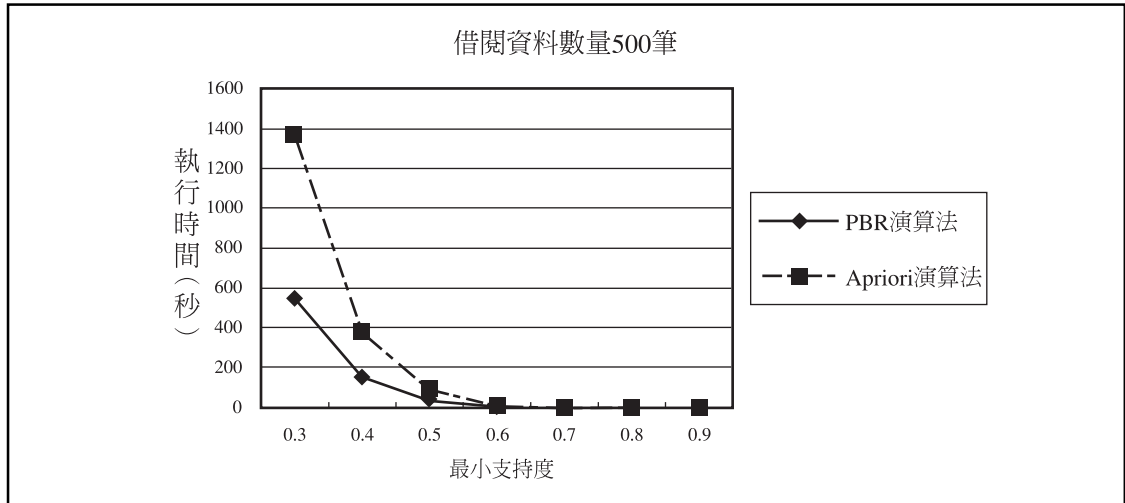
表四 實驗平台

CPU	CPU-Pentium III 850
Main memory	256 Mbytes
作業系統	Windows 2000 Sever
使用語言	C#

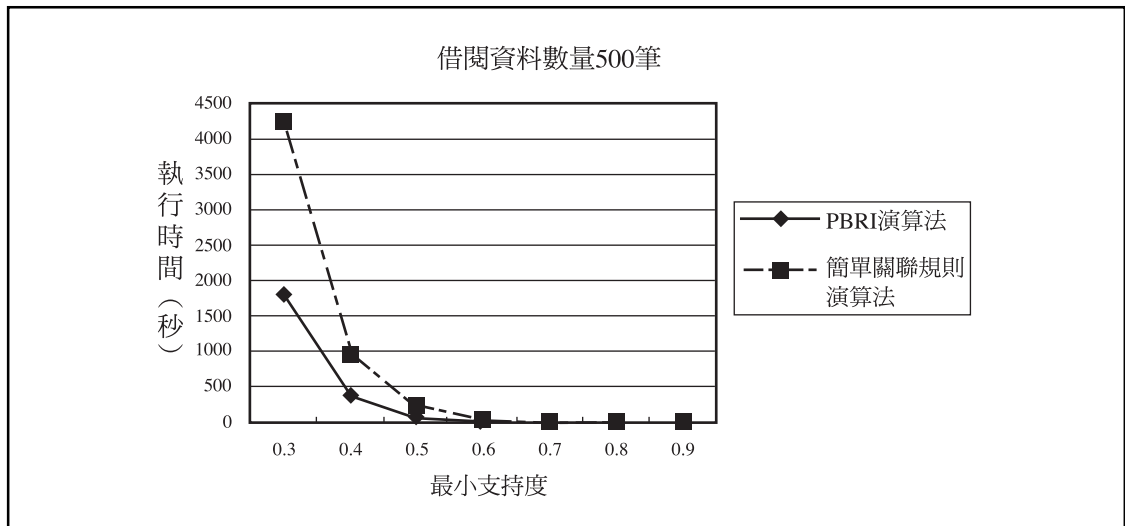
同最小支持度的條件下，分別評估使用擷取簡單關聯規則之演算法、及利用文中所設計之發掘包含有興趣度的讀者個人適性化書籍推薦的探勘演算法，稱之為PBRI (Personalized Book Recommendation for Interest)演算法，來找出前置項目組包含於此一讀者之借閱資料的關聯規則的執行時間，其執行結果如圖六。從圖六中顯示出當最小支持度設定愈低時，我們所設計之演算法愈優於擷取簡單關聯規則之演算法的執行效率。

## 陸、結論

讀者到圖書館依其需求來借閱書籍，從這些被借閱過的書籍中，可反映出讀者對書籍的偏好傾向特徵，若能從借閱資料中找出書籍彼此間的關聯性，對圖書館管理者在擬訂讀者個人化的書籍推薦時，必可提供相當有用的資訊。在本篇論文中，我們以某一讀者為探勘的目標，利用關聯規則分別考量書籍項目是否出現在借閱資料中、及書籍包含有興趣度，來找出此一讀者個人化的書籍推薦。在表五中，我們說明Apriori演算法、擷取簡單關聯規則演算法、及所提出方法之間的差異，從實驗評估的結果顯示，我們所提出之演算法都的執行效率都優於之前演算法的執行效能，來找出所要的關聯規則。從資料的蒐集、分析、方法設計、結果推導出的個人的



圖五 不同最小支持度的執行時間



圖六 包含興趣度之不同最小支持度的執行時間

化書籍推薦、及系統實作執行介面的雛型，顯示出我們所提出之探勘過程具有潛在應用與學術價值。

目前已有許多相關的研究，探討如何利用資料探勘技術以提昇圖書館對讀者個

人化的服務，但鮮少直接針對讀者個人化書籍推薦做分析與計算、及考量讀者對曾經借閱過之書籍的興趣度。因此，本篇論文利用關聯規則並考量讀者對書籍的興趣度，來探勘借閱資料中書籍項目與讀者的

表五 探勘演算法的差異

	Apriori演算法	PBR演算法	簡單關聯規則演算法	PBRI演算法
項目組合方式	任意	必須包含有欲探勘之讀者的借閱資料中任一項目	任意，包含項目組與自己的配對組合	必須包含有欲探勘之讀者的借閱資料中任一項目，包含項目組與自己的配對的組合
探勘步驟	先找出高頻項目組，然後計算符合最小信賴度的關聯規則	同左	同左	同左
關聯規則形式	任意	包含於欲探勘之讀者的借閱資料的項目組→項目組	任意	包含於欲探勘之讀者的借閱資料的項目組→項目組

關聯性，勢必能找出對讀者個人化更貼切與適性的書籍推薦。

## 參考書目

### 一、中文部份

- 卜小蝶(1998)。淺析個人化服務技術的發展趨勢對圖書館的影響。國立成功大學圖書館館刊，2。
- 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。國立交通大學資訊科學研究所碩士論文。
- 洪志淵(2001)。圖書流通記錄之一般化相關規則找尋之研究。國立中山大學資

訊管理研究所碩士論文。

- 張苑菁(2001)。以模糊理論建構之圖書推薦系統。淡江大學資訊工程研究所碩士論文。
- 湯春枝(2002)。從個人化服務行銷的理念談交通大學個人化數位圖書資訊服務(PIE@NCTU)系統。國立成功大學圖書館館刊，9。
- 陳慶瑄(2000)。學習社群對電子圖書館個人化服務之影響。國立中正大學資訊管理研究所碩士論文。
- 陳彥良、凌俊青、許秉瑜(2001)。在包裹式資料庫中挖掘數量關聯規則。資訊



- 管理學報，7 (2)，215-229。
- 辜曼蓉(1999)。讀者資訊尋求行為與以讀者為中心的圖書館行銷。書府，20，81-111。
- 孫冠華(2003)。應用資料探勘技術於數位圖書館之個人化服務及管理。南華大學資訊管理學研究所碩士論文。

## 二、英文部份

- Agrawal, R., Imielinski, T. and Swami, A. (1993), "Mining Association Rules between Sets of Items in Very Large Database," Proceedings of the ACM SIGMOD Conference on Management of Data, 207-216.
- Agrawal, R. and Srikant, R. (1994), "Fast Algorithms for Mining Association Rules in Large Database," Proceedings of the 20<sup>th</sup> International Conference on Very Large Data Bases, 487-499.
- Berry, M. J. A. and Linoff, G. (1997), Data Mining Techniques for Marketing, Sales, and Customer Support, New York: John Wiley.
- Chen, M. S., Han, J. and Yu, P. S. (1996), "Data Mining: An Overview from a Database Perspective," IEEE Trans. on Knowledge and Data Engineering, 8(6), 866-883.
- Han, J. and Kamber, M. (2000), Data Mining: Concepts and Techniques, Morgan Kaufmann. <http://mylibrary.e-lib.nctu.edu.tw/>
- Ou, J., Lin, S. and Li, J., "The Personalized Index Service System in Digital Library," Cooperative Database Systems for Advanced Applications, 92-99, 2001.
- Park, J. S., Chen, M. S. and Yu, P. S. (1997), "Using a Hash-Based Method with Transaction Trimming for Mining Association Rules," IEEE Transactions on Knowledge and Data Engineering, 9(5), 813-825.
- Srikant, R. and Agrawal, R. (1995), "Mining Generalized Association Rules," Proceedings of the 21<sup>th</sup> International Conference on Very Large Data Bases, 407-419.
- Srikant, R. and Agrawal, R. (1996), "Mining Quantitative Association Rules in Large Relational Tables," Proceedings of the ACM SIGMOD Conference on Management of Data, 1-12.

(收件日期：94年9月7日 接受日期：94年12月14日)

