利用資料探勘技術分析圖書館讀者個人化之書籍推薦

陳垂呈　　戴良安　　王筱薇

南台科技大學資訊管理研究所

E-mail: ccchen@mail.stut.edu.tw, { n9090023, m9390229}@webmail.stut.edu.tw

摘要

在讀者到圖書館借閱書籍的過程中，圖書館所扮演的角色往往是被動的，如何以主動地推薦書籍的方式來吸引讀者到館借閱，進而提昇書籍的借閱率與利用率，是圖書館管理上必須思考的問題之一。在本篇論文中，我們以讀者之借閱資料為探勘的資料來源，每一筆借閱資料記錄有讀者曾經借閱過的書籍項目，利用資料探勘(data mining)技術分別從以下兩方面來探討如何發掘讀者個人化的書籍推薦：一是考量書籍借閱無次序性，我們以某一讀者之借閱資料X為探勘的目標，探勘前置項目組為與X有相關的關聯規則$Y \rightarrow Z$, $Y \subseteq X$、$Y \cap Z = \varnothing$，X、Y、Z分別為包含一個或以上之書籍項目所形成的項目組，從關聯規則所顯示出的傾向特徵，可發掘此一讀者個人化最適性的書籍推薦。二是考量書籍借閱有次序性，我們仍以某一讀者之借閱資料$\langle X \rangle$為探勘的目標，探勘最大次序$\langle U, V \rangle$, $\langle U \rangle \subseteq \langle X \rangle$、$\langle U \rangle \cap \langle V \rangle = \varnothing$、$\langle V \rangle \cap \langle X \rangle = \varnothing$，$\langle X \rangle$、$\langle U \rangle$、$\langle V \rangle$分別為包含一個或以上之書籍項目所形成的有次序項目組，從最大次序所顯示出的傾向特徵，可發掘具有借閱次序之此一讀者個人化最適性的書籍推薦。我們根據所提出的方法，設計與建置一個讀者個人化最適性之書籍推薦系統。此探勘結果，對圖書館在擬訂最適性之讀者個人化書籍推薦時，可以提供非常有用的參考資訊。

關鍵詞：資料探勘、關聯規則、次序、圖書館、書籍推薦。

1. 簡介

藉由資訊技術的支援，電子化圖書館的經營方式也日漸成熟，圖書館所提供的服務也愈來愈多樣化，電子圖書、網際網路、儲存光碟、多媒體等科技媒體的出現，帶動了資訊儲存及檢索的新紀元。圖書館改變以往傳統搜尋書籍資料的方式，利用國內外網路上的豐富資源，配合各類電子媒介的輔助，能使讀者能以最少的時間，即可享受到最大的服務效益。但如何將傳統圖書館被動式的服務方式，轉變成以主動積極的方式來吸引讀者到館借閱，進而提昇讀者的借閱率及圖書館的利用率，是圖書館經營與管理上必須探討的課題之一。

圖書館中每天均有相當大量的書籍被借閱，在讀者曾經借閱過的書籍資料中，往往隱藏著書籍之間的關聯性，例如讀者借閱了一本「C語言程式設計」書籍，我們會發現其中也會有借閱「資料結構」書籍、或是一些相關書籍的傾向。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，即成為圖書館管理者必須思考的問題之一。在圖書館服務的項目中，以讀者個人化為中心的服務理念，是針對每一讀者個人化的資訊需求，調整圖書館的讀者服務，把最適性的館藏資訊主動告知讀者，進而提昇讀者的滿意度與對圖書館的忠誠度。

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，並以某一讀者之借閱資料為探勘的目標，利用資料探勘(data mining)技術分別從以下兩方面來發掘讀者個人化最適性的書籍推薦：

(1) 考量書籍借閱無次序性：我們以某一讀者之借閱資料X為探勘的目標，從借閱資料中探勘前置項目組為與X有相關的關聯規則$Y \rightarrow Z$, $Y \subseteq X$、$Y \cap Z = \varnothing$，X、Y、Z分別為包含一個或以上之書籍項目所形成的項目組。由於$Y \subseteq X$，表示關聯規則與此一讀者有相關性，並由關聯規則所顯示出的傾向特徵，可發掘出此一讀者個人化最適性的書籍推薦。

(2) 考量書籍借閱有次序性：我們仍以某一讀者之借閱資料$\langle X \rangle$為探勘的目標，從借閱資料中探勘最大次序$\langle U, V \rangle$, $\langle U \rangle \subseteq \langle X \rangle$、$\langle V \rangle \cap \langle X \rangle = \varnothing$，$\langle X \rangle$、$\langle U \rangle$、$\langle V \rangle$分別為包含一個或以上之書籍項目所形成的有次序項目組。由於$\langle U \rangle \subseteq \langle X \rangle$，表示最大次序與此一讀者有相關性，並由最大次序所顯示出的傾向特徵，可發掘具有借閱次序之此一讀者個人化最適性的書籍推薦。

經由上述兩種方法的探勘，我們可以發掘讀者曾經借閱過之書籍間的關聯性及次序性。我們根據所提出的方法，設計與建置一個讀者個人化最適性之書籍推薦系統。此探勘結果，對圖書館在擬訂最適性之讀者個人化書籍推薦，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。

本篇論文的架構如下：下一節中，我們介紹資料探勘技術、及其在圖書館服務之應用的相關研究；第3節中，我們考量書籍借閱無次序性，設計一個演算法來找出與某一讀者之借閱資料有相關的關聯規則，並藉此發掘此一讀者個人化最適性的書籍推薦；第4節中，我們考量書籍借閱有次序性，設計一個演算法來找出與某一讀者之借閱資料有相關的最大次序，並藉此發掘具有借閱次序之此一讀者個人化最適性的書籍推薦；第5節中，我們根據所提出的方法，設計與建置一個讀者個人化最適性之書籍推薦系統；最後，我們在第6節中做一結論。

2. 相關研究

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，其可完成以下工作或更多：分類(classification)、關聯規則(association rules)、分群(clustering)、次序相關分析(sequential pattern analysis)、及預測(forecasting)等[6]，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[7]。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，也造成書籍的借閱率不佳。

利用資料探勘技術於圖書館經營服務之應用的相關研究有：[1]曾利用*k*-means的方法來形成學習社群，以支援電子圖書館之個人化服務；[2]曾利用關聯規則於數位圖書館的個人化服務及管理；[3]曾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，以提昇圖書館之借閱率與讀者忠誠度等目的；[4]曾利用資料探勘技術來找出讀者與圖書之間的一般化關聯規則，做為讀者之新書推薦的依據；[5]曾以模糊理論(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，如何貼切地提供讀書個人化的書籍推薦，已成為提昇圖書館之經營與服務最重要研究的課題之一，也是資料探勘技術重要的應用主題之一。

Agrawal等人[8]首先提出擷取關聯規則來顯示出項目之間的關聯性，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$X \cap Y = \varnothing$，其中X稱之為前置項目組，Y稱之為後置項目組。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：在所有的交易集合中，同時包含有$X \cup Y$的比率值，即s=(同時包含有$X \cup Y$的交易數量)/(總交易數量)；信賴度c的定義為：在包含有X的交易集合中，也同時包括有Y的比率值，即c=(同時包含有$X \cup Y$的交易數量)/(包含有X的交易數量)。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。

在眾多擷取關聯規則的方法中，以Apriori演算法[9]是最具代表性，其探勘過程主要分成兩個階段：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就稱之為大項目組(large itemsets)，若某k-項目組滿足最小支持度，即稱之為大k-項目組(large k-itemsets)，$k \geq 1$，以$large_k$表示之；然後，就根據前階段所找出的大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為大3-項目組，假如關聯規則AB→C滿足最小信賴度，則此關聯規則成立，擷取關聯規則的相關研究可參考[10, 11]。

我們說明Apriori演算法[10]擷取關聯規則的步驟如下：

(1) 找出$large_{k-1}$，k>1，若為∅，則停止執行。
(2) 由(1)中組合任兩個有k-2項目相同的$large_{k-1}$，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(1)中，假如成立就保留此$itemset_k$，否則就刪除。
(4) 再檢查由(3)所擷取的$itemset_k$是否滿足最小支持度，假如符合就成為$large_k$，否則就刪除。
(5) 計算$large_k$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(1)找$large_{k+1}$，直到無法產生大項目組為止。

傳統關聯規則只考量其項目是否出現，並未考量項目出現的次序性，但考量現實的交易情形往往是包含了購買時間的前後關係。次序相關分析是在一群有次序的交易資料中，找出會時常出現的次序型樣(sequential pattern)，藉此可分析消費者有次序性的購買行為，其特點在於型樣中每個項目的前後位置為型樣的識別特徵之一，也就是型樣中的項目之間是含有次序性，因此在尋找循序樣式時，會有一個用以決定項目先後次序的衡量方式，如時間，所有的項目或項目集合依據該衡量方式做次序排列，而尋找次序型樣就是要在這些有次序的資料中，找出有趣的規則。例如，某一讀者在時點1借閱了A-書籍，在時點2又借閱了B、C兩本書籍，在時點3又借閱了D-書籍，經過排序與分析所有讀者的借閱資料之後，如果ACD是找出的次序型樣，表示許多讀者在借閱了A-書籍之後，會有借閱C、D-書籍的傾向。次序相關分析的相關研究可參考[12, 13]。

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，利用資料探勘技術分別從下兩方面來探討如何發掘讀者個人化最適性的書籍推薦：一是考量書籍項目之借閱無次序性；二是是考量書籍項目之借閱有次序性。藉由找出的規則型樣所顯示出的傾向特徵，可發掘此一讀者個人化最適性之書籍推薦、及發掘具有借閱次序之此一讀者個人化最適性的書籍推薦。

3. 發掘讀者個人化之書籍推薦

在此一章節中，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，其中的書籍項目都已事先排序過，並以某一讀者為探勘的目標，利用資料探勘技術做為發掘讀者個人化之書籍推薦的依據。此章節共分為兩小節如下：第3.1節中，我們設計一個演算法來發掘讀者個人化最適性的書籍推薦；第3.2節中，我們以一實例做說明。

3.1 擷取與某一讀者有相關的關聯規則

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：

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

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$Y \subseteq X$，表示關聯規則與此一讀者具有相關性，且若Y愈相似於X，則關聯規則的相關性愈強，其借閱的傾向性也愈強。因此，藉由關聯關規則的傾向性，可做為發掘此一讀者個人化之書籍推薦的依據。

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Apriori演算法，直接組合X中之項目與非X中之項目而形成的項目組，並判斷這些項目組是否為大項目組，擷取的過程說明如下：

(1) 從X中及非X中找出$large_1$，且必須至少各包含有一項。
(2) 由(1)中，組合包含於X中之任一$large_1$與非包含於X中之任一$large_1$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$large_2$，否則就刪除。
(3) 找出所有的$large_{k-1}$，$k>2$。
(4) 由(3)中，組合任兩個有k-2項目相同的$large_{k-1}$，形成$itemset_k$。
(5) 判斷由(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(3)中(若$itemset_{k-1} \cap X = \varnothing$、或$itemset_{k-1} \cap$非$X = \varnothing$，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(5)所找出的$itemsert_k$是否滿足最小支持度，假如符合就成為$large_k$，否則就刪除。
(7) 計算$large_k$所形成的關聯規則，其形式為：$Y \rightarrow Z$, $Y \subseteq X$、$Y \cap Z = \varnothing$，$Y \cup Z$是$\in large_k$。
(8) 跳至(3)繼續找出$large_{k+1}$，直到無法產生大項目組為止。

從以上演算法的(2)步驟開始，我們所擷取出的$large_k$，必定為$large_k \cap X \neq \varnothing$，表示與此一讀者之借閱資料有相關性，計算大項目組所形成的關聯規則$Y \rightarrow Z$, $Y \subseteq X$，若滿足最小信賴度，則我們做成以下的定義：藉由關聯規則$Y \rightarrow Z$所顯示出的借閱傾向，Z為此一讀者個人化最適性之書籍推薦，且Y愈相似於X，則借閱的傾向性也愈強。

3.2 實例說明

我們以表1的借閱資料庫D來進行分析，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，最小支持度為40% (即最小支持數量為2)，最小信賴度設為60%。假設目前欲探勘之讀者為R_5，其借閱資料為BC，以下我們說明發掘此一讀者最適性之書籍推薦的探勘過程。

表 1 借閱資料庫 *D*

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

擷取與 BC 有相關之大項目組的過程如下：

$itemset_1$
B
C
A
D
E

掃瞄 D →

$itemset_1$	出現次數
B	4
C	4
A	2
D	1
E	3

≥2 →

$large_1$	出現次數
B	4
C	4
A	2
E	3

$itemset_2$
AB
BE
AC
CE

掃瞄 D →

$itemset_2$	出現次數
AB	1
BE	3
AC	2
CE	2

≥2 →

$large_2$	出現次數
BE	3
AC	2
CE	2

$itemset_3$
BCE
ACE

步驟(5)、(6) →

$itemset_3$	出現次數
BCE	2
ACE	1

≥2 →

$large_3$	出現次數
BCE	2

無 4-項目組

我們以大 3-項目組 BCE 為例，計算形成的關聯規則 BC→E，其信賴度為 2/3=67%，滿足最小信賴度，因此，此一讀者最適性之書籍推薦為 E。

4. 考量具有借閱次序之讀者個人化的書籍推薦

對於讀者在不同時間下借閱過之書籍項目，往往也反映出書籍被借閱的次序性，其可能顯示讀者在需求上或是書籍專業上的次序性。在此章節中，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曾經借閱過的書籍項目，其中的書籍項目都依借閱的時間事先排序過，並以某一讀者為探勘的目標，利用資料探勘技術做為發掘具有借閱次序之讀者個人化的書籍推薦的依據。此章節共分為兩小節如下：第4.1節中，我們設計一個演算法來發掘具有借閱次序之讀者個人化最適性的書籍推薦；第4.2節中，我們以一實例做說明。

4.1 擷取與某一讀者有相關之最大次序

我們說明一些名詞的定義如下：

(1) 次序(sequence)：為有次序性之k個項目組所組成的集合，$k\geq 1$，以$sequence_k$表示之，例如一讀者先借閱A-書籍與B-書籍，然後再借閱C-書籍、D-書籍與E-書籍，則以<AB, CDE>來表示一個次序，任何一讀者有次序性之借閱資料可視為一次序。

(2) 次序支持度=包含有某一次序之交易資料的數量/全部交易資料的數量。若某一次序的支持度大於或等於最小次序支持度，則稱之為大次序(large sequence)，一大次序若包含有k個項目組，則稱為大k-次序，以$Lsequence_k$表示之。

(3) 最大次序(maximal sequence)：若一大次序$<X_1, X_2, \ldots, X_n>$包含於另一大次序$<Y_1, Y_2, \ldots, Y_m>$中，則$X_1\subseteq Y_{i1}$, $X_2\subseteq Y_{i2}$, …, $X_n\subseteq Y_{in}$，其中$i1<i2< \ldots <in\leq m$，例如一大次序<A, CD>包含於另一大次序<AB, CDE, F>中。若一大次序不包含於任何的大次序中，則稱此大次序為最大次序。

次序相關分析的目的，就是要找出所有能滿足最小支持度的最大次序之集合，並分析最大次序中的項目集合，以找出項目的先後次序關係。此一小節的目的就是：從讀者有次序性的借閱資料中，找出滿足最小次序支持度的最大次序，並從最大次序所顯示出的借閱傾向特徵，來發掘具有借閱次序之讀者最適性的書籍推薦。

假設欲探勘之讀者的借閱資料為$<X>$，$<X>$為一次序，根據次序的定義，我們必須找出以下形式的最大次序：

$<U, V>$：$<U>\subseteq<X>$, $<U>\cap<V>=\varnothing$, $<V>\cap<X>=\varnothing$, $<U>$、$<V>$皆為一次序，$<U, V>$為最大次序。

其顯示出的借閱傾向為：若讀者曾經借閱過$<U>$書籍，則接下來也會有借閱$<V>$書籍的傾向。由於$<U>\subseteq<X>$，表示最大次序與此一讀者具有相關性，且若$<U>$愈相似於$<X>$，則最大次序的相關性愈強，其借閱的傾向性也愈強。因此，藉由最大次序的傾向性，可做為發掘具有借閱次序之此一讀者個人化之書籍推薦的依據。

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，而修改次序相關演算法的步驟[12, 13]，直接組合$<X>$中之項目與非$<X>$中之項目而形成的項目組，並判斷這些項目組是否為大次序，擷取的過程說明如下：

(1) 對所有讀者之借閱資料中的書籍項目，依借閱時間做遞增或遞減的方式排序。

(2) 從$<X>$中及非$<X>$中找出$Lsequence_1$，且必須至少各包含有一項。

(3) 從(2)中組合兩個$Lsequence_1$形成$Lsequence_2$，其中前面項目$\subseteq<X>$，而後面項目$\not\subset<X>$。

(4) 找出$Lsequence_{k-1}$，$k>2$。

(5) 由(4)中排列組合任兩個有k-2 項目相同的$Lsequence_{k-1}$，形成$sequence_k$且其中前面項目$\subseteq<X>$。

(6) 判斷由(5)所找出的$sequence_k$其所有包括的$sequence_{k-1}$之子集合是否都出現在(4)中，若$sequence_{k-1}\cap<X>=\varnothing$、或$sequence_{k-1}\cap$非$<X>=\varnothing$，則不列入考慮，假如成立就保留此$sequence_k$，否則就刪除。

(7) 再檢查由(6)所擷取的$sequence_k$是否滿足最小次序支持度，假如符合就成為$Lsequence_k$，否則就刪除。

(8) 跳至(4)找$Lsequence_{k+1}$，直到無法產生大次序為止。

(9) 確認$Lsequence_k$之最大化，即若一大次序包含於另一大次序中，則刪除掉。

從以上演算法的(3)步驟開始，我們所擷取出的$sequence_k$，其型樣必定為$<U, V>$且$<U>\subseteq<X>$，表示與此一讀者之借閱資料有相關性，且在(5)步驟中所排列組合形成的$sequence_k$，可以產生不同的項目次序。我們計算最大次序所形成的型樣$<U, V>$，做成以下的定義：藉由最大次序$<U, V>$所顯示出的借閱傾向，$<V>$為具有借閱次序之此一讀者個人化最適性的書籍推薦，且$<U>$愈相似於$<X>$，則借閱的傾向性也愈強。

4.2 實例說明

我們以表 2 的借閱資料庫D來進行分析，I={A, B, C, D, E}為所有書籍項目的集合，$R=\{R_1, R_2, R_3, R_4\}$為 4 筆讀者之借閱資料的集合，設定最小次序支持度為 50% (即最小支持數量為 2)。假設目前欲探勘之讀者為R_4，其借閱資料為<A, B>，以下我們說明發掘具有借閱次序之此一讀者最適性書籍推薦的探勘過程。

表 2 借閱資料庫 *D*

借閱資料	書籍項目
R_1	<A, BC, E>
R_2	<A, C, D>
R_3	<A, BC, D, E>
R_4	<A, B>

擷取與<A, B>有相關之大次序的過程如下：

$sequence_1$
<A>
<B>
<C>
<D>
<E>
<BC>

掃瞄 D →

$sequence_1$	出現次數
<A>	4
<B>	3
<C>	3
<D>	2
<E>	2
<BC>	2

≥2 →

$Lsequence_1$	出現次數
<A>	4
<B>	3
<C>	3
<D>	2
<E>	2
<BC>	2

$sequence_2$
<A, C>
<A, D>
<A, E>
<A, BC>
<B, C>
<B, D>
<B, E>

掃瞄 D →

$sequence_2$	出現次數
<A, C>	3
<A, D>	2
<A, E>	2
<A, BC>	2
<B, C>	2
<B, D>	1
<B, E>	2

≥2 →

$Lsequence_2$	出現次數
<A, C>	3
<A, D>	2
<A, E>	2
<A, BC>	2
<B, C>	2
<B, E>	2

$sequence_3$
<A, C, D>
<A, C, E>
<A, B, C>
<A, D, E>
<A, BC, E>
<A, B, E>

步驟(6)、(7) →

$sequence_3$	出現次數
<A, C, D>	2
<A, C, E>	2
<A, B, C>	2
<A, D, E>	1
<A, BC, E>	2
<A, B, E>	2

≥2 →

$Lsequence_3$	出現次數
<A, C, D>	2
<A, C, E>	2
<A, B, C>	2
<A, BC, E>	2
<A, B, E>	2

$sequence_4$
<A, C, D, E>
<A, B, C, D>
<A, B, C, E>

步驟(6)、(7) →

$sequence_4$	出現次數
<A, B, C, E>	2

≥2 →

$Lsequence_4$	出現次數
<A, B, C, E>	2

無$sequence_5$。

刪除被包含於其他大次序的大次序，可得到最大次序為：<A, B, C, E>及<A, C, D>。我們以最大次序<A, B, C, E>為例，其所顯示出的傾向特徵，可發掘出具有借閱次序之讀者R_4個人化最適性的書籍推薦為<C, E>。

5. 讀者個人化最適性之書籍推薦系統

我們利用前面章節所描述的方法，應用到探勘讀者個人化最適性之書籍推薦的系統實作上。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，以亂數隨機產生每一讀者的借閱資料，共產生 200 筆借閱資料，以下為此一系統探勘過程的執行畫面。圖 1 為系統的借閱資料，包含有「讀者編號」與「曾經借閱過的書籍」等欄位資料。

借閱資料 : 資料表

讀者編號	曾經借閱過的書籍
R0001	A,C,E,H,B,P,X
R0002	B,GU,I,Z
R0003	C,E,Q,J
R0004	C,D,E,H,V,N
R0005	E,H,G
R0006	K,B,C,E,F
R0007	E,C,H,L,P,N,B
R0008	A,G,E,B,H,V,P,X
R0009	D,E,C
R0010	B,I,D

記錄: 1 之 200

圖1 借閱資料

圖2為探勘畫面，其中包含有兩項功能選項：「發掘讀者個人無次序性之書籍推薦」與「發掘讀者個人有次序性之書籍推薦」。假設目前點選「發掘讀者個人無次序性之書籍推薦」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，經由第3章節演算法的探勘過程，可在「關聯規則」欄位中顯示出找到的關聯規則，並在「個

人化最適性的書籍推薦」欄位中顯示出探勘的結果，如圖2。

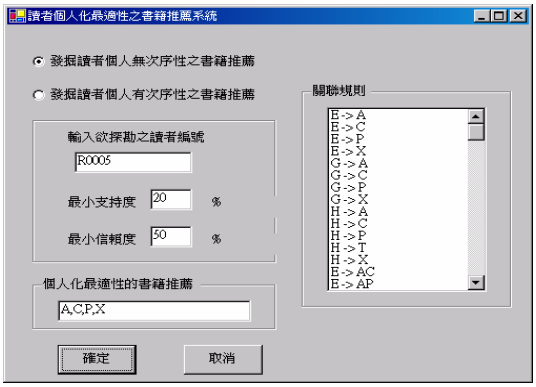

圖2 發掘讀者個人化最適性之書籍推薦的執行畫面

若點選「發掘讀者個人有次序性之書籍推薦」功能，則可在「個人化最適性的書籍推薦」欄位中顯示出包含有借閱次序之書籍推薦的探勘結果，如圖 3。

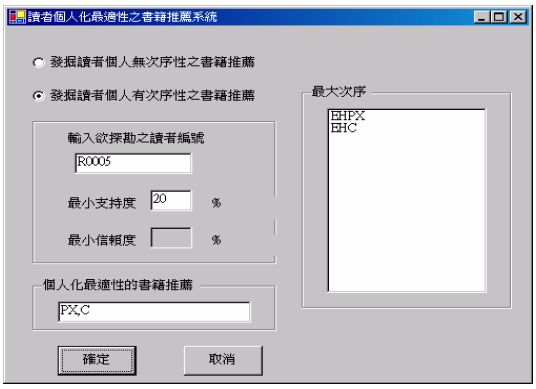

圖3 發掘具有借閱次序之個人化最適性的書籍推薦的執行畫面

6. 結論

圖書館除了蘊藏豐富的書籍與其他多樣的資料，也儲存了大量讀者借閱書籍的記錄，如何有效利用這些大量的借閱資料，找出對讀者有用的資訊，進而以主動積極的方式來吸引讀者到館借閱，是圖書館管理者必須考量的問題之一。在本篇論文中，我們以讀者之借閱資料為探勘的資料來源，並以某一讀者為探勘的目標，利用資料探勘技術分別從以下兩方面來探討如何發掘此一讀者個人化的書籍推薦：一是考量書籍借閱無次序性；二是考量書籍借閱有次序性。我們從探勘出的關聯規則及最大次序，可分別發掘出此一讀者個人化最適性的書籍推薦、及發掘具有借閱次序之此一讀者個人化最適性的書籍推薦。探勘的過程中，從資料的蒐集、分析、方法的設計、及結果的推導，可顯示出我們所設計之方法的有用性及創新。本篇論文之探勘結果，對圖書館在擬訂讀者個人化最適性的書籍推薦時，必可提供非常有用的參考資訊。

參考文獻

[1] 陳慶瑄。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，2000。

[2] 孫冠華。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，2003。

[3] 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。國立交通大學資訊科學研究所碩士論文，2001。

[4] 洪志淵。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，2001。

[5] 張苑菁。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，2001。

[6] M. S. Chen, J. Han, and P. S. Yu, “Data Mining: An Overview from a Database Perspective,” IEEE Trans. on Knowledge and Data Engineering, pp. 866-883, 1996.

[7] M. J. A. Berry and G. Linoff, Data Mining Techniques for Marketing, Sales, and Customer Support, New York: John Wiley, 1997.

[8] R. Agrawal, T. Imielinski, and A. Swami, “Mining Association Rules between Sets of Items in Very Large Database,” Proceedings of the ACM SIGMOD Conference on Management of Data, pp. 207-216, 1993.

[9] R. Agrawal and R. Srikant, “Fast Algorithms for Mining Association Rules in Large Database,” Proceedings of the 20th International Conference on Very Large Data Bases, pp. 487-499, 1994.

[10] J. S. Park, M. S. Chen. and P. S. Yu, “Using a Hash-Based Method with Transaction Trimming for Mining Association Rules,” IEEE Transactions on Knowledge and Data Engineering, Vol. 9, No. 5, pp. 813-825, 1997.

[11] R. Srikant and R. Agrawal, “Mining Generalized Association Rules,” Proceedings of the 21th International Conference on Very Large Data Bases, pp. 407-419, 1995.

[12] R. Agrawal and R. Srikant, “Mining Sequential Patterns,” Proceedings of the International Conference on Data Engineering (ICDE), 1995.

[13] R. Srikant and R. Agrawal, “Mining Sequential Patterns: Generalizations and Performance Improvements,” Proceedings of the Fifth International Conference on Extending Database Technology (EDBT), 1996.