

利用關聯規則發掘讀者適性化之書籍推薦

Using Association Rules to Find Adaptive Book Recommendations for the Reader

陳 垂 呈

Chui-Cheng Chen

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

E-mail : ccchen@mail.stut.edu.tw

【摘要 Abstract】

本論文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讀者曾經借閱的書籍，並以某一讀者為探勘的目標，利用關聯規則(Association rules)分別從以下兩方面發掘此讀者適性化的書籍推薦：一是探勘關聯規則，其前置項目組必須包含於此讀者的借閱資料中，根據關聯規則所顯示出的傾向特徵，做為發掘此讀者適性化之書籍推薦的依據；二是在考量讀者借閱興趣的加權情況下，探勘興趣加權關聯規則(Association rules with weighted interest-items)，其前置項目組必須包含於此讀者的借閱資料中，根據興趣加權關聯規則所顯示出的傾向特徵，做為發掘此讀者適性化之書籍推薦的依據。根據文中所提出的探勘方法，我們設計與建置一個發掘讀者個人適性化的書籍推薦系統。此探勘結果，對圖書館在擬定讀者個人適性化的書籍推薦服務時，將可以提供非常有用的參考資訊。

This paper uses the records of readers' borrowing history as the source data of exploration. The specific borrowing history of an arbitrarily selected reader is aimed at as the subject of exploration, to which a set of association rules is applied to find adaptive book recommendations for the reader from two aspects. From the aspect of applying association rules and that of applying association rules with weighted interest-items, antecedents of these two aspects are respectively contained in the records of the reader's borrowing history. The tendency revealed in the two aspects through the characteristics of the reader's reading history may provide the direction and principle for adaptive book recommendations. According to this methodology of exploration, a system for finding adaptive book recommendations is designed and constructed. The results of this exploration and construction will be particularly useful in the library service of book recommendations.

關鍵詞 Keyword

關聯規則 興趣加權關聯規則 借閱資料 書籍推薦

Association rules ; Association rules with weighted interest-items ; Borrowing history ;

Book recommendations

壹、前言

圖書館服務在資訊技術的支援下，可提供的應用服務愈加多樣化，例如電子圖書、線上資料擷取搜尋、多媒體瀏覽、儲存光碟、電子資料庫等服務，不僅帶動資訊儲存的豐富性，也帶來檢索的便利性及時效性。藉由網際網路(Internet)的連結，配合全球資訊網(WWW)的應用服務，使得圖書館在有限的空間及時間中，可提供讀者全世界豐富的館藏資源、及任何時間、地點的服務。在圖書館管理與經營上，有效善用資訊技術以支援創新管理及服務型態，進而提昇圖書館的有效利用，是圖書館管理者必須思考的問題之一。

在圖書館服務價值的理念中，除了書籍、資料的保存與管理，最重要的就是對讀者的服務。辜曼蓉(1999)曾指出圖書館的核心價值在於讀者和服務人員之間的互動，而館藏內容及資訊傳播扮演提昇圖書館服務品質的輔助角色。在數位圖書館之應用服務中使用者個人化的資訊服務模式，是針對使用者個人的特性而採用不同的查詢介面、查詢結果、或是推薦其適性的書籍資訊。(黃夙賢、柯皓仁、楊維邦，2001)因此，提供讀者一個滿意、便利的使用情境，對提昇讀者對圖書館的忠誠度與利用率將有很大的助益。在讀者服務的型態中，依據讀者個人之需求資訊的傾向特徵，善用個人化服務技術調整圖書館的資訊服務，是圖書館服務重要的行銷方式之一。(卜小蝶，民87；Ou, Lin & Li, 2001)

在圖書館個人化資訊服務中，交通大學的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)，讀者在此系統中可使用的服務項目有「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」、「個人書籤」、「個人借閱狀況」、「個人化桌面」及「使用者設定」等功能。在「個人化推薦」的服務功能中，其利用關聯規則(Association rules)及次

序相關(Sequences)等方法發掘讀者個人化的書籍推薦。因此，圖書館依據讀者個人的特徵傾向，將最貼切的書籍推薦給讀者，是讀者個人化資訊服務重要的功能之一。

圖書館每天均有相當大量的書籍被讀者借閱，在這些被借閱的書籍記錄中，除了顯示讀者借閱的傾向與需求，也隱藏著書籍之間的關聯性。例如若讀者借閱《網際網路行銷》書籍，則也會有借閱《電子郵件行銷DIY》書籍的傾向。因此，若能從這些大量的借閱資料中找出書籍項目之間的關聯性，則對提供讀者個人化的書籍推薦服務，必可提供相當有用的參考資訊。

資料探勘(Data mining)是從大量資料中找出潛在有用的資訊與知識，近年來資料探勘技術已被有效應用在許多的領域中(Han & Kamber, 2006)，例如顧客服務、選擇性行銷(Selective marketing)、銷售(Sales)、信用核證(Credit approval)、及醫療診斷(Medical diagnosis)等。在本論文中，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讀者曾經借閱的書籍項目，並以某一讀者為探勘目標，假設其借閱資料為 X ， X 為包含一個或以上之書籍項目所形成的項目組，利用關聯規則分別從以下兩方面發掘此讀者適性化的書籍推薦：

1. 只考量書籍是否出現在借閱資料中

文中設計一個探勘關聯規則的方法，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(Frequent itemsets)。從以上關聯規則所顯示出的借閱傾向為：若讀者曾經借閱書籍 Y ，則也會有借閱書籍 Z 的傾向特徵，且若當書籍 Y 愈相近於書籍 X ，則關聯規則愈能反映此讀者的借閱傾向特徵，借閱書籍 Z 的傾向性也愈強，藉此做為發掘此讀者適性化之書籍推薦的依據。

2. 加上考量讀者個人借閱興趣的因素

從此讀者之借閱資料中所反映出的借閱興趣傾向，利用興趣加權的計算方式，文中設計一個探勘興趣加權關聯規則(Association rules with weighted interest-items)的方法，找出以下形式的興趣加權關聯規則： $U \rightarrow V, U \subseteq X, U \cap V = \emptyset, U, V$ 為包含一個或以上之書籍項目所形成的項目組， $U \cup V$ 是高頻項目組。從以上興趣加權關聯規則所顯示出的借閱傾向為：若讀者曾經借閱書籍 U ，則也會有借閱書籍 V 的傾向特徵，且若當書籍 U 愈相近於書籍 V 、且包含的加權值愈大，則興趣加權關聯規則愈能反映此讀者借閱興趣的傾向特徵，借閱書籍 V 的傾向性也愈強，藉此做為發掘此讀者適性化之書籍推薦的依據。

我們以南部某一科技大學圖書館的借閱資料為例，根據所提出之方法設計與建置一個讀者適性化書籍借閱的推薦系統，並評估探勘結果的推薦效益。此探勘結果，對圖書館發掘讀者個人適性化的書籍推薦，將可提供非常有用的參考資訊。

本篇論文的架構如下：首先介紹資料探勘技術、及其在圖書館服務應用的相關研究；其次以某一讀者做為探勘的目標，設計一個方法探勘關聯規則，其前置項目組必須包含於此讀者的借閱資料中，藉此做為發掘此讀者適性化之書籍推薦的依據；接著以某一讀者做為探勘的目標，在考量讀者借閱興趣的加權情況下，設計一個方法探勘興趣加權關聯規則，其前置項目組必須包含於此讀者的借閱資料中，藉此做為發掘此讀者適性化之書籍推薦的依據；進而根據所提出的方法，設計與建置一個讀者個人適性化的書籍推薦系統；最後做一結論。

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，可完成的工作有：關聯規則、分類分析(Classification analysis)、分群(Clustering)、次序相關分析(Sequential pattern analysis)、及預測(Forecasting)等(Chen, Han, & Yu, 1996; Han & Kamber, 2006)，其探勘結果對企業從事行銷決策及市場預測等活動，可以提供非常有價值的參考資訊。(Berry & Linoff, 2004; Hui & Jha, 2000)

在圖書館書籍借閱的服務中，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，而圖書館只能被動等待讀者借閱書籍。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。近年來，已有許多利用資料探勘技術於圖書館服務應用的相關研究，其中包含有：陳慶瑄(2000)利用 k -means 方法找出學習社群，以支援電子圖書館之個人化服務；孫冠華(2003)利用關聯規則做為提昇數位圖書館之個人化服務及管理的方法依據；吳安琪(2001)利用資料探勘技術發掘讀者的社群關係，做為吸引讀者借閱書籍的依據，對提昇圖書館的借閱率與讀者忠誠度可以提供很大的助益；洪志淵(2001)利用資料探勘技術找出讀者與書籍之間的一般化關聯規則，藉此做為對讀者之新書推薦的依據；鄭玉玲(2003)曾利用關聯規則，在數位圖書館上建構一個讀者個人化的檢索與推薦系統；張苑菁(2001)利用模糊理論(Fuzzy)與資料探勘技術分析讀者的借閱資料，以推薦讀者相關的書籍作參考。

從交易資料中探勘關聯規則首先由Agrawal, Imielinski and Swami (1993)提出，做為顯示項目之間關聯性的型樣，關聯規則的定義說明如下：假設 I 是全部項目的集合， 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 $X \cap Y = \emptyset, X$ 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：在全部交易資料中同時包含有 $X \cup Y$ 的比率值，即 $s = (\text{包含有 } X \cup Y \text{ 之交易資料的數量}) / (\text{全部交易資料數量})$ ；信賴度 c 的定義為：在包含有 X 的交易資料中，也同時包含有 Y 的比率值，即 $c = (\text{包含有 } X \cup Y \text{ 之交易資料的數量}) / (\text{包含有 } X \text{ 之交易資料的數量}) = (X \cup Y \text{ 的支持度}) / X \text{ 的支持度}$ 。找出的關聯規則，其支持度與信賴度必須大於或等於所設定的最小支持度與最小信賴度，這樣的關聯規則才成立。

關聯規則的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，若關聯規則 $A \rightarrow BC$ 滿足最小信賴度，則此關聯規則成立。可參考探勘關聯規則的相關研究(Agrawal & Srikant, 1994; Coenen, Leng, & Ahmed, 2004; Han, Pei, & Yin, 2000; Han, Pei, Yin, & Mao, 2004; Li, He, & Lei, 2005; Liu, Li, & Zhao, 2004; Park, Chen, & Yu, 1997; Srikant & Agrawal, 1995; Tsay & Chang-Chien, 2004; Tsay & Chiang, 2005)，以及利用關聯規則於個人化服務應用的相關研究(Huang, Duyang, & Ke, 2001)。在探勘關聯規則的方法中，Apriori演算法(Agrawal & Srikant, 1994)是最具代表性的方法之一，其特性為簡單、容易瞭解及實作。Apriori演算法探勘關聯規則的步驟說明如下：

1. 找出 $frequent_{k-1}$ ， $k > 1$ ，若為 \emptyset ，則停止執行。
2. 由步驟1中組合任兩個有 $k-2$ 項目相同的 $frequent_{k-1}$ ，形成 $itemset_k$ 。

3. 判斷由步驟2所找出的 $itemset_k$ ，其所有包含的 $itemset_{k-1}$ 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，若符合就成為 $frequent_k$ ，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。

Tsay and Chang-Chien (2004) 提出 CDAR (Cluster-decomposition association rule)演算法探勘關聯規則，在探勘的過程中，先將包含有相同項目個數的交易資料歸屬於同一群組中，並利用類似Apriori演算法探勘關聯規則。在判斷 k -項目組是否為高頻 k -項目組的過程中， $k \geq 2$ ，此演算法只須掃描包含項目個數 $\geq k$ 的群組即可，而非如Apriori演算法需要掃描全部交易資料，實驗並顯示其執行效率優於Apriori演算法。

在過去的研究中，產品加權是考量一些獲利率高，但交易次數不頻繁的產品，可能因出現次數未能滿足設定值，而未能被挖掘出來，因此藉由加權來改變其在總交易資料數量中出現次數的計算值。經過加權之後所找出的關聯規則，稱之為加權關聯規則(Association rules with weighted items)，可參考相關研究(顏秀珍、邱鼎穎、李御璽, 2001; Cai, Fu, Cheng, & Kwong, 1998; Yue, Tsang, Yeung, & Shi, 2000)。探勘加權關聯規則的步驟說明如下：(Cai, et al., 1998)

1. 找出 $itemset_1$ 。
2. 找出各項目對應的加權值。
3. 找出 $itemset_{k-1}$ ， $k > 1$ 。
4. 從步驟3中組合任兩個有 $k-2$ 項目相同的 $itemset_{k-1}$ 找出 $itemset_k$ ，跳至步驟3直到找出所有項目組。

5. 利用加權支持度公式，計算出各項目組加權後的支持度。
6. 檢查項目組的支持度，若滿足最小支持度者，即成爲高頻項目組。
7. 計算高頻項目組所形成之加權關聯規則，若滿足最小信賴度，則加權關聯規則成立。

在以上演算法的計算過程中，在找出項目組之後並不立即判斷其是否爲高頻項目組，而是找出全部項目組並計算加權值之後，再判斷其是否爲高頻項目組。因爲若某一 k - 項目組若未能滿足最小支持度而被刪除，可能導致無法找出其組成某一 $(k+1)$ - 項目組，其加權之後卻能滿足最小支持度。因此先找出所有項目組，並計算其支持度的加權值，再判斷是否刪除，以將獲利率高或是有興趣的產品提高其支持度。

本論文以讀者之借閱資料做爲探勘的資料來源，並以某一讀者做爲探勘的目標，在探勘書籍項目之間的關聯性時，若能藉由此讀者曾經借閱的書籍來反映出其借閱興趣，對此讀者有借閱興趣之書籍進行加權的計算，其探勘結果必定可以較容易找出、且提昇適性化的書籍推薦服務。文中利用關聯規則分別從以下兩方面發掘此讀者適性化的書籍推薦：一是只考量書籍是否出現在讀者的借閱資料中；二是加上考量此讀者借閱興趣的因素。

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爲探勘的資料來源，每一筆借閱資料包含讀者曾經借閱過的書籍項目，且已依據字元由小至大做排序，例如 $A < B < C$ ，並以某一讀者爲探勘的目標，找出前置項目組必須包含於此讀者之借閱資料的關聯規則，藉此做爲發掘此讀者適性化之書籍推薦的依據。本節共分爲兩部分，首先說明發掘某一讀者適性化之書籍推薦的探勘過程，再以一實例做說明。

一、探勘方法設計

假設 X 爲目前欲探勘之某一讀者的借閱資料， X 爲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：

$$Y \rightarrow Z, Y \subseteq X, Y \cap Z = \emptyset, Y \cup Z \text{ 是高頻項目組}$$

其中 Y 、 Z 爲包含一個或以上之書籍項目所形成的項目組。從以上形式之關聯規則所顯示出的借閱傾向爲：若讀者曾經借閱過書籍 Y ，則也會有借閱書籍 Z 的傾向。由於 $Y \subseteq X$ ，表示此關聯規則可表現出此讀者的借閱傾向特徵，且當 Y 愈相近於 X 時，此關聯規則愈能反映出此讀者的借閱傾向特徵，其借閱書籍 Z 的傾向性也愈強。因此，藉由探勘以上形式的關聯規則，可做爲發掘此讀者適性化之書籍推薦的依據。

爲了避免計算與 X 無關的項目組及配合探勘的需要，我們修改 Apriori 演算法及 CDAR 演算法的執行步驟，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。探勘以上形式之關聯規則的過程說明如下：

1. 從借閱資料庫中找出 $frequent_1$ ，其中必須至少一個 $frequent_1 \subseteq X$ 、及至少一個 $frequent_1 \not\subseteq X$ ，若借閱資料 $\cap X = \emptyset$ ，則刪除此借閱資料，否則保留此借閱資料。同時將包含項目個數爲 i 、且字首(第一個項目)爲 j 的借閱資料歸屬於群組 G_{ij} 中， $i > 1$ ， $j \in C$ ， i 爲該群組中借閱資料包含之書籍項目的個數， C 爲全部書籍項目的集合。
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 $itemset_2$ ，從群組 G_{ij} 的借閱資料中檢查 $itemset_2$ 是否滿足最小支持度， $i > 1$ ， $j \leq itemset_2$ 的第一個項目，若符合就成爲 $frequent_2$ ，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 $itemset_k$ ，其包括的所有子集合 $itemset_{k-1}$ 是否都出現在步驟3中(若 $itemset_{k-1} \cap X = \emptyset$ 、或 $itemset_{k-1} \cap \text{非}X = \emptyset$ ，則將不予列入計算)，若成立就保留此 $itemset_k$ ，否則就刪除。
- 6.從群組 G_{ij} 的借閱資料中， $i \geq k$ ， $j \leq itemset_k$ 的第一個項目，檢查由步驟5所找出的 $itemset_k$ 是否滿足最小支持度，若符合就成為 $frequent_k$ ，否則就刪除。
- 7.計算 $frequent_k$ 所形成之關聯規則的形式為： $Y \rightarrow Z, Y \subseteq X, Y \cap Z = \emptyset, \{Y \cup Z\} \in frequent_k$ 。
- 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 $frequent_k \cap X \neq \emptyset$ ，如此將可避免探勘與目標之項目組型樣無關的計算。從步驟1整理後之群組 G_{ij} 的借閱資料中， $i \geq k$ ， $j \leq itemset_k$ 的第一個項目，檢查 $itemset_k$ 是否為高頻項目組，如此將可避免掃描未包含 X 任一項目的借閱資料、及項目個數小於 k 的借閱資料，並且只掃描字首小於等於 $itemset_k$ 之第一個項目的借閱資料。相較於原始

Apriori演算法、及CDAR演算法，我們所設計之探勘方法將可更有效率找到所要的項目組。然後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。藉由關聯規則 $Y \rightarrow Z$ 所顯示出的借閱傾向，可發掘出以下的傾向特徵：

關聯規則 $Y \rightarrow Z$ 成立，且 $Y \subseteq X, Y \cap Z = \emptyset$ ，則 Z 為此讀者適性化的書籍推薦，且 Y 愈相近於 X ，則借閱書籍 Z 的傾向性也愈強。

根據以上關聯規則所顯示的傾向特徵，圖書館可以主動推薦書籍 Z 給此讀者，如此將可減少此讀者搜尋書籍的時間，並提昇書籍借閱的頻率。

二、實例說明

我們以表1之借閱資料庫 D_1 為例，說明發掘讀者個人適性化之書籍推薦的探勘過程。 $\{A, B, C, D, E\}$ 為全部書籍項目的集合， $\{R_1, R_2, R_3, R_4, R_5, R_6\}$ 為6筆讀者之借閱資料的集合，最小支持度為50%(即最小支持數量為3)，最小信賴度為70%。假設目前欲探勘之讀者的借閱資料為BE，以下說明發掘此讀者個人適性化之書籍推薦的探勘過程。

表 1
借閱資料庫 D_1

借閱資料編號	書籍項目
R_1	ACD
R_2	BCE
R_3	ABCE
R_4	BE
R_5	BCE
R_6	ABD

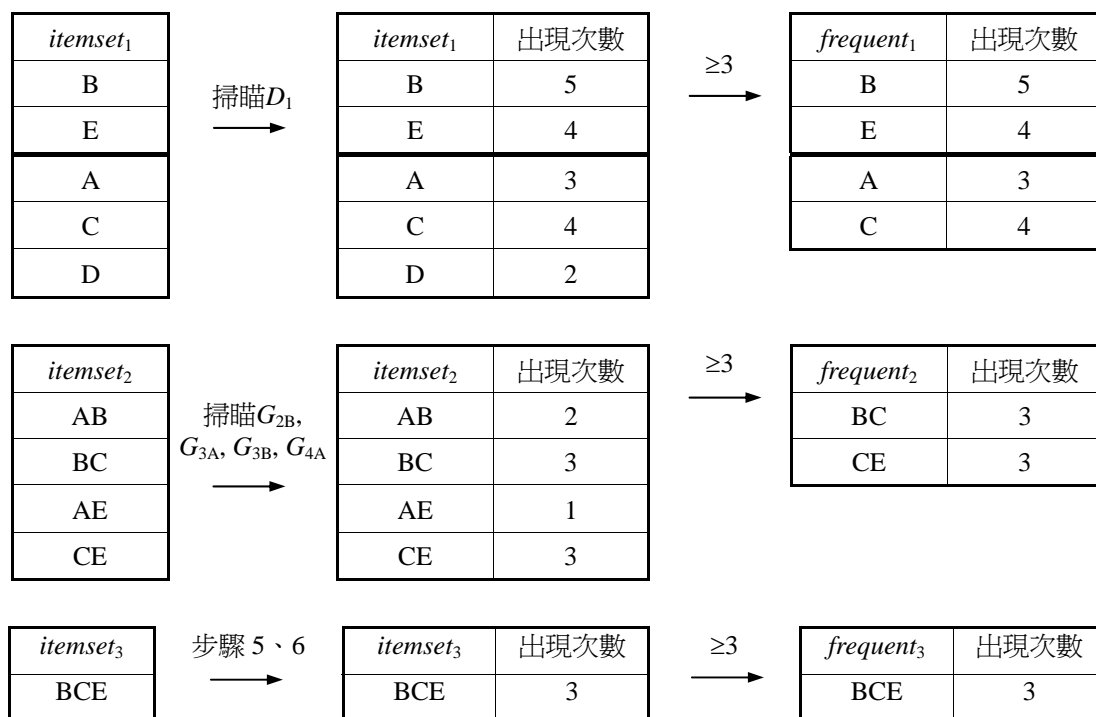
在擷取 $frequent_1$ 的過程中，將不包含BE任一項目之借閱資料刪除，並將包含相同項目個數的借閱

資料歸屬於同一群組中，並再將包含相同字首的借閱資料歸屬於同一子群組中，而形成表 2 之借閱資料庫 D_2 。

表 2
借閱資料庫 D_2

群組	借閱資料編號	書籍項目
G_{2B}	R_4	BE
G_{3A}	R_6	ABD
G_{3B}	R_2	BCE
	R_5	BCE
G_{4A}	R_3	ABCE

擷取包含 BE 中任何項目之高頻項目組的過程如下：



無 4-項目組。

我們以高頻 3-項目組 BCE 為例，計算形成的關聯規則 $BE \rightarrow C$ ，其信賴度為 $3/4=75\%$ ，滿足最小信賴度，此關聯規則成立。因此，藉由關聯規則 $BE \rightarrow C$ 所顯示出的借閱傾向特徵，可發掘出 C 為

此讀者適性化的書籍推薦。依據相同的方法，可找出其他高頻項目組所發掘的書籍推薦。

肆、考量借閱興趣發掘讀者個人適性化之書籍推薦

加權關聯規則是針對在探勘關聯規則的過程中，將項目組的支持度進行調整，使其有興趣的產品較容易符合最小支持度，以突顯這些項目的重要性。支持度是判定產品項目佔市場的規模度，若某一產品項目的支持度過低，則將導致無法找出其相關的關聯規則，如此便不易瞭解此產品與其他產品之間的關聯性。在圖書館讀者個人化之書籍推薦的應用中，基於欲探勘之讀者的借閱興趣，提升此讀者有借閱興趣之書籍的支持度權值，使其較容易成為高頻項目組，進而找出其相關的興趣加權關聯規則，其發掘出之書籍推薦將能更貼切於此讀者的借閱傾向。

此節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讀者曾經借閱過的書籍項目，且已依據字元由小至大做排序，例如 $A < B < C$ ，並以某一讀者為探勘的目標，在考量讀者借閱興趣的情況下，探勘前置項目組必須包含於此讀者之借閱資料的興趣加權關聯規則，藉此做為發掘此讀者適性化之書籍推薦的依據。此節共分為兩小節，首先說明在考量讀者借閱興趣的情況下，發掘某一讀者適性化之書籍推薦的探勘過程，再以一實例做說明。

一、探勘方法設計

假設目前欲探勘之讀者的借閱資料為 X ， X 為包含一個或以上之書籍項目所形成的項目組，從 X 包含的書籍項目中可反映出此讀者的借閱興趣或是需求，若為同一類別的書籍，則具有相同或是相似的分類特徵。例如《網際網路行銷》及《電子郵件行銷DIY》兩本書籍同歸屬於「網路行銷」之主題類別或是分類號為“496”。因此在發掘書籍推薦

的過程中，若能加強此讀者曾經借閱過之書籍的類別特徵，其書籍推薦將更能貼切於讀者的借閱興趣或是需求。

假設借閱資料 X 中的書籍項目共歸屬於 $c_1, c_2, c_3, \dots, c_m$ 等 m 個類別，其中各類別包含之書籍數量佔 X 包含之書籍數量的比例值分別為 $p_1, p_2, p_3, \dots, p_m$ ，且 $p_1 + p_2 + p_3 + \dots + p_m = 1$ 。在計算項目組的加權支持度時，若其中有項目 $\subseteq X$ 或項目 $\not\subseteq \{c_1, c_2, c_3, \dots, c_m\}$ ，則忽略這些項目的加權值，因為這些項目為此讀者已借閱過的書籍、或是不需要進行加權的書籍項目。對於需要加權的項目，可根據項目歸屬之類別所設定的加權值進行加權，其加權值的設定方法有以下兩種。

一是依據各類別的比例值 $p_1, p_2, p_3, \dots, p_m$ ，直接分別給定其加權值為 $w_1, w_2, w_3, \dots, w_m$ 。例如，假設某一借閱資料 X_1 包含有5個書籍項目ABCDE，其中A歸屬於類別 c_1 、BCDE歸屬於類別 c_2 ，計算各類別 c_1 、 c_2 之書籍項目所佔的比例值分別為20%及80%，其可反映出此讀者對類別 c_1 及 c_2 之書籍佔其借閱興趣或是需求分別為20%、及80%。因此，對非ABCDE且歸屬於類別 c_1 及類別 c_2 的書籍項目，管理者(或是使用者)可依其經驗或是主觀上的認定，而直接設定其加權值，例如分別為2及5。

二是管理者(或是使用者)可先設定 t 個分別為 $w_1, w_2, w_3, \dots, w_t$ 的加權值，其中 $w_1 \leq w_2 \leq w_3 \leq \dots \leq w_t$ ， $t \geq 1$ ，我們將加權值劃分成 t 個級數，採用0至1平均分佈的方式，將其分割成 t 個區間，並藉此對應各區間的加權值，使得位於分佈較後面區間具有較大的加權值，即

- 區間1：0 < 區間1 ≤ 1/t，其加權值為 w_1
- 區間2：1/t < 區間2 ≤ 2/t，其加權值為 w_2
- 區間3：2/t < 區間3 ≤ 3/t，其加權值為 w_3
- ⋮
- ⋮

區間 t ： $(t-1)/t < \text{區間} \leq 1$ ，其加權值為 w_t 。

若讀者的借閱資料為 X ，其包含書籍項目共歸屬於 m 個類別，各類別之書籍項目的比例值分別為 $p_1, p_2, p_3, \dots, p_m$ ，則依據各比例值位於那一區間，分別設定其類別之書籍項目所對應的加權值。例如假設加權值分別為1、2、3，其區間分佈可分割成： $0 < \text{區間} 1 \leq 33\%$ 、 $33\% < \text{區間} 2 \leq 67\%$ 、及 $67\% < \text{區間} 3 \leq 100\%$ ，對應的加權值分別為1、2、及3。若某一借閱資料 X_1 包含有5個書籍項目ABCDE，其中書籍A歸屬於類別 c_1 、書籍BCDE歸屬於類別 c_2 ，計算各類別 c_1 、 c_2 之書籍項目所佔的比例值分別為20%及80%，其可反映出此讀者對類別 c_1 及 c_2 之書籍佔其借閱興趣或是需求分別為20%、及80%，並分別位於區間1及區間3。因此，對非ABCDE且歸屬於類別 c_1 及類別 c_2 的書籍項目，分別設定其加權值為1及3。

本論文將採用第二種加權設定的計算方式，其優點有：一是先設定加權值，不必事先知道讀者的借閱資料中各類別之書籍項目的比例值，可適用於普遍性；二是依據各類別之書籍項目的比例值所位於的區間，其對應設定的加權值，仍可客觀顯示讀者之書籍借閱興趣的不同比例值，及其反映出之加權的差異性及重要性。在考量讀者借閱興趣的情況下，我們必須找出以下形式的興趣加權關聯規則：

$Y \rightarrow Z, Y \subseteq X, Y \cap Z = \emptyset, Y, Z$ 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。

從以上興趣加權關聯規則所顯示出的借閱傾向為：若讀者曾經借閱過書籍 Y ，則也會有借閱書籍 Z 的傾向特徵。由於 $Y \subseteq X$ ，表示此興趣加權關聯規則可表現出此讀者的借閱傾向特徵，且當 Y 愈相近於 X 、及 Y 包含書籍項目的加權值愈大，此興趣加權關聯規則愈能反映出此讀者的借閱傾向特徵，其借閱書籍 Z 的傾向性也愈強。因此，藉由探勘以上形式的興趣加權關聯規則，可做為發掘此讀

者適性化之書籍推薦的依據。

在探勘興趣加權關聯規則的過程中，我們對有加權值之項目進行以下的處理：若一借閱資料中有一書籍項目的加權值為 w ， w 為大於等於1的正整數，則在此借閱資料中複製此書籍項目成為 w 個，藉此以增加其所形成之項目組的支持度。例如一借閱資料為ABCDE，若A、C的加權值分別為3及2，則此借閱資料須變更為AAABCCDE。我們修改前一節所描述之演算法，直接組合 X 中的項目與非 X 中的項目而形成項目組，並加入項目的加權值，以判斷這些項目組是否為高頻項目組。探勘以上形式之興趣加權關聯規則的過程說明如下：

1. 首先設定各加權值，依據欲探勘之讀者的借閱資料所顯示的借閱興趣，設定需要加權之項目的加權值。
2. 從借閱資料庫中找出 $itemset_1$ 的出現次數，並依據項目的加權值調整 $itemset_1$ 的加權支持度，若滿足最小支持度，則成為 $frequent_1$ ，其中必須至少一個 $frequent_1 \subseteq X$ 、及至少一個 $frequent_1 \not\subseteq X$ 。若借閱資料 $\cap X = \emptyset$ ，則刪除此借閱資料，否則保留此借閱資料。同時將包含項目個數為 i 且字首(第一個項目)為 j 的借閱資料歸屬於群組 G_{ij} 中， $i > 1, j \in C, i$ 為該群組中借閱資料包含之書籍項目的個數， C 為全部書籍項目的集合。在各群組 G_{ij} 中，若借閱資料包含加權的書籍項目，則依據其加權值複製項目出現在借閱資料中的個數。
3. 由步驟2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 $itemset_2$ ，從群組 G_{ij} 的借閱資料中檢查 $itemset_2$ 是否滿足最小支持度， $i > 1, j \leq itemset_2$ 的第一個項目，若符合就成為 $frequent_2$ ，否則就刪除。
4. 找出所有的 $frequent_{k-1}, k > 2$ 。
5. 由步驟4中，組合任兩個有 $k-2$ 項目相同的 $frequent_{k-1}$ ，形成 $itemset_k$ 。

6. 判斷由步驟5所找出的 $itemset_k$ ，其包括的所有子集合 $itemset_{k-1}$ 是否都出現在步驟4中(若 $itemset_{k-1} \cap X = \emptyset$ 、或 $itemset_{k-1} \cap \bar{X} = \emptyset$ ，則將不予列入計算)，若成立就保留此 $itemset_k$ ，否則就刪除。
7. 從群組 G_{ij} 的借閱資料中， $i \geq k, j \leq itemset_k$ 的第一個項目，檢查由步驟6所找出的 $itemset_k$ 是否滿足最小支持度，若符合就成為 $frequent_k$ ，否則就刪除。其中若有多個 $itemset_k$ 包含相同的項目，則合併計算其支持度。
8. 計算 $frequent_k$ 所形成之關聯規則的形式為：
 $Y \rightarrow Z, Y \subseteq X, Y \cap Z = \emptyset, \{Y \cup Z\} \in frequent_k$ 。
9. 跳至步驟4繼續找出 $frequent_{k+1}$ ，直到無法產生高頻項目組為止。

在以上演算法的步驟2，我們依據書籍項目的加權值複製項目於借閱資料中的個數，以反映出項目權重、及計算項目組的加權支持度。從步驟3開始，我們所擷取出的 $frequent_k$ ，必定為 $frequent_k \cap X \neq \emptyset$ ，如此將可避免探勘與目標之項目組型態無關的計算。從步驟2整理後之群組 G_{ij} 的借閱資料中，檢查 $itemset_k$ 是否為高頻項目組， $i \geq k, j \leq itemset_k$ 的第一個項目，如此將可避免掃描未包含 X 任一項目的借閱資料、及項目個數小於 k 的借閱資料，並且只掃描字首小於等於 $itemset_k$ 之第一個項目的借閱資料。然後計算高頻項目組所形成的興趣加權關聯規則 $Y \rightarrow Z$ ，且 $Y \subseteq X$ ，若滿足最小信賴度，則興趣加權關聯規則成立。藉由興趣加權關聯規則 $Y \rightarrow Z$ 所顯示出的借閱傾向，可發掘出以下的特徵：

興趣加權關聯規則 $Y \rightarrow Z$ 成立，且 $Y \subseteq X, Y \cap Z = \emptyset$ ，則 Z 為此讀者適性化的書籍推薦，且 Y 愈相近於 X 、及 Y 包含書籍項目的加權值愈大，則借閱的傾向性也愈強。

根據以上興趣加權關聯規則所顯示的傾向特

徵，從讀者曾經借閱過的書籍項目中，圖書館依據讀者個人的借閱興趣，利用以上所描述的加權探勘方法，可以主動推薦書籍 Z 給此讀者，其將更貼切於此讀者的借閱興趣與需求。

二、實例說明

我們仍以表1之借閱資料庫 D_1 為例，說明在考量借閱興趣的加權情況下，發掘讀者個人適性化之書籍推薦的探勘過程。 $\{A, B, C, D, E\}$ 為書籍項目的集合，其中 $\{A, B\}$ 屬於類別 c_1 、 $\{D\}$ 屬於類別 c_2 、及 $\{C, E\}$ 屬於類別 c_3 ，加權值分別為 1, 2, 3， $\{R_1, R_2, R_3, R_4, R_5, R_6\}$ 為 6 筆讀者之借閱資料的集合，最小支持度為 50%，即最小支持數量為 3，最小信賴度為 70%。假設目前欲探勘之讀者的借閱資料為 BE ，以下說明在考量讀者借閱興趣的情況下，發掘此讀者適性化之書籍推薦的探勘過程。

首先計算加權值各區間的平均分佈、及其對應的加權值為：

1. 區間1：0 < 區間1 ≤ 33%，其加權值為 1；
2. 區間2：33% < 區間2 ≤ 67%，其加權值為 2；
3. 區間3：67% < 區間3 ≤ 100%，其加權值為 3。

計算此讀者之借閱興趣的加權值為：

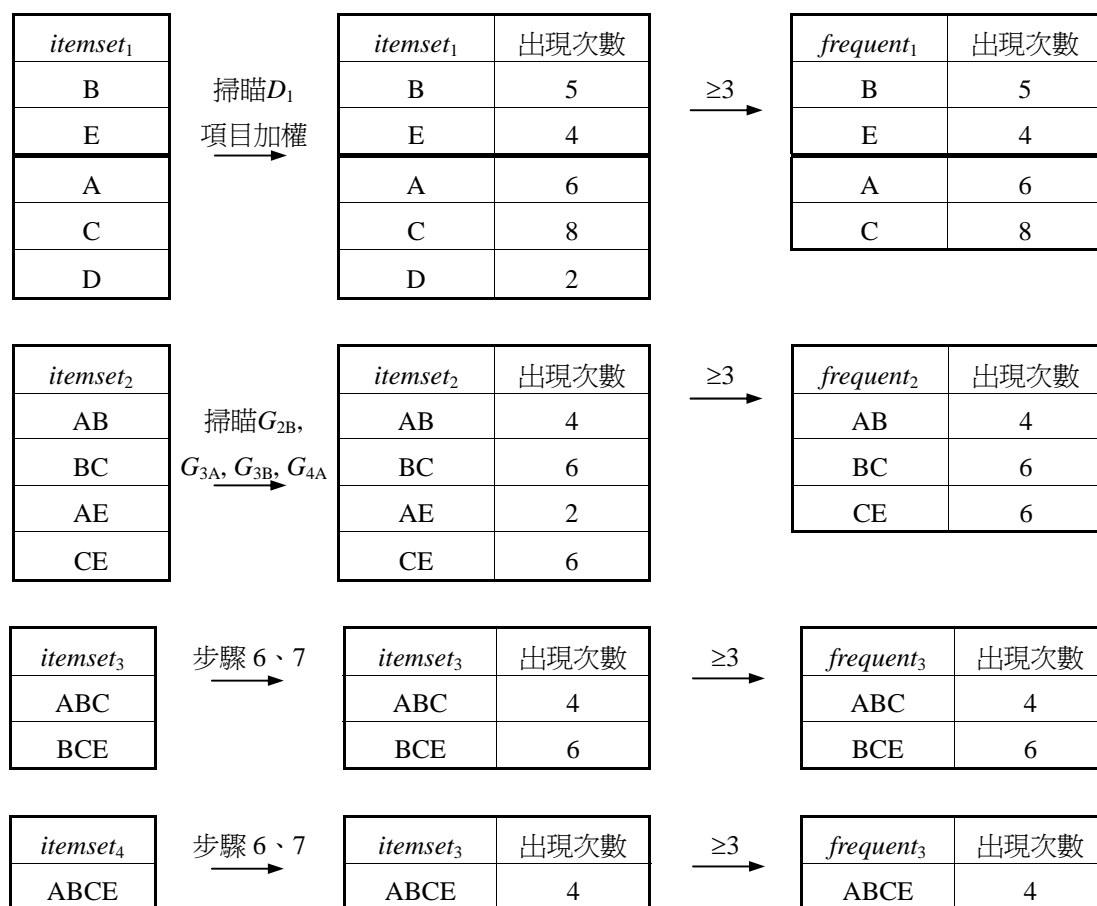
1. 類別 c_1 之書籍的加權值 = 2，由於書籍 B 佔此讀者之借閱資料的比例值為 $1/2 = 50\%$ ，位於區間 2；
2. 類別 c_2 之書籍不計算其加權值，由於此讀者未曾借閱過類別 c_2 的書籍；
3. 類別 c_3 之書籍的加權值 = 2，由於書籍 E 佔此讀者之借閱資料的比例值為 $1/2 = 50\%$ ，位於區間 2。

然後將未包含 BE 中任一項目的借閱資料刪除，並根據項目的加權值，在借閱資料中複製需要加權的項目，而形成表 3 之借閱資料庫 D_3 。

表 3
借閱資料庫 D_2

群組	借閱資料編號	書籍項目
G_{2B}	R_4	BE
G_{3A}	R_6	AABD
G_{3B}	R_2	BCCE
	R_5	BCCE
G_{4A}	R_3	AABCCE

擷取包含 BE 中任何項目之項目組的過程如下：



無 5-項目組。

我們以高頻 4-項目組 ABCE 為例，計算形成的興趣加權關聯規則 $BE \rightarrow AC$ ，其信賴度 $=4/4 = 100\%$ ，滿足最小信賴度，此興趣加權關聯規則成立。因此，藉由興趣加權關聯規則 $BE \rightarrow AC$ 所顯示出的借閱傾向特徵，可發掘出 AC 為此讀者適性化的書籍推薦。依據相同的計算方法，可找出其他高頻項目組所形成的興趣加權關聯規則，及其所發掘出的書籍推薦。

在以上探勘的過程中，若未對項目組進行加權的計算，則最多只能擷取到高頻 3-項目組 BCE，其形成的關聯規則 $BE \rightarrow C$ ，可發掘 C 為此讀者適

性化的書籍推薦，卻無法找出書籍 A。因此針對讀者個人的借閱興趣進行的加權計算，確實可找出原本未能顯示出的傾向特徵，其探勘結果將更能貼切符合讀者個人的傾向需求。

伍、發掘讀者適性化之書籍推薦系統的設計與實作

我們將前述的探勘方法，設計與建置一個發掘讀者適性化之書籍推薦的探勘系統，系統探勘過程的模型如圖 1，表 4 為探勘系統的開發平台。

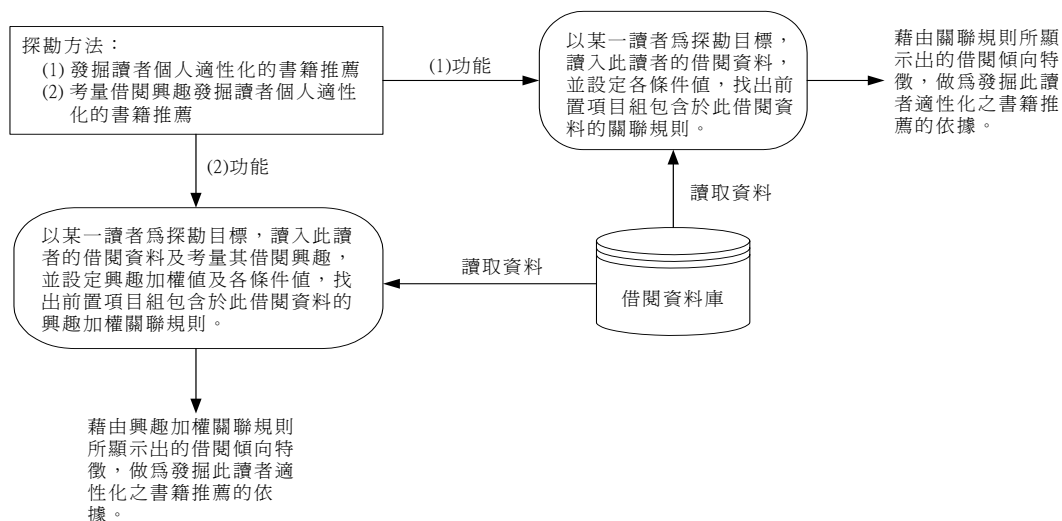


圖 1 系統探勘過程模型

表 4
系統開發平台

作業系統	Windows XP Professional Edit
CPU	Intel Pentium-4 1.7GHz
主記憶體	512M SDRAM
程式語言	VB.NET
網頁設計	ASP.NET
資料庫	Access 2003

本文以南部某一科技大學圖書館之讀者的借閱資料，如表 5，做為所設計與建置之探勘系統的資料來源。在 6 年的借閱資料中，我們以前面 5 年(2000-2004)之借閱資料做為探勘計算的訓練資料，合併其中相同讀者的借閱資料之後，顯示有 16,033 位不同讀者的借閱資料。以最後一年(2005)之借閱資料做為探勘計算的驗證資料，其中在前面 5 年也有出現的讀者共有 6,532 位。

料，合併其中相同讀者的借閱資料之後，顯示有 16,033 位不同讀者的借閱資料。以最後一年(2005)之借閱資料做為探勘計算的驗證資料，其中在前面 5 年也有出現的讀者共有 6,532 位。

表 5
借閱資料

	年份	讀者人數	書籍數量
訓練資料	2000	967	6,494
	2001	2,172	15,196
	2002	4,424	23,638
	2003	7,050	34,232
	2004	9,350	40,255
	2000-2004(合併相同讀者的借閱資料)	16,033	73,305
驗證資料	2005	8,666	37,850
	在訓練資料和驗證資料都出現	6,532	26,368

圖 2 為借閱資料的原始資料，包含有書籍的「條碼號」、「讀者證號」、「借閱日期」、「借閱時間」、「歸還日期」、「歸還時間」、及「書名」等欄位資料，每一筆資料是以每一本書籍為一個記錄來

儲存。因此，在進行探勘計算之前，我們必須先將曾經被同一讀者借閱過的書籍，彙整成一筆屬於此讀者的借閱資料。

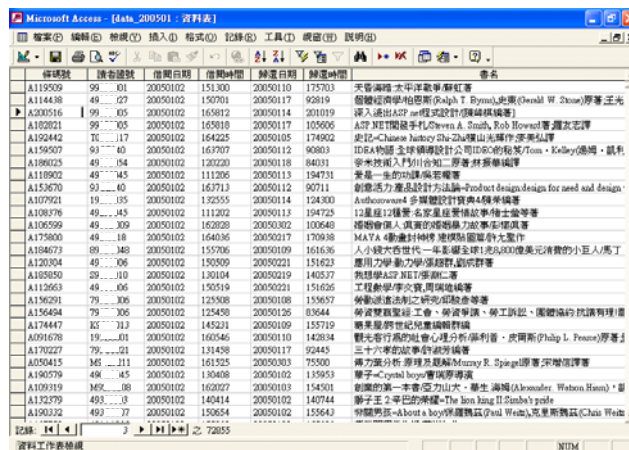


圖 2 原始借閱記錄

圖 3 為書籍分類資料，包含有書籍的「條碼號」、「分類號」、及「書名」等欄位資料，其中分

類號共分為 0, 100, 200, ..., 900 等 9 大分類，每一筆資料包含每一本書籍的條碼號、分類號、及書

名。例如書名《網際網路行銷》的條碼號及分類號分別為「A149401」及「496/4432」。在進行探勘計算時，可藉由條碼號來找出書籍歸屬的分類號。本

研究將分類號前面兩個數字相同的書籍，視為讀者具有相同借閱興趣傾向的書籍，例如書籍分類號490-499，其也可反映出中國圖書分類法的規則。

條碼號	索書號	書名
A175571	490.59/3333	電子商務理論與實務陳宏基主編
A149401	496/4432	網際網路行銷(第三版)Paul Hesson著 李廷平譯
A149425	496/4449	物流與運籌管理 觀念、構架與整合廖建義作
A167652	496/4448	物流與運籌管理 觀念、構架與整合廖建義作
A175568	496/4449	物流與運籌管理 觀念、構架與整合廖建義作
A149424	494/2274	談判學Roy J. Lewicki, David M. Saunders, John W. Minton著 馮建君編譯
A150469	494/2274	談判學Roy J. Lewicki, David M. Saunders, John W. Minton著 馮建君編譯
A170710	494/2274	談判學Roy J. Lewicki, David M. Saunders, John W. Minton著 馮建君編譯
A149406	494/11237	策略管理個案集2001/司徒達賢主編
A183009	494/11237	策略管理個案集2001/司徒達賢主編
A191710	494/11237	策略管理個案集2001/司徒達賢主編
A191723	494.5/4402	設施規劃與物流中心設計林立千著
A149404	494.5/4402	設施規劃與物流中心設計林立千著
A174988	494.5/4402	設施規劃與物流中心設計林立千著
A191722	494.5/4402	設施規劃與物流中心設計林立千著
A149578	494.57/7613	全球物流發展趨勢下之策略策略與實務(第二版)[編]
A149381	498.93/5276	歐洲連鎖店經營 德、英、西班牙、葡萄牙、波蘭、法國、意大利(中華民國對外貿易發展協會)[編]
A149380	498.2/4438	歐亞零售連鎖發展現況及趨勢研究(義大利、西班牙及葡萄牙)林靜美、陳 煥芳[著]
A149382	498.9305549/262	歐亞零售連鎖店市場與拓銷策略(台灣研究)
A084868	498.9305549/262	歐亞零售連鎖店市場與拓銷策略(台灣研究)
A149379	496/2248	如何設立海外行銷據點(台北地區)外貿協會研究
A149277	496.3/2173	如何透過大陸百貨商標關稅聯合拓銷大陸市場(國際/學術研討)
A149392	496.5/3481	顧客關係管理深度解析 執行以客戶為中心的企業轉型策略ARC企業管理顧問公司編
A149395	496.5/3481	顧客關係管理深度解析 執行以客戶為中心的企業轉型策略ARC企業管理顧問公司編
A149398	496.3/1656	顧客關係管理深度解析 執行以客戶為中心的企業轉型策略ARC企業管理顧問公司編
A149399	496/49702	顧客關係管理深度解析 執行以客戶為中心的企業轉型策略ARC企業管理顧問公司編
A149426	496/4122	國際行銷實務(第二版) 艾德蒙·柯利(Jeffrey Edmund Curry)著 吳曉暉譯

圖 3 書籍的條碼號、分類號(索書號)、及書名資料

接下來說明所建置之推薦系統，其在探勘訓練資料的執行過程。圖 4 為在「探勘方式」欄位中，點選「一般性探勘」功能的探勘畫面，分別在「探勘讀者編號」欄位中填入欲探勘的讀者編號、及在「最小支持度」和「最小信賴度」欄位中填入數值，

分別以 0.5% 及 60% 為例，經由第 3 節所描述之演算法的探勘過程，並以前置項目組包含有最多項目之關聯規則做為推薦的依據，可在「書籍推薦」欄位中顯示出探勘的結果，如圖 5。

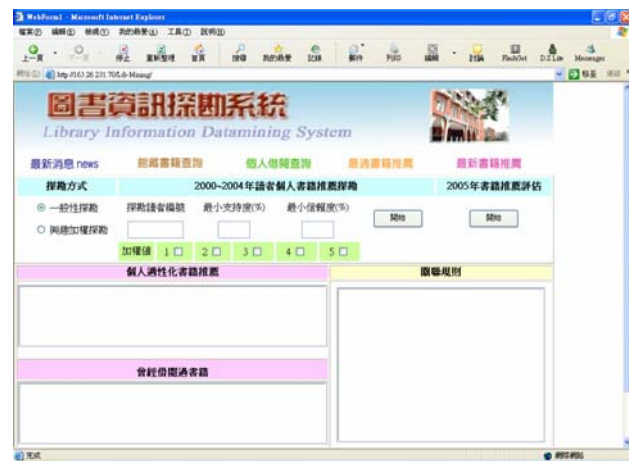


圖 4 發掘讀者適性化之書籍推薦的探勘執行畫面

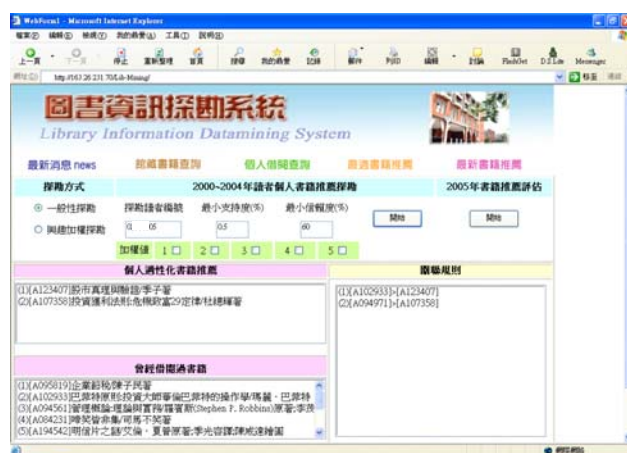


圖 5 發掘讀者適性化之書籍推薦的探勘結果畫面

圖 6 為在「探勘方式」欄位中，點選「興趣加權探勘」功能的探勘畫面，分別在「探勘讀者編號」欄位中填入欲探勘的讀者編號、在「最小支持度」和「最小信賴度」欄位中填入數值，分別以 0.5% 及 60% 為例，點選加權值的數目及值(以分別點選

加權值 1、2、及 3 為例)，經由第 4 節所描述之演算法的探勘過程，在作為推薦依據的關聯規則中，其前置項目組包含的項目數量為大於等於前一探勘方法之前置項目組包含的項目數量，可在「書籍推薦」欄位中顯示出探勘的結果，如圖 7。

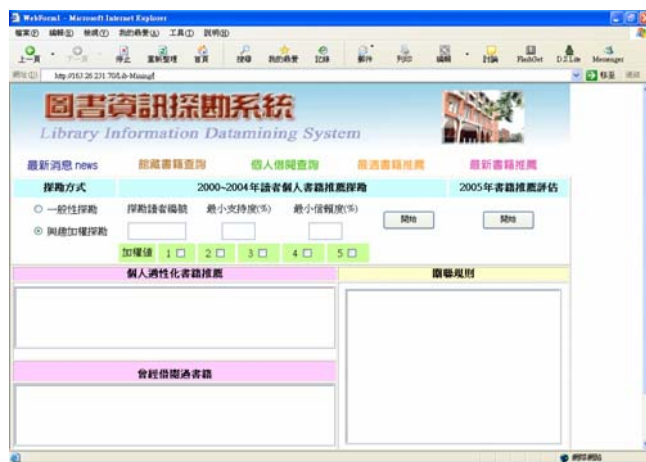


圖 6 考量借閱興趣發掘讀者適性化之書籍推薦的探勘執行畫面

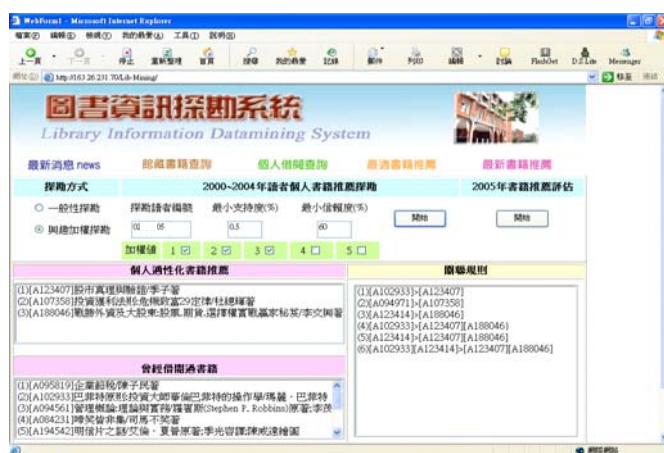


圖 7 考量借閱興趣發掘讀者適性化之書籍推薦的探勘結果畫面

我們以最後一年(2005)的借閱資料做為驗證資料,以評估在前面訓練資料中所找出之書籍推薦的結果成效。本探勘系統評估驗證資料的執行過程如下:圖 8 為在「一般性探勘」功能的探勘畫面中,並點選「2005 書籍推薦評估」欄位中的「開始」功能。在 6 年借閱資料中,會同時出現在前面 5

年及最後之借閱資料中的讀者共有 6,532 位,我們以此 6,532 位讀者做為評估的對象。在這 6,532 位讀者中,共有 1,075 位讀者在最後一年的借閱書籍中,包含有任一本書系統利用第 3 節之探勘方法所推薦的書籍,其推薦出現率為 16.5%(1075/6532)。



圖 8 發掘讀者適性化之書籍推薦的評估執行畫面

圖 9 為在「興趣加權探勘」功能的探勘畫面中,點選「2005 書籍推薦評估」欄位中的「開始」

功能,仍以此 6,532 位讀者做為評估的對象。在這 6,532 位讀者中,共有 1,217 位讀者在最後一年的

借閱書籍中，包含有任一本系統利用第 3 節之探勘方法所推薦的書籍，其推薦出現率為 18.6% (1217/6532)。在探勘的計算過程中，由於有些項目

組是經由加權之後成為高頻項目組，找出的關聯規則相較於前一探勘方法為多，其書籍推薦的數量也會較多，因此推薦出現率也會提升。



圖 9 考量借閱興趣發掘讀者適性化之書籍推薦的評估執行畫面

本探勘系統分別以發掘讀者個人適性化之書籍推薦及考量借閱興趣發掘讀者個人適性化之書籍推薦做為探勘的目標，從以上驗證資料的結果顯示，本研究對讀者個人之書籍推薦所設計的探勘方法，確實可反映出書籍推薦的成效，其中若能進行實際的書籍推薦作業，讓讀者接受到書籍推薦的訊息、及延長驗證的期間，則書籍推薦的出現率應可再提升。在探勘的計算過程中，設定參數值的不同，往往會影響找出的書籍借閱的傾向特徵，對書籍推薦時所導致的出現率也會受影響。因此，圖書館管理者可依其經驗及需求而設定各參數值，其發掘出的書籍推薦，將可得到更令讀者滿意的結果。

陸、結論

在圖書館服務中書籍借閱是最常被讀者利用的項目之一，讀者依其興趣及需求借閱書籍，這些曾經被借閱過的書籍中，除了可反映出讀者對書籍的借閱興趣傾向，也可顯示書籍之間的關聯性，若

能從借閱資料中找出書籍之間的關聯性，做為發掘讀者個人適性化之書籍的推薦依據，必可提供相當有用的參考資訊。目前已有許多相關研究探討利用資料探勘技術以提昇圖書館的經營服務(吳安琪，2001；黃夙賢等，2001；陳慶瑄，2000；孫冠華，2003)，但鮮少直接針對讀者個人、及考量讀者個人之借閱興趣進行加權的分析與計算。

本文以讀者之借閱資料做為探勘的資料來源，並以某一讀者做為探勘目標，設計兩個探勘關聯規則的方法，分別發掘讀者個人適性化的書籍推薦、及考量借閱興趣發掘讀者個人適性化的書籍推薦。在第一部份的探勘計算中，我們設計一個方法探勘關聯規則，其前置項目組必須包含於此讀者之借閱資料中，相較於 Apriori 演算法及 CDAR 演算法，我們所設計之探勘方法將可更有效率找到所要的項目組。在第二部份的探勘計算中，我們從讀者之借閱資料中考量其借閱興趣，提出一個方法探勘興趣加權關聯規則，其前置項目組必須包含於此讀

者之借閱資料中，相較於以前探勘加權關聯規則的演算法，我們所設計之探勘方法將可更容易、有效率找到所要的項目組。根據以上興趣加權關聯規則

所顯示的傾向特徵，其書籍推薦將更貼切於此讀者的借閱興趣與需求。

(收稿日期：2007 年 12 月 21 日)

參考文獻

- 卜小蝶 (民 87 年, 10 月)。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》，2，63-73。
- 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。
- 洪志淵 (2001)。《圖書流通記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。
- 黃夙賢、柯皓仁、楊維邦 (2001)。交通大學個人數位圖書資訊環境。在交通大學圖書館編，《新世紀數位圖書館與數位博物館趨勢研討會論文集》(頁 5-17)。新竹市：編者。
- 陳慶瑄 (2000)。《學習社群對電子圖書館個人化服務之影響》。未出版之碩士論文，國立中正大學資訊管理研究所，嘉義縣。
- 張苑菁 (2001)。《以模糊理論建構之圖書推薦系統》。未出版之碩士論文，淡江大學資訊工程研究所，台北縣。
- 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》，9，33-49。
- 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》。未出版之碩士論文，南華大學資訊管理學研究所，嘉義縣。
- 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，頁 81-111。
- 鄭玉玲 (2003)。《運用資料探勘技術實作數位圖書館上個人化之檢索與推薦服務-以南華大學圖書館為例》。未出版之碩士論文，南華大學資訊管理學研究所，嘉義縣。
- 顏秀珍，邱鼎穎，李御璽 (2001)。從大型資料庫中挖掘加權的關聯規則。在中國文化大學編，《全國計算機會議論文集》(頁 I117-I128)。台北市：編者。
- Agrawal, R., Imielinski, T. & Swami, A. (1993). Mining association rules between sets of items in very large database. In Peter Buneman and Sushil Jajodia (Eds.), *Proceedings of the ACM SIGMOD Conference on management of data* (pp. 207-216). New York: ACM Press.
- Agrawal, R. & Srikant, R. (1994). Fast algorithms for mining association rules. In Jorge B. Bocca, Matthias Jarke & Carlo Zaniolo (Eds.), *Proceedings of the 20th international conference on very large data bases* (pp. 487-499). San Mateo, California: Morgan Kaufmann.
- Berry, M. J. A. & Linoff, G. S. (2004). *Data mining techniques for marketing, sales, and customer support* (2nd ed.). New York: John Wiley.
- Cai, C. H., Fu, W. C., Cheng, C. H. & Kwong, W. W. (1998). Mining association rules with weighted items. In Barry Eaglestone, Bipin C. Desai, and Jianhua Shao (Eds.), *Proceedings of the international database engineering and applications symposium (IDEAS)* (pp. 68-77). Piscataway, NJ: IEEE.

- Chen, M. S., Han, J., & Yu, P. S. (1996). Data mining: An overview from a database perspective. *IEEE Transactions on Knowledge and Data Engineering*, 8(6), 866-883.
- Coenen, F., Leng, P., & Ahmed, S. (2004). Data structure for association rule mining T-trees and P-trees. *IEEE Transactions on Knowledge and Data Engineering*, 16(6), 774-778.
- Han, J. & Kamber, M. (2006). *Data mining: concepts and techniques*. (2nd ed.). San Mateo, California: Morgan Kaufmann.
- Han, J., Pei, J., & Yin, Y. (2000). Mining frequent patterns without candidate generation. In Weidong Chen, Jeffery F. Naughton and Philip A. Bernstein (Eds.), *Proceedings of the 2000 ACM SIGMOD international conference on management of data table of contents* (pp. 1-12). New York: ACM Press.
- Han, J., Pei, J., Yin, Y., & Mao, R. (2004). Mining frequent patterns without candidate generation: A frequent-pattern tree approach. *Data Mining and Knowledge Discovery*, 8(1), 53-87.
- Hui, S. C. & Jha, G. (2000). Data mining for customer service support. *Information and Management*, 38, 1-13.
- Huang, Y. P., Ouyang, C. P., & Ke, Y. H. (2001). A fuzzy approach to fulfilling personalized service through association rules derived from large databases. In Institute of Electrical and Electronics Engineers and IEEE Neural Networks Council (Eds.), *Proceedings of the 10th IEEE international conference on fuzzy systems* (pp. 272-275). Piscataway, NJ: IEEE.
- Liu, P. Q., Li, Z. Z., & Zhao, Y. L. (2004). Effective algorithm of mining frequent itemsets for association rules. In IEEE Systems, Man and Cybernetics Technical Committee on Cybernetics (Ed.), *Proceedings of the third international conference on machine learning and cybernetics* (pp. 1447-1451). Piscataway, NJ: IEEE.
- Li, Z. C., He, P. L., & Lei, M. (2005). A high efficient AprioriTid algorithm for mining association rule. In IEEE Systems, Man and Cybernetics Technical Committee on Cybernetics (Ed.), *Proceedings of the fourth international conference on machine learning and cybernetics* (pp. 1812-1815). Piscataway, NJ: IEEE.
- Ou, J., Lin, S., & Li, J. (2001). The personalized index service system in digital library. In Hongjun Lu (Ed.), *Proceedings of the third international symposium on cooperative database systems for advanced applications* (pp. 92-99). Los Alamitos, Calif.: IEEE Computer Society.
- Park, J. S., Chen, M. S., & 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. & Agrawal, R. (1995). Mining generalized association rules. In Umeshwar Dayal, Peter M. D. Gray and Shojiro Nishio (Eds.), *Proceedings of the 21th international conference on very large data bases* (pp. 407-419). San Francisco, Calif. Morgan Kaufmann.
- Tsay, Y. J. & Chang-Chien, Y. W. (2004). An efficient cluster and decomposition algorithm for mining association rules. *Information Sciences*, 160, 161-171.
- Tsay, Y. J. & Chiang, J. Y. (2005). CBAR: An efficient method for mining association rules. *Knowledge-Based Systems*, 18, 99-105.
- Yue, S., Tsang E., Yeung, D., & Shi, D. (2000). Mining fuzzy association rules with weighted items. In IEEE Systems, Man, and Cybernetics Society (Ed.), *Proceedings of the 2000 IEEE international conference on systems, man, and cybernetics* (pp. 1906-1911). Piscataway, NJ: IEEE.