

土木建材採購分類關聯分析之探討

陳 鈞 華 一 屏東科技大學土木工程系副教授、水利技師

鄭 富 尚 一 屏東科技大學土木所、碩士班

李 煜 基 一 屏東科技大學土木所、碩士班

蔡 瀚 德 一 屏東科技大學土木所、碩士班

張 武 訓 一 交通部高速鐵路工程局副局長、水利技師

甘 俊 二 一 七星農田水利研究發展基金會董事長、水利技師

摘 要

分類表是材料採購組織資訊的重要工具，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。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：土木建材、資料探勘、關聯規則探勘

一、研究動機與目的

土木構造物所使用的一切材料均可稱為「土木材料」，常用的土木材料有：木材、石材、水泥、塑膠、橡膠、混凝土、金屬材料、瀝青材料等種類甚多。因此重要的加以分門別類可以有效的降低錯誤發生機率及實質材料的需求度，達到供給不缺乏才到品質。在目前分類的機制的體制下圖書館是一個很成熟的管理體制依此本文想藉由這項管理體制運用在土木材料上，依材料用途分類可分為：構造物主體

材料與副材料，然而琳瑯滿目的土木材料要如何有效的以及有系統性的整合。

二、研究材料及方法

2.1 資料收集

初步以瞭解依生產方式分類為主，因此排除其中的依化學組成分類記錄後，共計有機材料及無機材料兩大類別。

2.2 資料分析

一般而言，若要分析相似的土木材料，是需要相當大量的營造廠商有相似的

採購模式才比較可行，同時也需長期收集觀察才較有意義。然而根據一些材料採購使用統計分析研究指出，材料採購中多營造廠商利用材料分類的人並不多。

2.3 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。舉例來說，如果有足夠的樣本，材料之間或分類號之間的關聯就可以透過使用者 (即社群) 的相似採購行為來獲得。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 (卜小蝶，2002)。

2.4 相似性比對方法評估

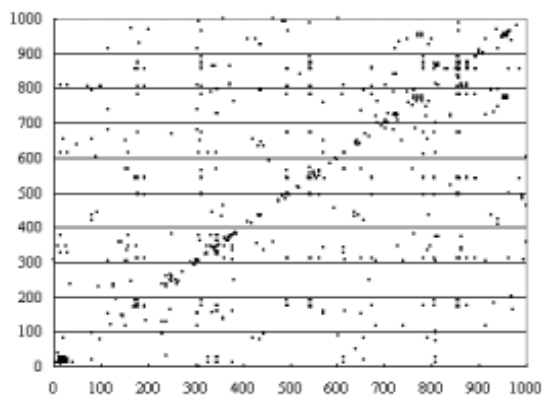
實驗目的之一是評估條件機率與餘弦夾角為基礎等二種不同相似性比對方法，其所獲得類號關聯的品質。首先我們隨機挑選了 10 個採購數量中等的分類號，作為測試組，針對其中每個分類號，再分別由三種來源建立其正確的關聯類號，包括分類表中同屬其分類層級的類號 (hierarchically related)、非分類層級的分類號如參見的分類號 (non-hierarchically related)、及由上述二種方法所獲得經人工

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採購頻率低的類號所影響，這些類號雖然只有少數人借閱、且採購次數也很低，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。

三、結果與討論

3.1 營造廠商認知的類號關聯分佈

為了比較廠商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，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。



資料來源：(卜小蝶，2002)

Cosine 相似值大於 0.5 之關聯類號組分佈情形

3.2類號關聯特性分析

3.2.1層級式類號關聯

透過資料尋找同時發生的事件(event)或記錄(record)並加以分析且以規則的形式來表達搜尋結果從所有事件中決定哪些相關事件應該放在一起。歷史資料中，找出哪些事件總是相伴發生，在大多數的時候我們會稱之為關類別層級（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廠商採購與需求分類架構之間關係的比較基礎聯規則。

3.2.2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突發或臨時性的變更需要所造成無可避免的分散情形外，另一方面也可能是因價格或是特殊別需求不同所造成。

表二 非層級式關聯類號舉例

分類號	(類名)	分類號	(類名)
19	木材	811	鋼筋
143	石材	549	鋼
284	水泥	870	橡膠
367	塑膠	445	人工瀝青材料
435	混凝土	929	強體材料
521	金屬材料	176	管線材料
610	天然瀝青材料	820	合成材
731	磚	803	玻璃
876	瓦	147	再生材料
992	砂漿	427	不鏽鋼

3.2.3必須材料類號關聯

第三種類型包括了一些必須材料的採購類別，如（水泥）、（鋼筋）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。

3.2.4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，

四、研究限制與應用討論

本文介紹如何藉由分析讀者的相似採購行為，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，材料採購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。由於材料採購的情境與一般商業機構不同。

五、結論與建議

本文嘗試以營造廠商角度來瞭解層級式土木建材分類系統中類號的關聯，透過廠商的相似採購行為比對，材料採購可以自動擷取分類表中廠商所認知的類號關聯。這些關聯類號的發掘對於材料需求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，特別是有關採購記錄的稀疏性問題。最後，針對不同性質與規模的材料採購。

參考文獻

- 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）。
- 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.
- 3.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，第 7 期，頁 81~94。

投稿 102.04.12
修改 102.05.10
定稿 102.05.13