

圖書資訊學刊
第七期(民國91年12月):頁81-94
National Taiwan University
Journal of Library
and Information Studies
No. 17 December 2002, PP. 81-94.
ISSN 1606-7509

使用者導向之圖書分類關聯分析研究

An Empirical Study on User-oriented Association Analysis of Library Classification Schemes

卜小蝶*

Hsiao-Tieh Pu

摘要

分類表是圖書館組織圖書資訊的重要工具，然而這類系統多半是以知識的學科分類來建立體系，著重在知識的層屬關係，較少關注使用者角度。本文即嘗試瞭解在分類表的層級結構(Hierarchical Structure)中，發掘讀者認知的分類號關聯(Association)其可能性。初步是以世新大學圖書館所提供兩年大量的借閱記錄為基礎，利用自動化方法分析讀者的相似借閱行為，藉以擷取類號之間的關聯。實驗結果顯示，根據讀者的借閱行為，許多相關類號不僅集中在分類表所設計的類別層級，同時也有不少類號是存在於其它不同的類別層級。針對這些關聯的特性，本文也加以分析歸納，並探討其限制與應用範疇。

關鍵詞：借閱記錄分析、關聯分析、圖書分類系統

* 世新大學資訊傳播學系副教授 (Associate Professor, Department of Information and Communications Shih Hsin University)

Abstract

Library classification schemes are mostly organized based on disciplines with a hierarchical structure. From the user point of view, some highly related yet non-hierarchical classes may not be easy to perceive in these schemes. This paper is to discover hidden associations between classes by analyzing users' usage of library collections. The proposed approach employs collaborative filtering techniques to discover associated classes based on the circulation patterns of similar users. Many associated classes scattered across different subject hierarchies could be discovered from the circulation patterns of similar users. The obtained association norms between classes were found to be useful in understanding users' subject preferences for a given class. Classification schemes can, therefore, be made more adaptable to changes of users and the uses of different library collections. There are implications for applications in information organization and retrieval as well. For example, catalogers could refer to the ranked associated classes when they perform multi-classification, and users could also browse the associated classes for related subjects in an enhanced OPAC system. In future research, more empirical studies will be needed to validate the findings, and methods for obtaining user-oriented associations can still be improved.

Keywords: Circulation Log Analysis; Association Analysis; Library Classification Scheme

壹、前言

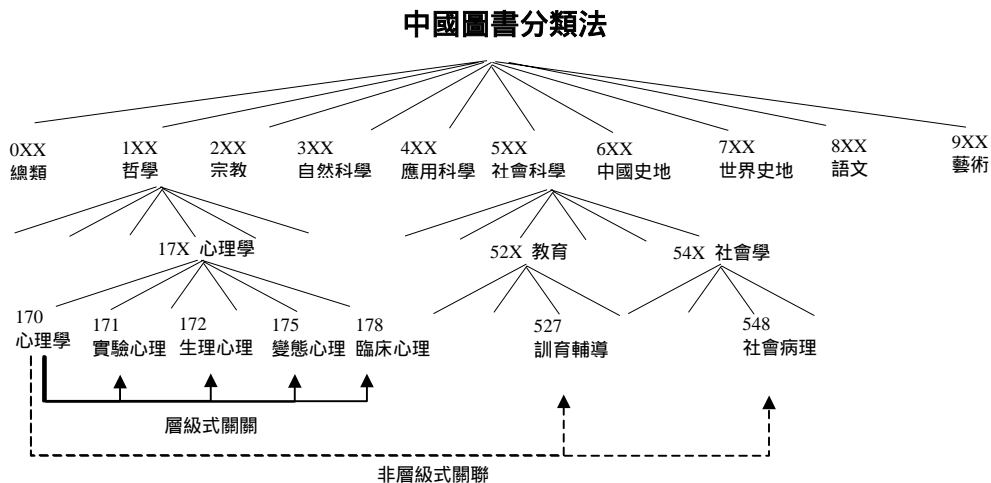
圖書分類表 (Library Classification Scheme) 是圖書館組織圖書資訊的重要工具，包括如美國的杜威 進分類法 (DDC, Dewey Decimal Classification) 美國國會圖書分類法 (LCC, Library of Congress Classification) 及台灣地區常用的中國圖書分類法 (CDC, Chinese Decimal Classification) 等，皆是歷史悠久、使用廣泛的系統。(註 1) 然而長久以來，這些分類表所扮演的角色多半以圖書的排架管理為主。(註 2) 事實上，隨著電子化、網路化環境的蓬勃發展，分類表在資訊組織及檢索方面的應用已逐漸擴大，例如 NetFirst、CyberDewey、CyberStacks 等計畫就是以 DDC 或 LCC 等分類表來組織網路資源；(註 3) 此外，也有人期望擴展類表中分類知識體系的價值，嘗試提昇

Web 線上公用目錄的分類瀏覽功能等。(註 4) 這些應用都顯示出分類表的重要性,也透露分類表的建立及維護必須具有彈性,才能因應多元變動的資源及使用者需求。然而這類系統多半是以知識的學科分類來建立體系,著重在知識的層屬關係(Hierarchical Relationships),(註 5)而較少以使用者角度來思考類表的設計。誠如 Solomon 所言:「由於欠缺對使用者使用語言、知識背景、及使用情境的瞭解,分類表經常不是那麼受到歡迎。但如何以使用者角度來發展分類表的方法及測試工作則更顯不足。」(註 6) 本文即嘗試瞭解在分類表的層級結構中,發掘讀者認知的分類號關聯其可能性。初步是以世新大學圖書館所提供二年大量的借閱記錄為基礎,利用自動化方法分析讀者的相似借閱行為,藉以擷取類號之間的關聯。實驗結果顯示,根據讀者的借閱行為,許多相關類號不僅集中在分類表所設計的類別層級,同時也有不少類號是存在於其它不同的類別層級。針對這些關聯的特性,本文也加以分析歸納,並探討其限制與應用範疇。

一般圖書分類表多採層級式架構(Hierarchical Structure),對使用者而言,分類表中有許多相關類別不一定會被安排在同一層級(Hierarchy)。換言之,某些相關類別並不易由分類的層級關係來看出其關聯。如圖一所示,CDC 是典型的層級式架構,最上層分為 大類(Class),每類代表一種主要學科,再細分為 小類(Division),每一小類再細分為目(Section)等等。當查詢有關 170(一般心理學)類別的書目時,與其在相同層級 17X(心理學)的類號,如 171(實驗心理學)、172(生理心理學)等,理想上應是找尋相關書目的最佳類別(圖中實箭號)。但透過讀者借閱記錄的相似性比對後,發現 170 与其它不同層級的類號,如 52X(教育)、54X(社會學)中的類號其實也有關聯,如 527(訓育輔導)或 548(社會病理學)等(圖中虛箭號)。若能獲取這些分散在其它層級的相關類號,則不僅有助於對讀者借閱興趣的掌握,同時對以讀者角度瞭解分類表的類號關聯也會有所助益。

筆者曾以世新大學圖書館所提供兩年大量的圖書借閱記錄為範圍,利用關聯(Association)及分群(Clustering)等資料探勘方法,初步分析圖書與讀者、讀者與讀者、及圖書與圖書間的隱藏關聯與規則,這些決策規則(Decision Rules)對瞭解並進而掌握讀者興趣有相當的參考價值。(註 7)而本文則進一步利用借閱記錄來瞭解分類表中類別之間的關聯。本研究所使用的分類表是以 CDC 為基礎,分析對象則以分類號的第三層級為主,即 000-999 的類號。而類號之間的關聯分析則是利用相似性比對方法,來推估(Similarity Estimation)相似借閱行為所反映出的類號關聯。這類方法普遍運用於協力式過濾(Collaborative Filtering)及資訊檢索等應用,(註 8)(註 9) 本文主要利用條件機率(Conditional-probability-based)及餘弦夾角(Cosine-based)等兩種方法進行實驗及比較。而研究方法的基本假設是:兩個類號的相似性可藉由分

析借閱此二類圖書的讀者來估計求得。



圖一 層級式分類表中相關類號分散情形舉例

實驗結果顯示，根據讀者的借閱行為，許多相關類號不僅集中在 CDC 所設計的類別層級，同時也有不少類號存在於其它不同的類別層級。本文進一步分析這些類號關聯，並依據其特性歸類為四種關聯，分別為層級式關聯（Hierarchical Association）、非層級式關聯（Non-Hierarchical Association）、熱門關聯（Popular Association）及錯誤關聯（Erroneous Association）等，並深入探討其特性及成因。

本研究所提出的方法主要是嘗試以讀者借閱行為的角度，來擷取各分類號之間的關聯。研究結果對於圖書資訊組織及檢索可以有相當廣泛的應用，例如編目人員可參考各分類號的相關類號排序，以作為多重分類的依據；而在編製或更新相關索引時，也可以針對各類號關聯而有所參酌；再者，當讀者在查詢資料時，系統也可擴展各類號的相關類別，以協助讀者瀏覽更多的分類知識及相關圖書；此外，這些相關類別也可作為發展新書推薦服務的基礎，例如將相關類別圖書推薦給借閱行為相似的讀者等。

貳、相關研究

有關分類學或分類表的研究向來是圖書資訊學的研究核心之一。（註 10）過去分類表多應用在圖書資訊的組織整理，但近年來隨著網路資源組織與知識管理研究的興

起，分類表的重要性也就再度受到關注，畢竟分類表是具體而微的人類知識表徵，其類目設計與結構的一致性都受到相當的肯定。(註 11)若能以此為基礎，再配合新時代、新環境的需求來加以改善，則將有助於其成為資訊組織或知識組織的關鍵性角色。例如 OCLC 就積極地擴展 DDC 成為知識組織的重要工具之一，(註 12)例如其利用一些重要的主題資源來增加 DDC 的詞彙 (Vocabulary)，包括將美國國會標題表 (LCSH) 或其它標題表 及重要的文件資料庫等，抽取其中詞彙並對應到合適的 DDC 類別之中。這樣的作法無非是想將 DDC 建構為一知識資料庫 (Knowledge Base)，再藉著這樣的基礎，來發展出更豐富多元的應用。例如可作為組織網路資源的分類架構，讓網路資源的整理能更為一致、有品質等。此外也可作為 OPAC 瀏覽分類知識的來源，讓使用者瞭解知識的層級關係與跨學科領域的主題分佈等。(註 13)

借閱記錄是瞭解讀者利用圖書館資源的重要線索，其應用範圍相當廣泛。一般而言，以其為對象的研究目的多在提供館藏發展 (Collection Development) 所需之管理資訊，如分析讀者的借閱類別及數量，以便調整館藏發展政策中各主題館藏的分佈，(註 14) 或分析各類讀者的閱讀模式 (Reading Patterns)，以便瞭解圖書館服務客群的需求與特性等。(註 15) 也有一些研究關心圖書架位的管理，如統計目前的借閱數量，以便預測未來的借閱情況。(註 16) 然而以借閱記錄來瞭解資訊組織效益的研究則並不多見，而這正也是本研究所欲探討的方向之一。

本文所使用類號關聯的相似性比對方法，其實在資料探勘 (Data Mining)、協力式資訊過濾 (Collaborative Information Filtering) 及資訊檢索 (Information Retrieval) 等領域都有相當多的探討。在資料探勘中有所謂的關聯規則分析，最常應用在商店的交易資料探勘 (Market Basket Analysis)。其功能主要是分析顧客的交易行為模式，特別是挖掘出那些商品可能經常會被一起購買。但這方面的分析在圖書資訊領域的應用相當少見，(註 17) 主要原因可能來自於交易記錄不易取得，加上圖書借閱記錄有相當多的限制，並不易獲得大量有意義的關聯，(註 18) 對此，本文會有深入探討。而在協力式資訊過濾的研究中，則相當重視推薦技術的應用，(註 19) 例如系統可長期、主動收集使用者的喜好 (Preference)，再加以分析及群集相似的使用者，以作為推薦的基礎。而在資訊檢索的研究中，文件與檢索需求的相關性比對向來是其關心重點，例如在向量空間 (Vector Space) 的檢索模式中，以 $TF*IDF$ (Term Frequency Theta Inverse Document Frequency) 權重公式來計算文件 (Document) 或檢索 (Query) 的相似性權值，並依據此來判斷文件與文件、文件與檢索之間的相似性，(註 20) 這也是本研究計算相似類號關聯所利用的方法之一。

參、研究方法與實驗結果

一、研究方法

(一) 資料收集

本文主要以世新大學圖書館所提供 1996/09/23 至 1998/09/22 兩年共計 171,284 筆的圖書借閱記錄為研究範疇。初步以瞭解中文類號關聯為主，因此排除其中的英文圖書借閱記錄後，共計有 166,065 筆中文圖書借閱記錄。研究進行當時的館藏規模約有 130,898 冊圖書可供流通，讀者數量約有 10,876 人。而由世新大學圖書館自動化系統所擷取的借閱記錄格式則包含了讀者証號、圖書登錄號、流水號、及借閱日期等四項欄位資料。再由讀者檔及書目檔分別擷取了讀者系所代碼及圖書分類號的前三碼等二項欄位資料，並去除流水號及將讀者証號轉換為單一序號，以維護讀者個人資料隱私。最後再將前述資料一併整合為實驗資料檔。

(二) 資料分析

一般而言，若要分析相似的借閱行為，是需要相當大量的讀者有相似的借閱模式才比較可行，同時也需長期收集觀察才較有意義。然而根據一些圖書館使用統計分析研究指出，圖書館中多數圖書利用的人並不多。(註 21) 而根據本研究實際分析上述二年的借閱記錄發現，在 130,898 筆可供流通的圖書中，僅有 33.1% 曾被借閱，而其中 32.1% 的圖書僅被借閱過一次，這些曾被借閱的圖書平均借閱率為 3.83 次。此外，這二年借書的讀者佔所有教職員生的比例並不低 (71.5%)，但平均每人借書率僅有 5.57 冊。若針對分類號來看，893 種類號的館藏圖書中，有 74.58% 的類號曾被借閱，平均每個類號有 110.34 位讀者借閱，但其標準差 274.02 卻相當高，顯示每個類號的借閱頻率落差頗大，許多借閱圖書多集中在某些類號。例如 857 (小說) 及 312 (電腦科學) 類就佔了超過 18% 的借閱率，這些類號可視為圖書館中的熱門借閱類別。從另一方面來看，每位讀者平均只借了 9.48 種類別的圖書，其標準差為 8.48，說明每位讀者的借閱類別種數其差異並不明顯。

此外，實驗中也針對資料的稀疏性 (Sparseness) 進行分析，主要作法是將測試資料轉換為一讀者與分類號對應的二維矩陣 (User-Class Matrix)，每位讀者對應的是一組其曾借閱的分類號，而每個分類號其實可視為一組各讀者借閱該類頻率的向量集合。實驗中分別建立了二個矩陣，以比較不同讀者層級對類號關聯分析的影響，一為以讀者個人為單位的 MI 矩陣，另一則為以讀者系所為單位的 MD 矩陣。表一所列即為二者的稀疏程度比較，很明顯地，以個別讀者為基礎的矩陣其稀疏程度相當高，而

即使是以讀者系所為基礎，其稀疏程度仍然偏高，這對關聯分析的進行自然形成相當阻礙，此部分將於以下章節討論。

表一 不同讀者層級的資料稀疏性分析結果

| 測試資料集 | 讀者數 | 分類號數量 | 有借閱記錄的數量 | 稀疏性 |
|---------------|-------|-------|----------|---------|
| MI (以個別讀者為基礎) | 7,779 | 666 | 73,697 | 0.98578 |
| MD (以讀者系所為基礎) | 58 | 666 | 8,009 | 0.79266 |

(三) 關聯分析方法

本研究的關聯分析假設其實與協力式資訊過濾的概念類似，亦即利用社群 (Community) 的共同興趣來決定資訊的價值及預測重要趨勢。(註 22) 舉例來說，如果有足夠的樣本，圖書之間或分類號之間的關聯就可以透過使用者 (即社群) 的相似借閱行為來獲得。而這些關聯初步可透過各種相似性比對方法來加以自動擷取。本研究主要採用以條件機率為基礎 (Conditional-Probability-based) 及以餘弦夾角為基礎 (Cosine-based) 等兩種方法。這些方法普遍為許多相關研究所採用，同時也相當簡便有效。有關方法的細節及優缺可參考筆者文章。(註 23)

二、實驗結果

(一) 相似性比對方法評估

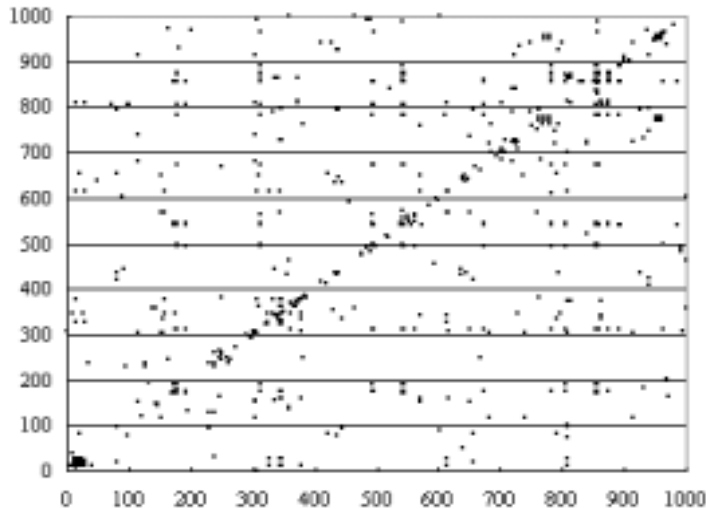
實驗目的之一是評估條件機率與餘弦夾角為基礎等二種不同相似性比對方法，其所獲得類號關聯的品質。首先我們隨機挑選了 10 個借閱數量中等的分類號，作為測試組，針對其中每個分類號，再分別由三種來源建立其正確的關聯類號，包括分類表中同屬其分類層級的類號 (hierarchically related)、非分類層級的分類號如參見的分類號 (non-hierarchically related) 及由上述二種方法所獲得經人工判斷相關的分類號等。本研究主要採用求全率 (Recall) 與求準率 (Precision) 作為評估指標，根據這二種方法所得到的關聯分類號，再分別與測試組作比較。實驗結果顯示餘弦夾角方法略勝一籌，主要原因是條件機率方法較易被借閱頻率低的類號所影響，這些類號雖然只有少數人借閱、且借閱次數也很低，但卻因與熱門類號在一起而很容易被視為關聯類號。反之，餘弦夾角方法對於具有較多相似借閱者的類號會給予較多的權重，因此比較不會受到低頻類號的影響。

而另一方面，本研究也嘗試觀察不同稀疏程度的交易記錄矩陣對於關聯分析的影響。如前述，在讀者與類號的對應矩陣中，MI 是以個別讀者為分析單位，而 MD 則以讀者系所為單位，二個矩陣中的讀者欄位分別為 7,779 人與 58 個系所。後者的稀疏程度雖比前者為低，而得以找出更多的關聯類號組，但也因此找出更多不易分析的關

聯類號。基於前述餘弦夾角方法的效益較佳，同時也較不受資料稀疏性的影響，本研究最後採用正規化餘弦夾角方法，(註 24) 並以個別讀者為擷取關聯類號的基礎。此外，實驗中也去除一些借閱頻率過低及過高的類號，以獲取較正確的類號關聯。

(二) 讀者認知的類號關聯分佈

為了比較讀者認知的類號關聯分佈與分類表中的類別層級分佈之差異，本研究設計了一 1,000x1,000 的類號矩陣來呈現其關聯類號的分佈情形，並加以觀察分析(如圖二)，圖中 X 軸與 Y 軸分別代表 000-999 的 1000 個類號。此地所顯示的是 52 組 Cosine 相似值大於 0.5 的高關聯類號，每組類號各有一組相同的對應類號組(如 020 與 023 為關聯類號組，則會有另一對應的類號組 023 與 020)。由圖中可看出借閱行為對類號層級結構的影響，很明顯地多數關聯類號組都出現在 45° 的中線附近；換言之，許多讀者的借閱行為模式與分類號架構其實是類似的。但是，也有一些關聯類號並不依循類號結構而分散在中線以外的位置。實驗中針對這些類號組，分析其所包含的書目，發現這些圖書的主題都頗為相關，顯示利用此方法所獲得的結果品質不錯。若將 Cosine 的相似門檻值(Threshold)降低，則可獲取更多的關聯類號組，至於門檻值的高低則依所需關聯類號組的多寡來決定。



圖二 Cosine 相似值大於 0.5 之關聯類號組分佈情形

(三) 類號關聯特性分析

本研究進一步分析了這些高關聯類號，並依其特性歸納出四種類型，討論如下：

1. 層級式類號關聯

第一種類型屬於中線附近的類號關聯，這些關聯其實與 CDC 中的類號架構類似，例如 023 (圖書館管理) 與 025 (專門圖書館) 的關聯在 CDC 中也存在，二者皆屬於 02X (圖書館學) 類別層級 (在本研究中，所謂的同類別層級是指類號的上一層類號也相同者)。這些關聯提供了一個觀察讀者借閱興趣與分類架構之間關係的比較基礎，例如，實驗結果顯示 020 (圖書館學) 與 023 (圖書館行政) 的關聯遠高於 028 (資料中心) 的關聯。此外，許多原屬於同一類別層級的類號，卻未發生關聯，例如 021 (圖書館建築與設備) 或 022 (圖書館行政) 就未與 020 (圖書館學) 產生關聯。

2. 非層級式類號關聯

第二種類型主要包括分散在中線以外地區的類號關聯，這些關聯一來是因為許多跨學域主題所造成無可避免的分散情形外，另一方面也可能是因本地讀者需求不同所造成。舉例來說，表二所列 組 Cosine 相似值大於 0.3 的關聯類號，其彼此主題都頗為相關，但在 CDC 中卻分屬不同的類別層級。

表二 非層級式關聯類號舉例

| 分類號 (類名) | 分類號 (類名) |
|---------------|----------------|
| 019 (讀書法) | 811 (著述翻譯及演說術) |
| 143 (近世哲學) | 549 (社會改革論) |
| 284 (西方神學) | 870 (西洋文學) |
| 367 (生態學) | 445 (市政及衛生工程) |
| 435 (園藝) | 929 (風景建築, 造園) |
| 521 (教育心理及教學) | 176 (心理學各論) |
| 610 (中國通史) | 820 (中國文學) |
| 731 (日本史地) | 803 (東方語言) |
| 876 (法國文學) | 147 (德奧哲學) |
| 992 (旅行觀光) | 427 (飲食烹飪) |

為了解決主題分散的問題，在 DDC 中設計了相關索引 (Relative Index) 的機制，而其作法就是將同一主題的所有可能相關類號群集在同一索引款目之下，以方便查檢。但實際上，這種方式並不易反映變動的知識結構，例如各地圖書館的文化脈絡 (Cultural Contexts) 與讀者興趣不盡相同，使用同一分類體系不一定能完全符合本地讀者的需求與知識結構。由實驗所獲得的 26,423 組關聯類號中，去除一些熱門的類號 (如 857、312 等) 之後，在所有 666 種有借閱記錄的類號中，就有 539 種類號具有

關聯類號，441 種類號擁有至少 10 個以上的關聯類號。從其中我們擷取了 395 組借閱頻率較高的關聯類號進行分析，有 285 組包含了非層級式類號關聯，平均每種類號有 3.11 個非層級式關聯類號。舉例來說，177（應用心理）與 497（廣告）的關聯可能來自於廣告系同學選修了心理學課程，因此經常同時借閱這二類圖書。除了因選修因素而造成的關聯外，177 同時也與 293（命相）產生關聯，這種關聯透露出本地讀者與分類體系中的知識結構其實是有所差異。

上述這些關聯分析顯示出圖書主題的跨學科與多元化特性，同時也反映出本地讀者變動的借閱興趣，而這些效果都不易由相關索引來達成。雖然本實驗所獲得的非層級式關聯類號並不是那麼大量，但若利用人工來檢視 1,000 個類號的關聯類號則將更為困難。事實上，我們還可以利用 Transitive 分析方法來串連更多的關聯類號，例如可將上述所提的 3.11 個非層級式類號作為起始類號（Seed），再串連其同屬層級的其它類號等。

3. 熱門類號關聯

第三種類型包括了一些熱門的借閱類別，如 857（小說）、312（電腦科學）等，這些類號和其它類號產生關聯的原因多半是因為其熱門程度，而不一定與二者的主題有關。換言之，這種類號多少會影響到其它類號的關聯權重，而造成觀察上的不便。但這些類號畢竟也反映出讀者的一般借閱興趣。

4. 錯誤類號關聯

最後一種類型則包含一些人為錯誤所造成的類號關聯，例如在 CDC 中，425 這個類號尚未被使用，但透過上述實驗，卻得到其與 805（日耳曼語系）有關聯。經查証其所含書目，發現 425 應為 424（美容）之誤植。

肆、研究限制與應用討論

一、研究限制與困難

本文介紹如何藉由分析讀者的相似借閱行為，來擷取層級式分類系統中類號的關聯，這種以使用者為導向的方法基本上有一些限制與困難需要注意。首先，圖書館借閱記錄一般而言其稀疏性並不小，對於進行關聯分析其實有相當困難。由於圖書館的情境與一般商業機構不同，前者館藏圖書相較後者的商品種類一般來說更為多元；同時，一旦圖書入館就很難撤架。除此，大學圖書館的讀者會因畢業等因素而無法繼續

借閱該館圖書。基於上述種種因素，圖書館中僅有一小部分的館藏有足夠的流通記錄，而且其平均的借閱率也不高；反之，商品的種類較少，無法售出的產品也會很快撤架，因此平均每種產品都能累積出較大量的交易記錄。

除了上述限制外，大學圖書館的特性也會產生一些問題。例如多數讀者所借閱圖書的類別並不多，也就是讀者的借閱興趣多半集中在某些類別，因此要得到所有類號的關聯類號其實並不容易。同時，其讀者群借閱圖書的動機可能來自於修課，因此並不易釐清類號之間的關聯是否基於主題性質上的關聯。加上以過去的借閱記錄作為基礎，也不一定能反映目前的借閱行為等。若存在上述任何一種問題，即便有大量的借閱記錄，對於關聯分析的幫助其實也不大。

二、應用範疇

本研究結果對於圖書資訊的組織及檢索有相當的應用價值。就圖書資訊組織而言，所獲取的關聯類號可提供多重分類時的參考。此外，由於 CDC 目前尚未有相關索引的編製，這些關聯類號也許可以作為其設計時的參考。若是應用在其它分類系統如 DDC，也可作為更新其相關索引的依據。例如表三所列為 435（園藝）的關聯類號舉例，這些關聯類號皆是以讀者借閱行為的角度所獲取，一方面可提供編目人員作為圖書多重分類的依據；另一方面，也可作為編修相關索引款目的參考。經查驗 六種生命科學領域期刊之影響力及每種期刊內二篇論文之作者數，可由目視顯著地看出作者數與期刊影響係數並無關聯性，證實假設一並不成立，即高影響力之期刊之作者人數並無顯著多於低影響力期刊的作者人數的現象。茲將期刊影響係數和作者人數列表對照如下：

表三 分類號 435（園藝）的關聯類號舉例

| 關聯類號（類名） | 書目之題名舉例 | 書目之標題 | Cosine |
|--------------|------------------|-----------------|----------|
| 436（森林） | 林業哲學與森林美學 | 林業-哲學，原理 | 0.267261 |
| 375（植物學） | 綠美化景觀植物：草花篇 | 植物-台灣 | 0.194225 |
| 374（經濟植物） | 香花植物 | 花卉；園藝 | 0.174816 |
| 427（飲食烹飪） | 歐式花果茶譜 | 茶；飲料；食譜 | 0.16533 |
| 929（風景建築，造園） | 植栽、綠化與景觀 | 造園；都市美化 | 0.149533 |
| 992（觀光旅遊） | 太魯閣國家公園步道解說手冊 | 國家公園-台灣；太魯閣國家公園 | 0.144265 |
| 445（衛生及市政建設） | 城市環境藝術：景觀與設施 | 城市計劃；景觀工程 | 0.134463 |
| 069（博物館） | 歐洲博物館巡禮 | 博物館-歐洲 | 0.095618 |
| 351（地形學） | 台灣的地形景觀 | 地形-台灣 | 0.094987 |
| 367（地質學） | 生機花園：與野生動物共享的花園觀 | 生態學；園藝 | 0.055728 |

而就圖書資訊檢索而言，本研究結果也可作為改善圖書檢索系統介面的參考，例如改善 OPAC 的分類瀏覽功能。舉例來說，當讀者查詢園藝類圖書時，系統不僅可以呈現檢索出的書目資料，同時也可將相關類號提供讀者瀏覽，包括屬於分類表中的層級式關聯類號（如園藝類屬於農業，而農業又包含了農作物、森林等類目），及非層級式的關聯類號（如觀光旅遊、飲食烹飪等類目）。上述分類號的關聯分析研究，其實還有相當多的應用可以嘗試，包括如提昇新書通告服務、及改善 OPAC 檢索結果的相似性排序等。（註 25）

伍、結語

本文嘗試以使用者角度來瞭解層級式圖書分類系統中類號的關聯，透過讀者的相似借閱行為比對，圖書館可以自動擷取分類表中讀者所認知的類號關聯。這些關聯類號的發掘對於圖書資訊的組織與檢索有很大的幫助，包括提供多重分類的依據、充實分類表的相關索引、及改善 OPAC 主題瀏覽介面等。然而，在運用這類方法也有一些限制需要加以留意，特別是有關借閱記錄的稀疏性問題。最後，針對不同性質與規模的圖書館，其讀者借閱行為的特性自然有所不同，對類號關聯的分析也就會有不同影響，亟需更多的實証研究。

註釋

註 1：L.M. Chan, "Classification, Present and Future," *Cataloging and Classification Quarterly*, 21:2 (1995): 5-17; 賴永祥, *中國圖書分類法, 增訂七版* (台北市: 商務, 民國 78 年)。

註 2：A.J. Tinker, et al., "The Dewey Decimal Classification and the Transition from Physical to Electronic Knowledge Organization," *Knowledge Organization*, 26:2 (1999): 80-96.

註 3：T.B. Hickey and D. Vizine-Goetz, "The Role of Classification in CORC," *The 23rd International Online Information Meeting*, edited by B. McKenna. (Oxford: Learned Information Europe, 2000): 247-50.

註 4：A.S. Pollitt, "The Application of Dewey Classification in a View-based Searching OPAC," *Proceedings of the 5th ISKO Conference* (1998): 25-9.

註 5：B.H. Kwasnik, "The Role of Classification Structures in Reflecting and Building Theory," *Advances in Classification Research*, edited by Raya Fidel et al. (Medford,

- NJ.: Learned Information, 1993): 63-81.
- 註 6 : P. Solomon, "User-based Methods for Classification Development," *Advances in Classification Research*, edited by Barbara Kwasnik et al. (Washington, DC, October, 1991): 163-170.
- 註 7 : 卜小蝶, 「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 期 (民 90 年 6 月), 頁 59-72。
- 註 8 : B. Sarwar, et al., "Item-based Collaborative Filtering Recommendation Algorithms," *WWW10 Conference*, Hong Kong (2001).
- 註 9 : G. Salton, *Automatic Text Processing: The Transformation, Analysis, and Retrieval of Information by Computer*. (Reading, MA: Addison-Wesley, 1989).
- 註 10 : 何光國, *圖書資訊組織原理* (台北市: 三民, 民國 79 年)。
- 註 11 : R. Thompson, K. Shafer, and D. Vizine-Goetz, "Evaluating Dewey Concepts as a Knowledge Base for Automatic Subject Assignment," (Dublin, OH.: OCLC, 1997). Available: http://orc.rsch.oclc.org:6109/eval_dc.html.
- 註 12 : J.S. Mitchell and D. Vizine-Goetz, "A Research Agenda for Classification," (Dublin, OH.: OCLC, 2000). Available: http://www.oclc.org/dewey/research/research_agenda.html.
- 註 13 : M. Yee, "Guidelines for OPAC displays prepared for the IFLA task force on guidelines for OPAC displays," (IFLA, 1998). Available: <http://www.ifla.org/VII/s13/guide/opac.htm>.
- 註 14 : L.H. Bertland, "Circulation Analysis as a Tool for Collection Development," *School Library Media Quarterly* 19:2(1991): 90-7.
- 註 15 : J.D. Eldredge, "The Vital Few Meet the Trivial Many: Unexpected Use Patterns in a Monographs Collection," *Bulletin of the Medical Library Association* 86:4 (1998): 496-503.
- 註 16 : A. Barr and H.S. Sichel, "A Bivariate Model to Predict Library Circulation," *Journal of the American Society for Information Science* 42:8 (1991): 546-53.
- 註 17 : K. Banerjee, "Is Data Mining Right for Your Library?," *Computers in Libraries* 18 (1998): 28-31.
- 註 18 : S.J. Cunningham and E. Frank, "Market Basket Analysis of Library Circulation Data," *Proceedings of the 6th International Conference on Neural Information Processing* (Perth, Western Australia, 1999): 825-30.
- 註 19 : A. Paepcke et al., "Beyond Document Similarity: Understanding Value-based

Search and Browsing Technologies,” *SIGMOD Records* 29:1 (2000): 80-92.

註 20 : 同註 9。

註 21 : A. Kent et al. *Use of Library Materials: the University of Pittsburgh Study* (Pittsburgh University, PA., 1979)

註 22 : G. Karypis, “Evaluation of Item-based Top-n Recommendation Algorithms,” *Proceedings of the 2001 ACM CIKM International Conference on Information and Knowledge Management* (Atlanta, Georgia, 2001): 247-54.

註 23 : Hsiao-Tieh Pu, "Discovery of User-Oriented Class Associations for Enriching Library Classification Schemes," *The 75th Annual Meeting of the American Society for Information Science & Technology* (Philadelphia, U.S.A., to appear.

註 24 : G. Salton and C. Buckley, “Term Weighting Approaches in Automatic Text Retrieval,” *Information Processing and Management* 24:5 (1988): 513-23.

註 25 : 同註 7。