

基於多層關聯探勘之新穎圖書 推薦系統

A Novel Book Recommendation System Based on
Multi-level Association Mining

鍾峰宜

Feng-I Chung

國立中正大學資訊管理學系博士生

Ph.D. Student

Department of Information Management

National Chung Cheng University

廖容佐

Rung-Tzuo Liaw

國立清華大學動力機械工程學系研究員

Research Fellow

Department of Power Mechanical Engineering

National Tsing Hua University

莊宗穎

Zong-Ying Jhuang

國立中正大學資訊工程學系研究生

Graduate Student

Department of Computer Science and Information Engineering

National Chung Cheng University

* 通訊作者：丁川康 ckting@pme.nthu.edu.tw

投稿日期：2018年6月21日；接受日期：2018年10月15日

洪新原

Shin-Yuan Hung

國立中正大學資訊管理學系教授

Professor

Department of Information Management
National Chung Cheng University

丁川康*

Chuan-Kang Ting

國立清華大學動力機械工程學系教授

Professor

Department of Power Mechanical Engineering
National Tsing Hua University

【摘要 Abstract】

在當今學術圖書館圖書借閱率逐年下降，圖書推薦系統顯的更加重要—它能有效協助機構促進圖書利用與決策支援。協同過濾為最成功與廣泛應用之推薦系統架構，協同過濾以相同閱讀興趣讀者對圖書的評價資料做為圖書推薦之計算基礎，但學術圖書館在典藏政策、讀者借閱行為、與營運模式均有別於一般商業書店，最為特殊之處在於學術圖書館借閱交易的資料稀疏性與分佈不平衡，這兩個現象嚴重影響圖書推薦預測之準確度與品質。本研究提出以學術圖書館標準分類系統作為基礎，並結合多層次關聯規則探勘演算法，以解決資料稀疏性及分佈不均的問題。此外，本研究設計最佳化折衷方案之圖書推薦策略，作為推薦圖書之挑選依據，再利用隱含性指標資訊之回饋機制調整計算權重，藉以提高圖書推薦之準確度及有效性。實驗證明所提出的方法能有效的降低資料稀疏程度，並且能更準確地發掘讀者感興趣之潛在圖書清單，進而達成有效的圖書推薦。

As the borrowing rate at libraries declines significantly year by year, the book recommendation system becomes increasingly

important in that it can help libraries to promote the borrowing and utilization of books. Collaborative filtering is one of the most successful and widely used technologies for recommendation systems. It leverages the similarity in reading taste between readers as the basis of recommendation. However, this method can't directly apply to book recommendation in academic libraries because their collection policy, reader's behavior, and business model are different from business bookstores. In particular, the distributions of borrowed books at academic libraries are sparse and imbalanced, which seriously affect the accuracy and quality of book recommendation. This paper proposes a new method for book recommendation at academic libraries. The proposed method incorporates multi-level association rule exploration algorithm and taxonomy tree to address the issues of sparsity and imbalance. In addition, we present a best compromise solution for selection of recommended books. The feedback from implicit indicator information is further used to improve the accuracy and effectiveness of recommendation. Experimental results show that the proposed method can handle the sparsity and imbalance in transactions and discover the books of the reader's potential interest, which validates the utility and effectiveness of recommendation.

【關鍵字 Keywords】

協同過濾、資料稀疏性、多層次關聯規則、圖書推薦、學術圖書館
Collaborative Filtering; Sparsity; Multiple-Level Association Rule Mining; Book Recommendation; Academic Library

壹、緒論

近年學術圖書館在個人化服務崛起及經營者重視的趨勢下，其經營策略已從過去豐富圖書資訊提供者的角色，逐漸演進為主動促進讀者社群與圖書館互動服務之媒介。尤其在資訊科技與人工智慧技術的廣泛應用潮流，學術圖書館積極於強化智慧化服務與應用，以邁向智慧化圖書館（Becker et al., 2017），藉此增強學術圖書館在服務讀者的競爭力，進而提升讀者關係管理與對學術圖書館之忠誠度。

根據報導（柯皓仁，2017；高鵬，2017），學術圖書館普遍面臨

實體圖書借閱率逐年下降的困境，在後現代資訊化社會（Information Society）的推波助瀾之下，讀者在圖書資訊使用與閱讀行為有了革命性的改變，使得實體圖書借閱率下滑更為劇烈。為解決這些問題，學術圖書館也開始思考如何運用人工智慧技術來協助決策者進行分析讀者閱讀偏好與模式，進一步主動推銷讀者有潛在興趣的圖書資源，以期追求圖書利用之最大化，並做為圖書規劃經營策略目的（Alharthi, Inkpen, & Szpakowicz, 2017; Becker et al., 2017; Boateng & Quan, 2014; Johnson, Becker, Estrada, & Freeman, 2015; Siguenza-Guzman, Saquicela, Avila-Ordóñez, Vandewalle, & Cattrysse, 2015）。基於以上想法，學術圖書館開始採用推薦系統理論（Recommender Systems）對讀者進行圖書推薦服務。自 Resnick & Varian（1997）提出推薦系統理論以來，推薦系統已被證實在電子商務領域具有顯著成效，例如 Netflix 電影推薦、Pandora 音樂推薦、Amazon 商品推薦、Facebook 社群推薦、Google 新聞文章推薦…等。

推薦系統方法可分為三大類（Adomavicius & Tuzhilin, 2005）：基於內容推薦（Content-based Recommendations）、協同推薦（Collaborative Recommendations）、與混合式方法（Hybrid Approaches）。基於內容推薦方法緣自資訊檢索（Information Retrieval）與資訊過濾（Information Filtering）之研究基礎（Baeza-Yates & Ribeiro-Neto, 1999; Belkin & Croft, 1992; Salton, 1989），此方法主要依據顧客個人過去的瀏覽紀錄與商品喜好認同資料，來進行顧客偏好推薦，由於這類的推薦方法只考慮顧客的過去購買行為，因此無法有效預測與挖掘顧客新的偏好（Huang, Chen, & Chen, 2016）。協同推薦方法則由 Goldberg, Nichols, Oki, and Terry（1992）所提出，此方法將具有相似喜好的顧客定義為特定群體，並定義相鄰群體之間的關係，藉由這樣的定義與運算進行評估和預測未來購買行為，以對顧客進行有效的推薦。

在當今推薦系統研究中，協同過濾方法（Collaborative Filtering, CF）被廣為討論並認為是較有效的推薦方法（Herlocker, Konstan, Borchers, & Riedl, 1999; Papagelis & Plexousakis, 2005; Wang, De Vries, & Reinders, 2006），而在實現協同過濾的方法中，一般可區分為記憶基礎協同過濾（Memory-based CF）、模型基礎協同過濾（Model-based CF）、及混合式協同過濾（Su & Khoshgoftaar, 2009）。記憶基礎協同過濾也稱為啟發式基礎協同過濾（Heuristic-based CF），此方法在使用者 / 產品項目資料庫（User-Item Dataset）中，分析和識別每一位使用者所具有的相似興趣群體標籤，並藉由群體中同儕之間（Peers 或

Neighbors) 的相似度評估，來做為使用者對新產品項目的偏好預測與聚合計算，推薦前 N 個最相似項目清單是其常見的用法。模型基礎協同過濾則是透過蒐集使用者評比歷史數據，設計一個可訓練並能識別複雜特徵的學習模型，透過有效的學習演算法去預測真實世界中的評比數據；這個方法的評估技術經常是基於機率模型產生推薦清單，常用的評估模型包含叢聚模型 (Cluster Models) 及貝氏網路 (Bayesian Networks)。而混和式協同過濾 (Hybrid CF) 通常結合基於內容 (Content-based) 的推薦系統技術，來擷取有用而重要的資訊，再以協同過濾推薦技術評估使用者在過去喜好項目之間的相似程度，進而給予適當的推薦；這種方式往往需要進階的文字探勘技術 (例如：TF-IDF)，以建立本文特徵與項目之間的關聯性。

基於以上演算及其應用特性，在記憶基礎協同過濾方法中，以項目基礎 (Item-to-Item 或稱為 Item-based) 協同過濾推薦法 (Deshpande & Karypis, 2004; Sarwar, Karypis, Konstan, & Riedl, 2001) 最被為廣泛採用在圖書推薦的相關應用，例如：Amazon.com 網路書店從交易數據中，計算最常出現的圖書購買項目組合，來做為推薦讀者購買圖書資訊 (Linden, Jabobi, & Benson, 2001; Linden, Smith, & York, 2003)。然而，我們從協同過濾推薦演算所需的交易資料角度來看，學術圖書館有三點特質迥異於商業書店。首先，典藏政策差異性：商業書店上架政策多以迎合大眾喜好的暢銷圖書為主要考量，因此其圖書交易資料範圍較集中；相反的，學術圖書館的圖書購置政策則是以滿足機構教學與研究為主要考量，特別是具有獨特領域學門用書，其圖書流通資料遠較商業書店疏散。其次，購買 / 借閱行為差異性：以營利事業型態為主的商業書店可透過電子商務營運模式，將評論機制導入業務流程，因此在取得讀者評論資料方面較為豐富；而非營利事業的學術圖書館的流通系統與流程多無此機制，因此幾乎無讀者評論資料可供分析 (Jomsri, 2014)，且大部分圖書都具有固定流通借期，在借出圖書尚未歸還之前，該圖書的流通會一直處於停滯狀態而流通量成長緩慢；也就是說，圖書館買了一本書，大家輪流看一段時間，這有別於商業書店交易模式，某一本熱門圖書可能在一天之內就賣出數千本來自數百個客戶。最後，業務模式差異性：商業書店 (特別是連鎖及電子商業書店)，其讀者群分布廣泛及多樣化，短時間內可快速累積評價與交易量，且讀者具有長期回購資料可供分析；相反的，學術圖書館則是服務特定對象，加上有部分來自特定經費來源下所購置之圖書，可能僅能用於某些特殊用途而長期處於流通停滯狀態；另外，學術圖書

館每年大約有 25% 的讀者會畢業，即停止借閱交易，而在如此高頻率與高比例的讀者流動現象之下，則無法持續累積個人圖書借閱紀錄。因此，對於協同過濾圖書推薦系統而言，圖書流通資料具有特別嚴重稀疏性 (Sparsity) 及不平衡資料集 (Imbalanced Dataset) 分布的問題，而這些問題除了可能會造成推薦結果有偏頗 (Biases) 之疑慮 (卜小蝶, 2002; Anderson, 2004; Koenigstein, Dror, & Koren, 2011)，也會造成圖書推薦準確度低落之現象 (Najafabadi & Mahrin, 2016)。

但是，目前以學術圖書館為主的圖書推薦系統相關研究中，幾乎未處理此稀疏性及不平衡資料集分布問題。

本研究提出一個可以有效降低資料維度，處理圖書借閱流通交易紀錄的新式圖書推薦方法，以解決上述圖書借閱資料之稀疏性與分佈不平衡問題。具體而言，本研究採用能有效處理分類結構的多層次關聯規則探勘演算法 (Han & Fu, 1999)，並提出利用學術圖書館編目分類系統之層級資料結構進行資料降維的策略與方法 (Huang, Chen, & Zeng, 2004; Huang, Ma, Cheng, & Wang, 2015; Najafabadi, Mahrin, Chuprat, & Sarkan, 2017; Zhou & Luo, 2010)，以善用所具有的嚴謹層級結構的特性及優勢；再者，我們採取最佳化折衷方案之圖書推薦策略，作為處理候選圖書之推薦依據，並透過讀者在資訊檢索過程中，經由系統所安排圖書推薦清單的點擊日誌隱含性回饋資料，來調整計算權重，以提高在圖書推薦之準確度及有效性。實驗證明，我們的方法比學術圖書館過去所採用圖書推薦方法，更能準確地發掘讀者潛在興趣之圖書，進而進行有效的圖書推薦。這些成果對於學術圖書館藉由分析圖書借閱流通交易紀錄，進行圖書推薦以提升圖書借用率，具有相當大的助益。

貳、文獻探討

一、協同過濾技術在圖書推薦應用現況

如何有效提升圖書利用 (流通) 率，一直是學術圖書館之重要目標，而利用讀者圖書借閱流通紀錄進行資料分析，除了可以瞭解讀者對於圖書利用的行為，更對圖書館員在圖書典藏與利用推廣策略規劃，提供高價值的參考資訊 (柯皓仁、楊雅雯、吳安琪、戴玉旻, 2002; 謝建成、林湧順, 2006; Blecic et al., 1998; Geyer-Schulz, Neumann, & Thede, 2003; Yan et al., 2010)。Goldberg 等人 (1992) 所提出之協同

過濾推薦技術至今已被圖書館廣泛使用於發掘與分析讀者圖書借閱行為。

協同過濾推薦技術藉由蒐集所有產品項目顧客群的偏好資料，預測新顧客可能喜好的主題或商品項目。舉例而言，假設有 m 個使用者集合 $\{u_1, u_2, u_3, \dots, u_m\}$ ，及 n 個商品項目集合 $\{i_1, i_2, i_3, \dots, i_n\}$ ，我們可以自某個使用者 u_i 對某商品項目的評價資料 I_{ui} ，推論為使用者 u_i 的偏好行為，其中，評價資料可以是明確性指標（Explicit Indications，例如：1-5 Scale 評價分數），也可以是隱含性指標（Implicit Indications，例如：購買或點擊次數）（Miller, Konstan, & Riedl, 2004）。這些讀者評價資料將轉換成讀者 - 項目評價矩陣（User-item Ratings Matrix），進行相關性與相似度計算，在記憶基礎協同過濾技術中採用的相似度計算包括：皮爾森相關度分析（Pearson Correlation）、餘弦相似度計算（Cosine Similarity）、調整式餘弦相似度計算（Adjusted Cosine Similarity）；而在模型基礎協同過濾（Model-based CF）常使用之相似度計算包括：單純貝氏分類計算（Naive Bayes）、閔可夫斯基距離分群計算（Minkowski Distance）、歐幾里得距離回歸計算（Euclidean Distance）、面向模型（Aspect Model）、潛在語意分析計算（Latent Semantic Analysis）（Adomavicius & Tuzhilin, 2005; Su & Khoshgoftaar, 2009）。推薦系統利用相關性與相似度計算找出具有相同興趣的讀者群體，達到資訊協同過濾的目的，進而將這些資訊推薦給類似的讀者。

然而，圖書推薦相關應用範圍廣泛，並非所有應用情境都有豐富的讀者 - 項目評價矩陣（User-item Ratings Matrix）資料可供系統計算，因而會造成圖書推薦準確度表現不佳之現象（Najafabadi & Mahrin, 2016）。討論協同過濾應用於圖書推薦之相關文獻，整理如表 1 所示。近年來研究學者致力於結合機器學習演算方法與運用隱含性指標（Implicit Indications）資料提升圖書推薦模型預測能以及圖書推薦準確度（Bauer & Nanopoulos, 2014）。一些學者研究應用關聯規則來進行偏好樣式（Pattern）探勘，以提供圖書項目相關度或相似度計算之參考基礎（Ali, Khusro, & Ullah, 2016; Parvatikar & Joshi, 2015; Tewari, Kumar, & Barman, 2014; Tewari & Priyanka, 2014; Tsuji et al., 2014）；部分研究引用額外數據資源（例如：語意網路及學習軌跡）作為圖書推薦項目相似度計算之評估準則（Huang et al., 2016; Noia, Ostuni, Tomeo, & Sciascio, 2016; Zhang, 2016）；亦有研究應用圖書主題探勘演算法（例如：LDA 或 LFM）結合內容基礎（Content-based）推薦技術，透過自動分群模式，以混合式圖書推薦方法產生圖書推薦清單（Mariana, Surjandari, Dhini,

表 1
協同過濾應用於圖書推薦相關文獻整理

相關文獻 (References)	相關 / 相似評估方法 (Correlation / Similarity Method)			機器學習評估方法 (Machine Learning Method)				
	皮爾森 / 迴歸 (Pearson / Regression)	餘弦 (Cosine)	距離 (Distance)	機率 (Probability)	關聯規則 (ARM)	分群 (Clustering)	分類 (Classification)	潛在語意分析 (Latent Semantic Analysis) 文本情感 (Opinion) 主題探勘 (LDA / LFM) 語意網 (Semantic Web)
Ali et al. (2016)		*			*		*	
Chen et al. (2015)			*					*
Huang et al. (2016)		*			*			
Jonsri (2014)								*
Krishnamurthy & Balasubramani (2014)					*			
Lu & Zhang (2015)						*		*
Martiana et al. (2017)					*			*
Misra, D., & Ushama, T. (2017)								*
Noia et al. (2016)				*				
Parvatikar & Joshi (2015)					*			
Peru & Ng (2013)								*
Priyanka et al. (2015)							*	*
Sohail et al. (2013)							*	*
Teng et al. (2014)					*			
Tewari et al. (2014)		*			*			*
Tewari & Priyanka (2014)					*			*
Tsuji et al. (2014)					*			
Xin et al. (2013)			*					*
Xin et al. (2014)	*					*		*
Zhang et al. (2016)							*	*
Zhang, F. (2016)			*				*	

註：調查 IEEE, ACM, Elsevier, Springer 等資料庫過去 5 年(2013-2017)協同過濾應用於圖書推薦相關文獻。

Rosyidah, & Prameswari, 2017 ; Pera & Ng, 2013 ; Xin, Haihong, Junde, Meina, & Junjie, 2013) ; 另外, 有一部分研究則採用讀者在文本情感資料 (Opinion, 或稱為書評資料) 進行情緒因素分析, 協助協同過濾在產生圖書推薦清單過程所需之權重計算 (Chen, Chen, & Wang, 2015; Priyanka, Tewari, & Barman, 2015; Sohail, Siddiqui, & Ali, 2013) 。

從過去的研究中瞭解, 關聯規則可以從大量的數據中識別出圖書借閱交易項目組合關聯, 以提供個人化 top-N 圖書推薦 (Sarwar, Karypis, Konstan, & Riedl, 2000) ; 此外關聯規則的導入具有降低資料維度效果及事先預測閱讀偏好樣式的優點, 並能有效提高圖書推薦準確度方面 (Najafabadi et al., 2017) 。

二、圖書交易資料稀疏性 (Sparsity) 問題與處理方法

資料稀疏性是協同過濾技術中存在已久的議題 (Huang et al., 2004) 。圖書借閱交易紀錄為高度稀疏之資料矩陣, 由於稀疏資料矩陣中可評估的評價資料分散, 在計算相關度與相似度時不易找到相似的項目或使用者的, 導致協同過濾推薦成效不彰。

解決資料稀疏性主要分為二種方式: 降低資料維度 (Dimensionality Reduction) 與探索附加資訊 (Additional Information) (Zhou & Luo, 2010) 。降低資料維度的主要概念是以最少的特徵值表現資料, 以降低使用者 - 項目 (user-item) 交易矩陣中無相關對應的資料量, 進而減少矩陣的稀疏性, 常見的方法包括: 主成份分析 (Principle Component Analysis, PCA) 、潛在語義分析 (Latent Semantic Indexing, LSI) 、及奇異值分解 (Singular Value Decomposition) ; 探索附加資訊則是利用探索或是附加額外有用資訊使交易資料矩陣具備描述密集 (Dense) 關聯之有意義資訊, 利於推薦運算時的有用線索, 例如: Sarwar et al. (2001) 從客戶過去的商品交易紀錄與回饋資料進行客戶購買行為分析, 附加資訊於項目基礎推薦方法以識別並推薦類似的商品給客戶。

過去有研究結合分類樹概念與探索附加資訊的方式, 處理資料稀疏性問題。Jomsri (2014) 依據圖書屬性 (Profile) 預先將圖書歸類於某個事先定義的圖書類別標籤 (Catalog) 之下, 並應用關聯規則演算法分析讀者借閱紀錄與圖書類別標籤之間的關聯性, 依據這些關聯規則提出適當的圖書推薦建議。Najafabadi et al. (2017) 利用使用者購買商品資料庫進行分群運算前處理, 接著以關聯規則探勘演算法預測讀者偏好規則 (特徵), 最後以偏好規則特徵為準則, 進行相似度計算及產生推薦清單。Kanagal et al. (2012) 則是建立一個具有分類知識之

潛在因素模型 (Taxonomy-aware Latent Factor Model) 解決商品購買資料稀疏性問題，以有效的推薦具有相同特徵的商品。

目前已有許多研究證明關聯規則探勘可有效實現探索附加資訊方法 (Huang et al., 2015; Huang et al., 2004; Kardan & Ebrahimi, 2013; Tyagi & Bharadwaj, 2013)。而多層次關聯規則探勘演算法對於處理如分類學 (Taxonomies) 之樹狀資料結構成效卓著 (熊馨、王卫平、叶跃祥, 2005; Han & Fu, 1999; Kanagal et al., 2012; Zhang, Chow, & Wu, 2016; Ziegler, Lausen, & Schmidt-Thieme, 2004)。基此，本研究將藉由學術圖書館嚴謹的編目分類系統建立樹狀結構模型，結合多層次關聯規則探勘技術獲取讀者借閱行為特徵，解決圖書流通資料稀疏性問題。

三、圖書推薦策略

典型的協同過濾推薦策略透過相似度計算與加權計算產生推薦清單 (Su & Khoshgoftaar, 2009)。典型的相似度計算方法包括：相關性基礎相似度計算 (Correlation-Based Similarity) 與餘弦向量基礎相似度計算 (Vector Cosine-Based Similarity)。前者可使用皮爾森相關度分析 (Pearson Correlation) 評估二個變數間是否線性相關，而後者則透過餘弦相似計算 (Cosine Similarity) 二個變數在向量空間中其夾角餘弦值，做為相似性評估標準。

上述方法皆須以使用者 - 項目評價矩陣 (User-item Ratings Matrix) 中的評價資料進行運算；然而，目前多數圖書交易機制中並無評價資料，若以借閱次數做為評價分數，由於一本書被同一位讀者重複借閱之比例極低，導致評價矩陣內多為二元值資料 (Binary Data)，則無論採用相關性基礎相似度計算方法 (Correlation-Based Similarity) 或是餘弦向量基礎相似度計算方法 (Vector Cosine-Based Similarity) 皆無法有效量化相似度，故無法準確進行推薦 (Geuens, Coussement, & De Bock, 2018)。為解決此議題，Geuens et al. (2018) 採用傑卡德相似係數 (Jaccard Similarity Coefficient) 作為項目之間相似度的評估標準，其實驗結果顯示，傑卡德 (Jaccard) 相似係數運算可以有效處理因客戶回饋或評價資料不足所造成的推薦失誤現象。傑卡德 (Jaccard) 相似係數計算方式如下：

$$J(A, B) = \frac{|A \cap B|}{|A \cup B|} = \frac{|A \cap B|}{|A| + |B| - |A \cap B|}, \text{ where } 0 \leq J(A, B) \leq 1 \quad (\text{公式 1})$$

在權重數值計算方面，本研究採用 Deshpande & Karypis (2004) 提出的平均互惠命中演算法 (Average Reciprocal Hit-Rank, ARHR) 作為產生推薦清單時的權重計算方法。平均互惠命中演算法 (ARHR) 依據使用者對推薦項目的點擊次數賦予各個項目權重，以提升下次推薦的準確度。平均互惠命中演算法 (ARHR) 計算方式如下：

$$ARHR = \frac{1}{n} \left(\sum_{i=1}^h \frac{1}{p_i} \right) \quad (\text{公式 2})$$

n ：推薦類別下所有圖書總數。

h ：被推薦之圖書所出現在推薦清單的順序值。

p_i ：出現在推薦清單 $top-N$ 中的位置 (i.e., $1 \leq p^i \leq N$)。

基於學術圖書館交易資料的特性，本研究採用傑卡德 (Jaccard) 相似係數計算圖書項目之間的相似度評估，並結合隱含性指標資訊 (Implicit Indications, 例如：點擊) 回饋機制 (Bauer & Nanopoulos, 2014)，依據平均互惠命中演算法 (ARHR) 賦予圖書項目權重數值，以提高圖書推薦系統的準確度。

參、研究方法

本研究提出一基於分類學之圖書推薦系統 (如圖 1)，此推薦系統包含兩個主要特點：一、利用學術圖書館之編目分類系統建構分類學 (Taxonomies) 樹狀資料結構，藉由多層次關聯規則探勘 (Multiple-Level Association Rules Mining) 演算法建構出圖書類別的關聯規則；二、在選定推薦之圖書類別內，本研究提出一最佳化折衷方案 (Best Compromise Solution) 之圖書推薦策略一對存在借閱紀錄之圖書以相似係數作為推薦依據，無借閱紀錄之圖書則採隨機挑選。分類系統的使用可以有效降低資料維度及獲取讀者借閱行為特徵，並同時解決資料矩陣的稀疏性問題；而結合不同策略選出推薦之圖書則可以提升推薦圖書的多樣性 (Diversity)。為進一步提升相似係數的有效性，本研究結合傑卡德 (Jaccard) 相似係數以及平均互惠命中 (ARHR) 法，從讀者隱含性指標資訊 (Implicit Indications) 中，計算出客觀的評估權重，用以調校傑卡德 (Jaccard) 相似係數。各方法的詳細內容，分述如下。

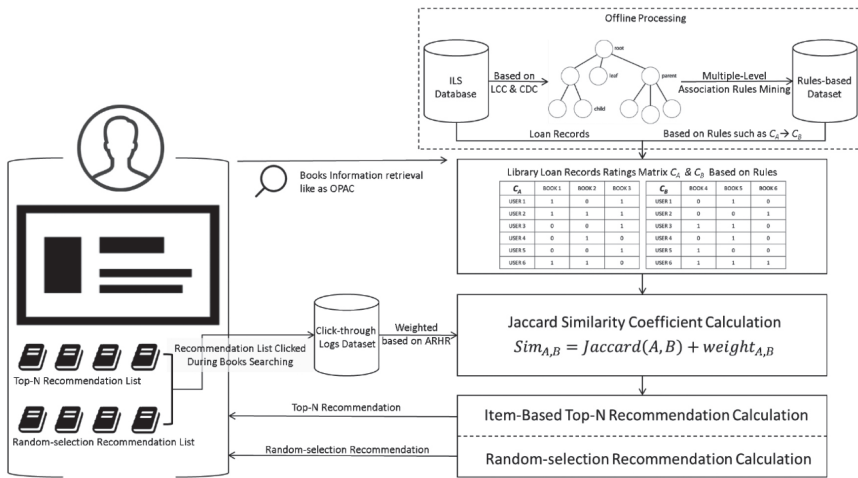


圖 1 基於分類學之圖書推薦系統架構

一、圖書流通資料前置處理

本研究採用的圖書流通資料取自某學術圖書館的圖書自動化系統 (ILS) 資料庫 D ，如表 2 所示，資料欄位包含：登錄號、索書號、圖書名稱、ISBN、經雜湊函數 (Hash) 重新編碼後的讀者代碼、借閱日期、到期日、及歸還日期，並排除非圖書之借閱紀錄如：設施借閱 (例如：討論室)、影音資料借閱 (例如：DVD)、機構典藏論文借閱、及期刊借閱紀錄等。

表 2-1
交易資料欄位 - 代碼對照表

欄位名稱	登錄號	索書號	圖書名稱	ISBN	讀者代碼	借閱日期	到期日	歸還日期
代碼	<i>barcode</i>	<i>callnum</i>	<i>title</i>	<i>isbn</i>	<i>uid</i>	<i>ld</i>	<i>dd</i>	<i>rd</i>

為進行關聯法則探勘，須將資料庫中的資料轉換為交易資料集 (Transactional dataset, TD)，在交易資料集中每筆資料由交易識別碼 (Transaction ID, TID) 及項目 (Items) 所組成，表示如下：

$$\{tid, items\},$$

本研究定義一筆交易資料為某位讀者某一次的借閱紀錄，基本定義 *tid* 由讀者識別碼 (*uid*) 及借閱日期 (*ld*) 組成：

$$tid = uid \& ld,$$

而 *items* 則由本次所借閱的圖書（索書號^註）組成：

$$items = \{callnum_1, callnum_2, callnum_3, \dots, callnum_n\}.$$

註：索書號 (Call number)，也稱分類索書號或稱謂號碼，基本上是由分類號 (碼) 和同類書區分號組成的一組號碼，例如：索書號 521.821_8247 在中國圖書分類法中所代表意義則為教育心理學：教科書分類 (521.8)，以利於讀者區別藏書中的每一冊書，以及作為書籍排列與尋獲之依據。

我們以範例來說明其具體作法，請考慮下表 2-2 所示某讀者 (F12****715) 原始交易資料如下 (原始交易資料格式，如前頁表 2-1 所示)：

表 2-2
讀者原始交易資料範例

編號	原始交易資料
1	C03885995, 885.7157 8733, 牧羊少年奇幻之旅, 957-13-2383-7 (平裝), F12****715, 20170125, 20180131, 20170828
2	C03909037, 873.59 8575-2 v.1, 納尼亞傳奇：魔法師的外甥, 957-455-867-3 (精裝), F12****715, 20170125, 20180131, 20170828
3	C04092435, 873.59 8575-2 v.2, 納尼亞傳奇：獅子、女巫、魔衣櫥, 957-455-890-8 (精裝), F12****715, 20170125, 20180131, 20170828
4	C03937623, 873.59 8575-2 v.3, 納尼亞傳奇：奇幻馬和傳說, 957-455-904-1 (精裝), F12****715, 20170125, 20180131, 20170828
5	C04435649, 873.59 8575-2 v.4, 納尼亞傳奇：賈思潘王子, 957-455-905-X (精裝), F12****715, 20170125, 20180131, 20170828
6	C04435672, 873.59 8575-2 v.7, 納尼亞傳奇：最後的戰役, 957-455-908-4 (精裝), F12****715, 20170125, 20180131, 20170828
7	C04435650, 873.59 8575-2 v.5, 納尼亞傳奇：黎明行者號, 957-455-906-8 (精裝), F12****715, 20170125, 20180131, 20170828

8	C03937634, 873.59 8575-2 v.6, 納尼亞傳奇:銀椅, 957-455-907-6 (精裝), F12****715, 20170125, 20180131, 20170828
9	C05433075, 803.22 8838, 我的第一本圖解韓語單字:輕鬆圖解一看就懂的韓語單字入門書, 9789866077296(平裝), F12****715, 20170607, 20180131, 20170814
10	C05480268, 861.57 8456:7 v.1, 化物語 = バケモノ ガタリ, 978-957-10-4309-8(上冊:平裝), F12****715, 20170712, 20180131, 20170823
11	C05480279, 861.57 8456:7 v.2, 化物語 = バケモノ ガタリ, 978-957-10-4309-8(上冊:平裝), F12****715, 20170712, 20180131, 20170828
12	C05480280, 861.57 8456:7-2 v.1, 偽物語 = ニセモノ ガタリ, 978-957-104-718-8(上冊:平裝), F12****715, 20170712, 20180131, 20170828
13	C05480291, 861.57 8456:7-2 v.2, 偽物語 = ニセモノ ガタリ, 978-957-104-718-8(上冊:平裝), F12****715, 20170712, 20180131, 20170828
14	C05480303, 861.57 8456:7-3, 傷物語 = キズモノ ガタリ, 978-957-104-559-7(平裝), F12****715, 20170712, 20180131, 20170918

為有效進行關聯法則探勘運算，我們依據上述定義，將以上原始交易資料 (Records)，轉換成交易資料集 (Transaction dataset)，演算法 1 為圖書交易資料前置處理程序，將在資料庫 D 中具有相同讀者識別碼及借閱日期之索書號組成一筆交易資料集的項目集合，根據這個範例展示，我們由原來的 14 筆原始交易資料，轉換成 3 筆交易資料集，如表 3 所示。

演算法 1 圖書交易資料前置處理

Input: Database D
 Output: Transactional database T
 For $i \leftarrow 1$ to $|D|$
 $tid \leftarrow ld(D_i) \& uid(D_i)$
 $items_{tid} \leftarrow items_{tid} \cup callnum(D_i)$
 End For

表 3
前置處理編碼後之交易資料集

$T_{F12****715 \& 20170125} = \{885.7157\ 8733, 873.59\ 8575-2\ v.1, 873.59\ 8575-2\ v.2, 873.59\ 8575-2\ v.3, 873.59\ 8575-2\ v.4, 873.59\ 8575-2\ v.5, 873.59\ 8575-2\ v.6, 873.59\ 8575-2\ v.7\}$
$T_{F12****715 \& 20170607} = \{803.22\ 8838\}$
$T_{F12****715 \& 20170712} = \{861.57\ 8456:7\ v.1, 861.57\ 8456:7\ v.2, 861.57\ 8456:7-2\ v.1, 861.57\ 8456:7-2\ v.2, 861.57\ 8456:7-3\}$

二、建立圖書交易資料集分類樹

針對交易資料集 TD ，本研究依據學術圖書館的編目分類系統建構出圖書交易資料集分類樹 (Taxonomy tree based on Transactional Dataset, TTD) 以及基於此分類樹之分類交易資料集 TTD 。圖書交易資料集分類樹的建立分為兩階段，第一階段選出基本候選類別，第二階段則是進行類別擴增與刪減。

演算法 2 為候選類別建立程序，給定交易資料集 TD 、索書號編碼長度 l 、以及閾值 ϵ ，首先選出 TD 中索書號編碼長度小於 l 之類別建立類別集合 C 作為基本候選類別，接著針對 C 中每一類別 c ，若該類別圖書數量大於閾值 ϵ 且存在子類別，則將 c 刪除並將其子類別 s 加入 C 中，最後傳回 C 作為候選類別。在候選類別建立程序中，透過閾值 ϵ 設定以控制分類樹的深度以及各類別間圖書數量之差異。

基於候選類別，演算法 3 將交易資料集 TD 轉換為交易資料集分類樹 TTD ，針對每一筆交易資料 TD_i ，將其中之索書號與候選類別進行最長前綴匹配 (Longest Prefix Match) 作為其在 TTD 中之新類別，將交易資料集中之所有資料重新賦予類別後便可建立如圖 2 之交易資料集分類樹 TTD 。

由演算法 2 與 3 建立交易資料集分類樹後，接著便可利用多層次關聯規則探勘建立圖書類別的關聯規則。

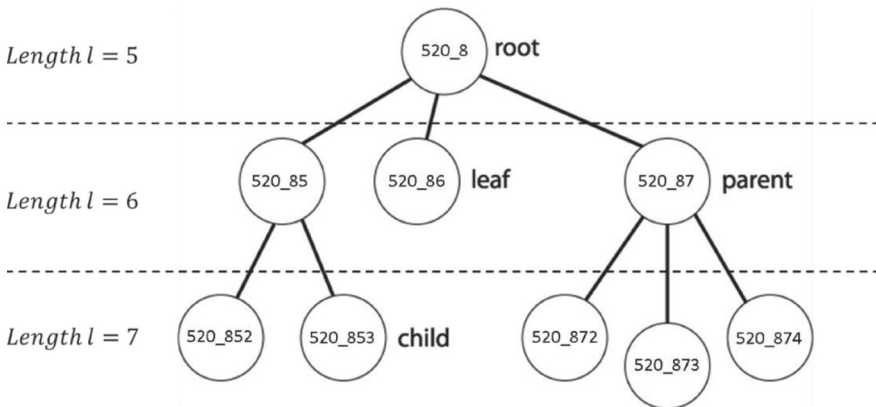


圖 2 交易項目逐碼剖析與項目層級關係圖

演算法 2 候選類別建立程序

```

Input: Transactional database TD, length l, threshold  $\epsilon$ 
Output: Candidate categories C
C  $\leftarrow \emptyset$ 
For each category c  $\in$  TD
    If length(c) < l then
        C  $\leftarrow$  C  $\cup$  {c}
    End If
End For
For each category c  $\in$  C
    If count(c, TD) >  $\epsilon$  or child(c)  $\neq \emptyset$  then
        C  $\leftarrow$  C  $\setminus$  {c}
        S  $\leftarrow$  child(c)
        C  $\leftarrow$  C  $\cup$  S
    End If
End For
    
```

演算法 3 分類樹資料集建立程序

```

For i  $\leftarrow$  1 to |TD|
    TTDi  $\leftarrow$  TDi
    For each c  $\in$  callnum(TTDi)
        c  $\leftarrow$  LongestPrefixMatch(c, C)
    End For
End For
    
```

三、多層次關聯規則探勘 (Multiple-Level Association Rules Mining)

傳統關聯規則探勘將圖書當作項目，由於項目的稀疏性使得關聯規則的發現極其困難。本研究採用多層次關聯規則探勘以解決上述議題。承前所述，在分類資料集 **TTD** 中每個項目為分類而非圖書，如此可大幅解決圖書項目稀疏性的問題。考慮以下圖書借閱交易資料表及關聯規則計算範例：

圖書借閱 交易資料表

T_1	{351.1_8547, 357_8569, 558.092_8774 }
T_2	{337.6_8476, 333.301_845.2, 317.6_8836, 520.11_8635}
T_3	{521.1_8585, 526.8_845, 521.821_8247, 313.3_8584}
T_4	{319.1_8753, 521.13_7665}
T_5	{BF455_G38_1993, BF445_E88_1994, BF317_L83_1986}

Level-1 minsup=4

Level-1 frequent 1-itemsets

Itemset	Support
{3****}	7
{5****}	6
{B****}	3

Level-1 frequent 2-itemsets

Itemset	Support
{3****} {5****}	4

Level-2 minsup=3

Level-2 frequent 1-itemsets Level-2 frequent 2-itemsets

Itemset	Support	Itemset	Support
{31****}	3	{31****} {52****}	3
{33****}	2	{31****} {BF****}	0
{35****}	2	{52****} {BF****}	0
{52****}	5		
{55****}	1		
{BF****}	3		

因此，我們可以推論出規則：
規則一、*Level1*(3**** → 5****)
規則二、*Level2*(31**** → 52****)

於上述範例中，使用傳統關聯規則探勘即使將最小支持度設為 2 依然無法找出任何關聯規則。然而若採用多層次關聯規則探勘，各項目將依其前綴作為分類並計算支持度，在 Level-1、最小支持度為 4 時可以找出關聯規則 3**** → 5****，而在 Level-2、最小支持度為 3 時則可以找出關聯規則 31**** → 52****。

此外，本研究考量各個關聯規則的信心程度 (Confidence) 值及相關度 (Lift) 值，在滿足最小信心度下，依據相關度依序進行推薦。當多於一個關聯規則符合時根據規則的相關度進行選擇，舉例如下：若 168.1* 類圖書同時符合兩項推薦規則：規則 1 (168.1* → 545.1*，相關度 1.85，信心程度 0.78) 與規則 2 (168.1* → 150.1*，相關度 1.08，信心程度 0.52)，則先採用高相關度的規則一，選擇 545.1* 類 (詳表分類：社區環境、都會都市) 的圖書進行推薦。信心程度值與相關度值計算，如公式 3、公式 4 所示。

$$Conf(A \rightarrow B) = \frac{Sup(A \cup B)}{Sup(A)} \quad (\text{公式 3})$$

$$Lift(A \rightarrow B) = \frac{Sup(A \cup B)}{Sup(A) \cdot Sup(B)}$$

If Lift = 1: correlative
If Lift > 1: positive (公式 4)
If Lift < 1: negative

四、產生圖書推薦清單

多層次關聯規則的推論結果會是某個類別的圖書，然而，若該類別內圖書數量大於一次能推薦的圖書數量時則需要進一步地選擇。基此，本研究提出一個最佳化折衷方案 (Best Compromise Solution) 之圖

書推薦策略，此折衷方案分別考量類別內有借閱紀錄及無借閱紀錄之圖書進行推薦。

針對推薦類別內有借閱紀錄之圖書，其推薦順序來自於圖書之相似度。在相似度計算上我們採用傑卡德 (Jaccard) 相似係數 (如圖 3)。為提升傑卡德相似係數之有效性，本研究使用讀者在推薦清單之點擊次數作為隱含性指標資訊對相似度進行回饋，回饋之權重則基於平均互惠命中演算法 (ARHR，如公式 5)。舉例說明：在某分類下 (假設該分類所有圖書為 100 本) 之某圖書出現在 *top-N* (假設 $N=10$) 推薦清單中的第 3 本，當讀者點選該圖書時 (假設該圖書在點擊日誌資料集中出現過 2 次，且同屬於該分類下有 5 筆點擊紀錄)，其權重計算如下：

$$\frac{1}{100} \times \left(\left(\frac{1}{1} + \frac{1}{2} + \frac{1}{3} \right) + \frac{2}{5} \right) = \frac{2.23}{100} = 0.0223$$

$$weight_{i,k} = \frac{1}{n} \left(\left(\sum_{m=1}^h \frac{1}{p_m} \right) + t \right) \tag{公式 5}$$

n ：推薦類別下所有圖書總數。

h ：被推薦之圖書所出現在推薦清單的順序值。

p_m ：出現在推薦清單 *top-N* 中，第 m 順序位置 (i.e., $1 \leq p_m \leq N$)。

t ：該推薦圖書出現在點擊日誌資料集 (Click-through Logs Dataset) 屬於該推薦類別之比例。

C_B	B_1	B_2	B_3	B_4	B_5
u_1	1	1	0	0	1
u_2	1	1	1	0	0
u_3	0	1	0	1	1
u_4	0	0	0	1	1
u_5	1	1	0	1	0

```

for i = 1 to m
  do {
    for k = 1 to n
      if  $B_i \neq B_k$ 
        then  $Sim_{i,k} = Jaccard(B_i, B_k)$ 
      else  $Sim_{i,k} \rightarrow 0$ 
    return ( $Sim_{i,k} + weight_{i,k}$ )
  }
    
```

$$\begin{aligned}
 J(B_1, B_2) &= \frac{3}{4} = 0.75 & J(B_1, B_5) &= \frac{1}{5} = 0.20 & J(B_2, B_5) &= \frac{2}{5} = 0.40 & J(B_4, B_5) &= \frac{2}{4} = 0.50 \\
 J(B_1, B_3) &= \frac{1}{3} = 0.33 & J(B_2, B_3) &= \frac{1}{4} = 0.25 & J(B_3, B_4) &= \frac{0}{4} = 0.00 \\
 J(B_1, B_4) &= \frac{1}{5} = 0.20 & J(B_2, B_4) &= \frac{2}{5} = 0.40 & J(B_3, B_5) &= \frac{0}{4} = 0.00
 \end{aligned}$$

圖 3 傑卡德 (Jaccard) 相似係數計算圖解範例

此外，在推薦類別內無借閱紀錄之圖書，則採隨機挑選（Random Selection）進行推薦。根據圖書自動化系統自 2006 至 2017 之圖書借閱紀錄中發現近六成的圖書無任何借閱紀錄，若只採用相似度作為標準，只能推薦約四成的圖書清單。因此，依據推薦規則，從推薦類別從未被借閱過的圖書中，進行隨機挑選可以有效提升推薦圖書的多樣性。

肆、實驗與評估

本研究將就圖書自動化系統資料庫中讀者的圖書借閱資料進行分析，並以實驗驗證本研究提出之圖書推薦系統與傳統關聯規則探勘相比有效性（Effectiveness）、及效率（Efficiency）上之表現。

一、資料分析

本研究採用的資料來自某國立大學圖書館圖書自動化系統自 2006 至 2017 共 2,966,759 筆讀者真實圖書借閱紀錄，表 4 為實驗資料之分析概況。經由前置處理共擷取出 831,555 筆交易紀錄（Transactions），其中，被借閱、不重複計算的圖書有 375,995 冊，佔全館圖書館藏的 40.97%，換言之，全館圖書館藏有近六成的圖書無任何借閱紀錄。

圖 4 為學術圖書館 2006-2017 原始圖書借閱次數統計圖，由此圖可以發現該學術圖書館自 2013 年起圖書借閱率逐年下降，2017 年之借閱次數相較於 5 年前驟減約 50%，顯示學術圖書館近年確有借閱率迅速下滑之趨勢。

表 4
實驗資料概況（Profile）表

實驗原始資料來源	某國立大學圖書館圖書自動化系統資料庫	
	讀者人口數：約 13,000 人	
實驗原始資料概況	資料擷取區間	2006 至 2017
	圖書借閱紀錄（Raw data）	2,966,759 筆資料
	館藏書目資料	917,731 筆資料
	交易資料集（Transaction dataset）	831,555 筆資料
前處理後實驗資料概況	不重複讀者人數	37,741 人
	不重複圖書冊數	375,995 冊

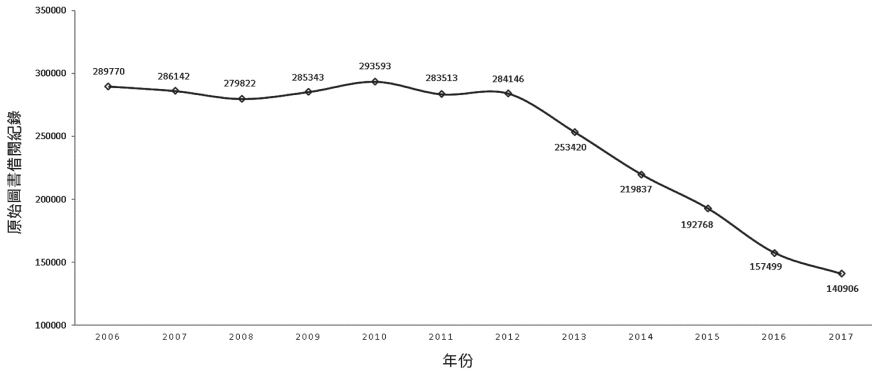


圖 4 學術圖書館 2006-2017 原始圖書借閱次數統計圖

圖 5 為圖書借閱冊數 - 次數統計圖。針對被借閱、不重複計算的 375,995 冊圖書進行分析發現，共有 115,915 冊圖書借閱次數為 1 次，約佔全部圖書的三成；而借閱次數低於 9 次以下的圖書則有 302,785 冊，亦即高達八成的圖書其流通次數不超過 10 次，顯見此圖書借閱資料的稀疏性。

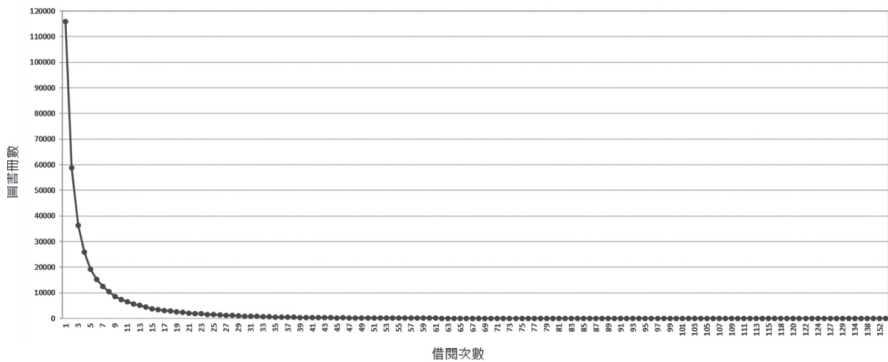


圖 5 圖書借閱冊數 - 次數統計圖

圖 6 為圖書分類借閱次數累加百分比統計圖。從圖書分類學角度切入分析發現，80% 的交易集中在約 30% 的大分類標籤上；相似地，從高頻率圖書借閱紀錄中發現，借閱量前 20% 的圖書佔約 80% 借閱次數，且落在僅約 1% 的大分類標籤上，亦即多數的借閱記錄僅分布於極少數的圖書分類下。由此可以證實，學術圖書館圖書借閱記錄確屬不平衡資料集 (Imbalanced Dataset)。

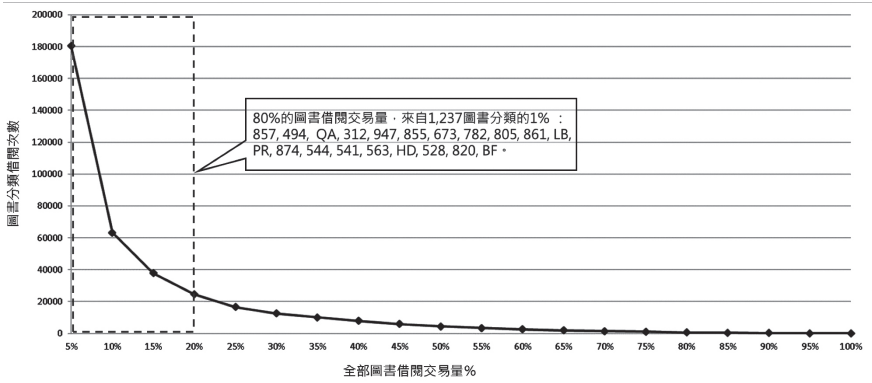


圖 6 圖書分類借閱次數累加百分比統計圖

二、測量指標

過去廣泛被應用於評估協同過濾推薦準確度的指標可分為三類 (Herlocker, Konstan, Terveen, & Riedl, 2004)：評價預測準確度測量 (Predictive Accuracy Metrics)、分類準確度測量 (Classification Accuracy Metrics)、以及排名準確度測量 (Rank Accuracy Metrics)。本研究採用精準度 (Precision)、召回率 (Recall)、及 F1 指標 (F-measure) 做為分類準確度測量測量指標。

精準度 (Precision)、召回率 (Recall)、及 F1 指標 (F-measure) 為二元分類中評估分類結果之常用指標。在二元分類中，依據分類結果可以區分為命中 (True Positive, TP)、正確拒絕 (True Negative, TN)、誤判 (False positive, FP)、以及未命中 (False Negative, FN)。圖書推薦系統亦可視為二元分類問題，即對於某位讀者某本書是否推薦，如表 5 所示，舉例而言，給定任一筆交易資料 TD_i 與圖書推薦之關聯規則 $X \rightarrow Y$ ，則 $X \in TD_i$ 為推薦、 $X \notin TD_i$ 為不推薦、 $Y \in TD_i$ 為有借閱、 $Y \notin TD_i$ 則為無借閱。基於此四種情況，精準度定義為所有推薦的圖書中正確推薦的比例 (如公式 6)，召回率之定義則為所有應推薦的圖書中正確推薦的比例 (如公式 7)，F1 指標則同時考量精準度與召回率 (如公式 8)。對於圖書推薦系統的評估，高精準度代表系統推薦的圖書符合讀者需求，高召回率則代表讀者偏好的圖書，系統都能加以推薦。

表 5
圖書推薦規則準確度測量描述

準確度測量	有借閱	無借閱
推薦	命中 (TP)	誤判 (FP)
不推薦	未命中 (FN)	正確拒絕 (TN)

$$Precision = \frac{TP}{TP + FP} \quad (\text{公式 6})$$

$$Recall = \frac{TP}{TP + FN} \quad (\text{公式 7})$$

$$F - measure = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (\text{公式 8})$$

此外，為驗證圖書推薦系統之效率，本研究以資料矩陣的尺度與稀疏性進行評估，參考 Sarwar 等人（2000）所提出的資料矩陣稀疏率評估方式，資料矩陣尺度之評估方式為讀者總數（ $nUsers$ ）與圖書總數（ $nItems$ ）之乘積，而資料矩陣稀疏率之計算，則是參與計算之交易紀錄總數佔矩陣尺度之比例（如公式 9）。

$$Sparsity\ measure = 1 - \frac{nR}{matrix\ size} \quad (\text{公式 9})$$

where $Matrix\ size = nUsers \times nItems$

nR ：參與計算之交易記錄總數。

$nUsers$ and $nItems$ ：出現在矩陣計算之讀者與圖書總數。

三、實驗設計與結果

(一) 實驗設計

為驗證圖書推薦系統之表現，本研究將 2006 至 2015 的年圖書交易資料集作為訓練資料建立圖書推薦系統，並利用 2016 至 2017 之圖書交易資料集作為測試資料進行驗證。實驗比較簡單關聯規則演算法（Simple Association Rule Mining, SARM）與本研究提出之基於多層次關聯規則之圖書推薦系統間各測量指標上的優劣，本研究並進一步探討多層次關聯規則中層數對推薦系統之影響。

參數設定上，根據初步實驗結果，本研究將最小支持度 minsup 設在 5；在建置圖書交易資料集分類樹上，由於原始資料中 80 次以下的交易佔全部交易量的 99.9%（如圖 5），因此本研究將分支閾值 ϵ 設為 80。建置後圖書交易資料集分類樹，如圖 7 所示，最高分支層數為 12 層。

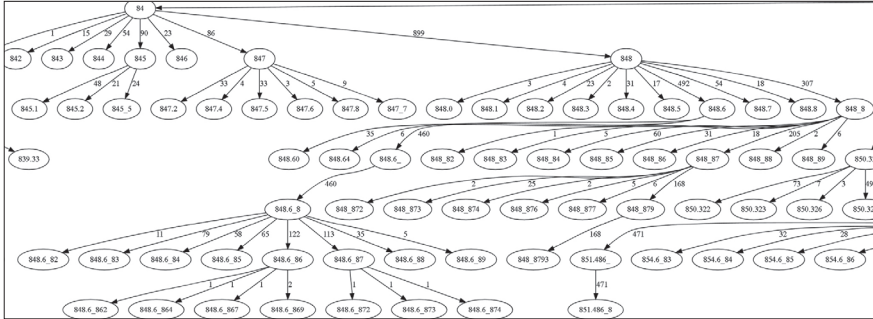


圖 7 圖書交易資料集分類樹 (TTD)，分支門檻值 $\epsilon=80$

(二) 實驗結果

表 6 為圖書推薦模型各項測量指標彙整表。在分類準確度指標上，本研究提出之 **TTD** Level-3 的推薦規則在精準度、召回率、及 F1 指標分別為 96%、59%、73%，均遠高於不分層 (NON-TTD) 推薦規則的 34%、3%、以及 5%。然而隨著層數增加，精準度、召回率、及 F1 指標亦隨之下降，但仍高於不分層之推薦系統。觀察準確度與召回率可以發現本研究提出的推薦系統可以大幅降低誤判 (FP) 的比例。

表 6
圖書推薦模型各項測量指標彙整表

層次 (Level)	精準度 (Precision)	召回率 (Recall)	F1 (F-measure)	稀疏率 (Sparsity-rate)	矩陣尺度 (Matrix-size)
TTD Level-3	0.96	0.59	0.73	0.4278	7.3×10^6
TTD Level-4	0.88	0.49	0.63	0.6179	5.7×10^7
TTD Level-5	0.72	0.38	0.49	0.8324	3.2×10^8
TTD Level-6	0.55	0.20	0.29	0.9581	1.3×10^9
SARM (NON-TTD)	0.34	0.03	0.05	0.9999	1.4×10^{10}

根據公式 6 對精準度的定義得知，推薦規則未能達到 100% 精準，關鍵在於 FP（誤判）的程度，然而，在基於同時考量過度訓練（Overfitting）的因素之下，探討其可能造成錯誤率的原因，我們推估可能的原因在於冷門類別難以產生有效之關聯規則進行圖書推薦所造成。舉例說明，假設我們由訓練資料集所得到的推薦規則為：804 → 805，而其中有 4% 的測試資料集是存在 804 → 806 或 804 → 807，這表示，在資料訓練過程中，有借 804 類圖書並同時 806 或 807 類的圖書的交易資料集並不顯著（低頻率、少數借閱行為），因此，這方面的誤差極有可能是少數冷門類別被借出的交易組合。為避免此遺珠之憾，本研究除了找出 N/2 具有高相關度的推薦圖書之外，並從推薦類別從未被借閱過的圖書中，進行二次隨機挑選，來推薦另外 N/2 本圖書，以提升推薦具有潛在興趣圖書的可能性。

本研究進一步探究關聯規則數對分類準確度指標的影響。圖 8 為各層次關聯規則數曲線圖。實驗結果顯示不分層方法其取得的關聯規則數遠低於 *TTD* 採用不同層次取得之關聯規則數；而 *TTD Level-3* 取得之關聯規則數亦較 *TTD Level-4* 與 *TTD Level-5* 高，但低於 *TTD Level-6*。由此可以發現關聯規則數與分類準確度指標之間存在正相關的關係。

在推薦系統效率上，採用不分層的方法矩陣稀疏率高達 99.99%、矩陣尺度亦高達 10^{10} ，相較之下 *TTD Level-3* 則取得最好的稀疏率與矩陣尺度。對照圖 9 聚合處理後各層次圖書冊數圖可以發現 *TTD Level-3* 透過類別聚合大幅降低矩陣尺度，並將矩陣稀疏率降至 42.78%。然隨著層數上升，*TTD* 的稀疏率與矩陣尺度也逐漸退化，原因在於類別聚合的程度隨層數下降，但表現仍較不分層佳。

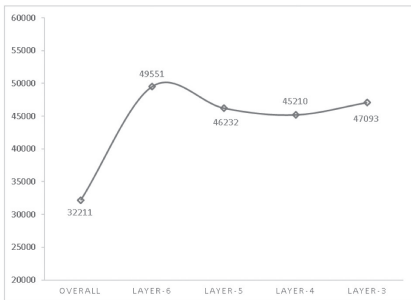


圖 8 各層次關聯規則數 (minsup=5)

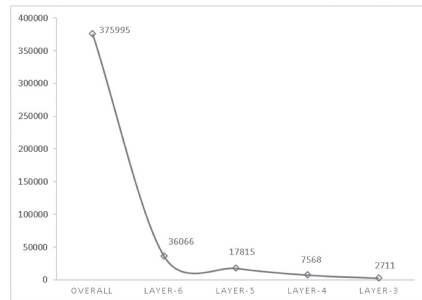


圖 9 聚合處理後各層次圖書冊數

上述實驗結果驗證本研究提出之圖書推薦系統能有效提升分類準確度各項指標包含精準度、召回率及 F1 指標，實驗結果亦顯示此系統能有效降低矩陣尺度與稀疏率，從而提升推薦系統的效率。

伍、討論與結論

傳統採用協同過濾推薦技術應用於圖書推薦之作法上，是以具相同閱讀興趣讀者（Neighbor Users）對圖書之評價資料，來做為圖書推薦計算基礎。但是，學術圖書館在典藏政策、讀者借閱行為、與營運模式，都有別於一般商業書店，造成圖書借閱交易資料矩陣稀疏性及不平衡資料集（Imbalanced Dataset）之現象特別嚴重，對圖書推薦預測之準確度造成負面影響。

基此，本研究提出一個最佳化折衷方案（Best Compromise Solution）之圖書推薦策略來對抗這個處境。透過建構基於圖書目錄學之交易資料集分類樹，解決圖書借閱交易資料矩陣稀疏性與不平衡資料集分布之問題，同時採用多層關聯規則演算法，建立基於讀者借閱行為之圖書類別推薦規則，並在推薦類別下，使用相似度計算及隨機挑選演算法，分別計算出有借閱紀錄與無借閱紀錄之圖書推薦清單，並運用讀者推薦圖書點擊次數結合平均互惠命中法回饋至相似度計算。

實驗結果顯示，本研究提出之圖書推薦系統可以有效提昇分類準確度指標，其中，精準度由 34% 提升至 96%，召回率由 3% 提升至 59%，F1 指標由 5% 提升至 73%。在效率方面，資料矩陣尺度由 10^{10} 降至 10^6 ，資料矩陣稀疏率由 99.99% 降至 42.78%。由實驗證實，本研究所提出的圖書推薦系統可有效應用於學術圖書館之圖書推薦服務。

本研究嘗試提出一個適用於學術圖書館圖書推薦技術創新方法的可能性，有別於傳統的圖書推薦模式，我們的方法除了能有效的克服高頻率圖書借閱項目對低頻率圖書借閱項目的干擾之外，也可以更精準、有效的預測適當的圖書推薦清單。此外亦能應用本研究的圖書推薦系統，將產生的圖書推薦清單圖書資訊檢索平台結合，提供讀者更多樣的圖書資源推薦途徑。未來，我們將持續運用多層關聯規則演算法跨層次之探勘特性，設計隱含性指標資訊回饋數值（例如：資訊檢索歷程日誌、讀者特徵值…等）於產生圖書推薦清單過程中之計算策略，以降低矩陣稀疏率；另一方面，我們也將嘗試結合其他輔助資訊（Auxiliary Information，例如：社交網絡、信任、或第三方書評資訊…等），發掘出更多潛在有趣的圖書推薦項目。

參考文獻

- 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。【Pu, H. T. (2002). An Empirical Study on User-oriented Association Analysis of Library Classification Schemes. *Journal of Library and Information Studies*, 17, 81-94. (in Chinese)】
- 柯皓仁 (2017)。大專校院圖書館, 中華民國一〇五年圖書館年鑑。臺北市: 國家圖書館, 頁 115。【Ke, H. R. (2017). Da jhuan siao yuan tu shu guan. In National Central Library (Eds.), *Jhong hua min guo tu shu guan nian jian: 105 nian*.(pp.115). Taipei: National Central Library. (in Chinese)】
- 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。【Ke, H. R., Yang, Y. W., Wu, A. C., & Tai, Y. M. (2002). Preliminary Study of Personalized and Group Services in Electronic Libraries. *National Central Library Bulletin*, 91(1), 161-195. (in Chinese)】
- 高鵬 (2017)。圖書館統計, 中華民國一〇五年圖書館年鑑。臺北市: 國家圖書館, 頁 354。【Kao, P. (2017). Tu shu guan tong ji. In National Central Library (Eds.), *Jhong hua min guo tu shu guan nian jian: 105 nian*.(pp.354). Taipei: National Central Library. (in Chinese)】
- 熊馨、王卫平、叶跃祥 (2005)。基于概念分层的个性化推荐算法。计算机应用, 25 卷 5 期, 1006-1008。【Xiong, X., Wang, W. P., & Ye, Y. X. (2005). Personal recommendation algorithm based on concept hierarchy. *Computer Applications*, 25(5), 1006-1008. (in Chinese)】
- 謝建成、林湧順 (2006)。書目探勘讀者使用圖書館之行為。教育資料與圖書館學, 44(1), 35-60。【Shieh, J. C., & Lin, Y. S. (2006). Bibliomining User Behaviors in the Library. *Journal of Educational Media & Library Sciences*, 44(1), 35-60. (in Chinese)】
- Adomavicius, G., & Tuzhilin, A. (2005). Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. *IEEE transactions on knowledge and data engineering*, 17(6), 734-749.
- Alharthi, H., Inkpen, D., & Szpakowicz, S. (2017). A survey of book recommender systems. *Journal of Intelligent Information Systems*, 1-22.

- Ali, Z., Khusro, S., & Ullah, I. (2016, May). A hybrid book recommender system based on table of contents (toc) and association rule mining. In *Proceedings of the 10th International Conference on Informatics and Systems* (pp. 68-74). ACM.
- Anderson, C. (2004). The long tail. *Wired magazine*, 12(10), 170-177.
- Baeza-Yates, R., & Ribeiro-Neto, B. (1999). *Modern information retrieval* (Vol. 463). New York: ACM press.
- Bauer, J., & Nanopoulos, A. (2014). Recommender systems based on quantitative implicit customer feedback. *Decision Support Systems*, 68, 77-88.
- Becker, S. A., Cummins, M., Davis, A., Freeman, A., Giesinger, C. H., Ananthanarayanan, V., ... & Wolfson, N. (2017). *NMC horizon report: 2017 library edition*. The New Media Consortium.
- Belkin, N. J., & Croft, W. B. (1992). Information filtering and information retrieval: Two sides of the same coin?. *Communications of the ACM*, 35(12), 29-38.
- Blecic, D. D., Bangalore, N. S., Dorsch, J. L., Henderson, C. L., Koenig, M. H., & Weller, A. C. (1998). Using transaction log analysis to improve OPAC retrieval results. *College & Research Libraries*, 59(1), 39-50.
- Boateng, F., & Quan Liu, Y. (2014). Web 2.0 applications' usage and trends in top US academic libraries. *Library Hi Tech*, 32(1), 120-138.
- Chen, L., Chen, G., & Wang, F. (2015). Recommender systems based on user reviews: the state of the art. *User Modeling and User-Adapted Interaction*, 25(2), 99-154.
- Deshpande, M., & Karypis, G. (2004). Item-based top-n recommendation algorithms. *ACM Transactions on Information Systems (TOIS)*, 22(1), 143-177.
- Geuens, S., Coussement, K., & De Bock, K. W. (2018). A framework for configuring collaborative filtering-based recommendations derived from purchase data. *European Journal of Operational Research*, 265(1), 208-218.
- Geyer-Schulz, A., Neumann, A., & Thede, A. (2003). An architecture for behavior-based library recommender systems. *Information technology and libraries*, 22(4), 165.
- Goldberg, D., Nichols, D., Oki, B. M., & Terry, D. (1992). Using collaborative

- filtering to weave an information tapestry. *Communications of the ACM*, 35(12), 61-70.
- Han, J., & Fu, Y. (1999). Mining multiple-level association rules in large databases. *IEEE Transactions on knowledge and data engineering*, 11(5), 798-805.
- Herlocker, J. L., Konstan, J. A., Borchers, A., & Riedl, J. (1999). An algorithmic framework for performing collaborative filtering. In *Proceedings of the 22nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval* (pp. 230-237).
- Herlocker, J. L., Konstan, J. A., Terveen, L. G., & Riedl, J. T. (2004). Evaluating collaborative filtering recommender systems. *ACM Transactions on Information Systems (TOIS)*, 22(1), 5-53.
- Huang, S., Ma, J., Cheng, P., & Wang, S. (2015). A hybrid multigroup coclustering recommendation framework based on information fusion. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 6(2), 27.
- Huang, T. C. K., Chen, Y. L., & Chen, M. C. (2016). A novel recommendation model with Google similarity. *Decision Support Systems*, 89, 17-27.
- Huang, Z., Chen, H., & Zeng, D. (2004). Applying associative retrieval techniques to alleviate the sparsity problem in collaborative filtering. *ACM Transactions on Information Systems (TOIS)*, 22(1), 116-142.
- Johnson, L., Becker, S. A., Estrada, V., & Freeman, A. (2015). *NMC horizon report: 2015 library edition* (pp. 1-54). The New Media Consortium.
- Jomsri, P. (2014, August). Book recommendation system for digital library based on user profiles by using association rule. In *Innovative Computing Technology (INTECH), 2014 Fourth International Conference on* (pp. 130-134). IEEE.
- Kanagal, B., Ahmed, A., Pandey, S., Josifovski, V., Yuan, J., & Garcia-Pueyo, L. (2012). Supercharging recommender systems using taxonomies for learning user purchase behavior. *Proceedings of the VLDB Endowment*, 5(10), 956-967.
- Kardan, A. A., & Ebrahimi, M. (2013). A novel approach to hybrid recommendation systems based on association rules mining for content recommendation in asynchronous discussion groups. *Information Sciences*, 219, 93-110.
- Koenigstein, N., Dror, G., & Koren, Y. (2011, October). Yahoo! music

- recommendations: modeling music ratings with temporal dynamics and item taxonomy. In *Proceedings of the fifth ACM conference on Recommender systems* (pp. 165-172). ACM.
- Linden, G. D., Jacobi, J. A., & Benson, E. A. (2001). *U.S. Patent No. 6,266,649*. Washington, DC: U.S. Patent and Trademark Office.
- Linden, G., Smith, B., & York, J. (2003). Amazon. com recommendations: Item-to-item collaborative filtering. *IEEE Internet computing*, 7(1), 76-80.
- Mariana, S., Surjandari, I., Dhini, A., Rosyidah, A., & Prameswari, P. (2017, October). Association rule mining for building book recommendation system in online public access catalog. In *Science in Information Technology (ICSITech), 2017 3rd International Conference on* (pp. 246-250). IEEE.
- Miller, B. N., Konstan, J. A., & Riedl, J. (2004). PocketLens: Toward a personal recommender system. *ACM Transactions on Information Systems (TOIS)*, 22(3), 437-476.
- Najafabadi, M. K., & Mahrin, M. N. R. (2016). A systematic literature review on the state of research and practice of collaborative filtering technique and implicit feedback. *Artificial intelligence review*, 45(2), 167-201.
- Najafabadi, M. K., Mahrin, M. N. R., Chuprat, S., & Sarkan, H. M. (2017). Improving the accuracy of collaborative filtering recommendations using clustering and association rules mining on implicit data. *Computers in Human Behavior*, 67, 113-128.
- Noia, T. D., Ostuni, V. C., Tomeo, P., & Sciascio, E. D. (2016). Sprank: Semantic path-based ranking for top-n recommendations using linked open data. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 8(1), 9.
- Papagelis, M., & Plexousakis, D. (2005). Qualitative analysis of user-based and item-based prediction algorithms for recommendation agents. *Engineering Applications of Artificial Intelligence*, 18(7), 781-789.
- Parvatikar, S., & Joshi, B. (2015, December). Online book recommendation system by using collaborative filtering and association mining. In *Computational Intelligence and Computing Research (ICCIC), 2015 IEEE International Conference on* (pp. 1-4). IEEE.
- Pera, M. S., & Ng, Y. K. (2013, October). What to read next?: making

- personalized book recommendations for K-12 users. In *Proceedings of the 7th ACM conference on Recommender systems* (pp. 113-120). ACM.
- Priyanka, K., Tewari, A. S., & Barman, A. G. (2015, April). Personalised book recommendation system based on opinion mining technique. In *Communication Technologies (GCCT), 2015 Global Conference on* (pp. 285-289). IEEE.
- Resnick, P., & Varian, H. R. (1997). Recommender systems. *Communications of the ACM*, 40(3), 56-58.
- Salton, G. (1989). Automatic text processing: The transformation, analysis, and retrieval of. *Reading: Addison-Wesley*.
- Sarwar, B., Karypis, G., Konstan, J., & Riedl, J. (2000, October). Analysis of recommendation algorithms for e-commerce. In *Proceedings of the 2nd ACM conference on Electronic commerce* (pp. 158-167). ACM.
- Sarwar, B., Karypis, G., Konstan, J., & Riedl, J. (2001, April). Item-based collaborative filtering recommendation algorithms. In *Proceedings of the 10th international conference on World Wide Web* (pp. 285-295). ACM.
- Siguenza-Guzman, L., Saquicela, V., Avila-Ordóñez, E., Vandewalle, J., & Cattrysse, D. (2015). Literature review of data mining applications in academic libraries. *The Journal of Academic Librarianship*, 41(4), 499-510.
- Sohail, S. S., Siddiqui, J., & Ali, R. (2013, August). Book recommendation system using opinion mining technique. In *Advances in computing, communications and informatics (icacci), 2013 international conference on* (pp. 1609-1614). IEEE.
- Su, X., & Khoshgoftaar, T. M. (2009). A survey of collaborative filtering techniques. *Advances in artificial intelligence*, 2009, 4.
- Tewari, A. S., & Priyanka, K. (2014, November). Book recommendation system based on collaborative filtering and association rule mining for college students. In *Contemporary Computing and Informatics (IC3I), 2014 International Conference on* (pp. 135-138). IEEE.
- Tewari, A. S., Kumar, A., & Barman, A. G. (2014, February). Book recommendation system based on combine features of content based filtering, collaborative filtering and association rule mining. In *Advance Computing Conference (IACC), 2014 IEEE International* (pp. 500-

- 503). IEEE.
- Tsuji, K., Takizawa, N., Sato, S., Ikeuchi, U., Ikeuchi, A., Yoshikane, F., & Itsumura, H. (2014). Book recommendation based on library loan records and bibliographic information. *Procedia-social and behavioral sciences*, 147, 478-486.
- Tyagi, S., & Bharadwaj, K. K. (2013). Enhancing collaborative filtering recommendations by utilizing multi-objective particle swarm optimization embedded association rule mining. *Swarm and Evolutionary Computation*, 13, 1-12.
- Wang, J., De Vries, A. P., & Reinders, M. J. (2006, August). Unifying user-based and item-based collaborative filtering approaches by similarity fusion. In *Proceedings of the 29th annual international ACM SIGIR conference on Research and development in information retrieval* (pp. 501-508). ACM.
- Xin, L., Haihong, E., Junde, S., Meina, S., & Junjie, T. (2013, December). Collaborative Book Recommendation Based on Readers' Borrowing Records. In *Advanced Cloud and Big Data (CBD), 2013 International Conference on* (pp. 159-163). IEEE.
- Yan, F., Zhang, M., Tang, J., Sun, T., Deng, Z., & Xiao, L. (2010, July). Users' book-loan behaviors analysis and knowledge dependency mining. In *International Conference on Web-Age Information Management* (pp. 206-217). Springer, Berlin, Heidelberg.
- Zhang, F. (2016). A personalized time-sequence-based book recommendation algorithm for digital libraries. *IEEE Access*, 4, 2714-2720.
- Zhang, H., Chow, T. W., & Wu, Q. J. (2016). Organizing books and authors by multilayer SOM. *IEEE transactions on neural networks and learning systems*, 27(12), 2537-2550.
- Zhou, J., & Luo, T. (2010, April). A novel approach to solve the sparsity problem in collaborative filtering. In *Networking, Sensing and Control (ICNSC), 2010 International Conference on* (pp. 165-170). IEEE.
- Ziegler, C. N., Lausen, G., & Schmidt-Thieme, L. (2004, November). Taxonomy-driven computation of product recommendations. In *Proceedings of the thirteenth ACM international conference on Information and knowledge management* (pp. 406-415). ACM.

A Novel Book Recommendation System Based on Multi-level Association Mining

Feng-I Chung

Ph.D. Student
Department of Information Management
National Chung Cheng University

Rung-Tzuo Liaw

Research Fellow
Department of Power Mechanical Engineering
National Tsing Hua University

Zong-Ying Jhuang

Graduate Student
Department of Computer Science and Information Engineering
National Chung Cheng University

Shin-Yuan Hung

Professor
Department of Information Management
National Chung Cheng University

Chuan-Kang Ting*

Professor
Department of Power Mechanical Engineering
National Tsing Hua University

Purpose

In an era of dramatic decreases of library loans year by year, the

book recommendation service is of an even more importance, because it effectively helps institutions' decision making in library usage strategies. Since libraries' collection strategies, patrons' book-loan behaviors and management models are different from bookstores in business sectors, the problem of sparse and imbalanced dataset of library transactions is particularly severe, resulting in negative impacts on accuracy and quality of predictions of book recommendation systems. In this study, the researchers proposed an innovative book recommendation system based on the concept of multi-level association rule mining, for helping resolve the problem of data sparsity and imbalanced dataset, and reveal potential book recommendation lists that interest library users for providing effective book recommendation services.

Method

In this study, the multiple-level association rule mining algorithm was adopted, and a taxonomy tree based transactional dataset was constructed for data dimensionality deduction, for resolving the problem of data sparsity and imbalanced. In this taxonomy tree, features and advantages of rigid hierarchical data structures of cataloging and classification systems used in academic libraries were incorporated. In the other hand, for generating book recommendation lists, the Jaccard's Similarity Coefficient algorithm and weight calculation adjustment based on implicit indication information were adopted for evaluating and generating the best compromised solution for book recommendation systems, aiming at enhancing the accuracy and effectiveness of book recommendations.

Findings

In this study, the library loan records of the library automation system at a national university library in Taiwan from year 2006 to 2017 were studied, with a total of 2,966,759 records, among which 375,995 borrowed and non-double-calculated books were analyzed. It was found that 115,915 were borrowed for just once, making up about 30% of borrowed books. The number of books being borrowed for less than 9 times was 302,785, meaning that 80% of the books had being circulated less than 10 times,

which proved the sparsity of book loan data of academic libraries. Taking it from the perspective of library classification, 80% of transactions focused on 30% of classification categories. From the library loan records of high-frequently borrowed items we could see that the top 20% of frequently borrowed books made up 80% of borrowings, and focused on only 1% of classification categories, that is, the majority of library loan records distributed in a limited number of classification categories. It was evident that the distribution of academic library loan records was of the problem of imbalanced dataset.

Based on the proposed solution in this study, we conducted a study on the 831,555 transactions we pre-processed and extracted, with Precision(P), Recall(R), and F-measure as the measuring indicators of classification accuracy degrees. The study results suggested that compared to the simple association rule mining algorithm, the proposed book recommendation system in this study could raise the Precision from 34% to 96%, the Recall from 3% to 59%, and F-measure from 5% to 73%. As to the efficiency, the scale of data matrix fell from 10^{10} to 10^6 , and the sparsity rate of data matrix fell from 99.99% to 42.78%. From the experiment we could see that the proposed book recommendation system could be effectively applied to book recommendation services in academic libraries.

Significance of the study

In this study we tried to propose an innovative book recommendation technology for academic libraries, which is different from traditional models of book recommendations. Through constructing a taxonomy tree based on transaction datasets of library classifications, we tried to solve the problem of sparsity of data matrix and imbalanced dataset distribution associated with library transactions. We also adopted the algorithm of multiple-level association rule mining to set rules for classification category recommendation based on patrons' book loan behaviors. In addition to overcoming disturbances of high-frequently borrowed items on low-frequently borrowed items, our proposed solution can also more accurately and effectively predict appropriate book recommendations. In addition, the proposed book recommendation system in this study can incorporate the generated book recommendation lists with the library information

retrieval systems, for providing library users with a variety of choices of library resource recommendations. In the future, the study results can be incorporated with other auxiliary information, such as social media, reputation or third-party book review information, for exploring and discovering more interesting library recommendation items.

