

第3屆台灣工業與應用數學會年會

(海報論文：大專生組)

Hodge排序的實作與應用

Yi-Hsun Lin(林奕勳), Chun-Ya Chu(朱君亞) Liang-Hui Huang(黃良惠), Hsing-Yu Lin(林星妤), Min-Hsiang Hsu(許閔翔) Pei-Zhen Lu(呂佩臻)
Department of Mathematics, National Chengchi University, Taiwan (101207420@nccu.edu.tw)
Advisor: Prof. Yen-Lung Tsai (蔡炎龍)

Abstract

排序可以應用在許多方面，例如產品推薦、電影評分等等。但原始資料常有不平衡及不完整的問題，使得這些資料在應用上有些困難。我們採用Qianqian Xu，Qingming Huang，Tingting Jiang，Bowe Yan，Weisi Lin 和Yuan Yao 等人，運用Hodge 理論發展的排序方式。處理傳統上難以排序的問題。我們的研究中，收集三組本身是不平衡、不完整的數據，這些資料在僅有兩組時較易決定高低的情況下，我們運用Hodge排序理論，使用Python程式語言計算，完成完整排序。

Problem description

假設我們有 n 個需要評分的項目(例如 n 部電影)，編號後我們可以設 $V = \{1,2,...,n\}$ 為所有被評分項目，也就是準備排序的集合。我們假設 m 有位評分者，以 $A = \{1,2,...,m\}$ 表之。一位評分者 $\alpha \in A$ 給各評分項目的評分可以以一個函數 $s_\alpha:V \rightarrow R$ 表之。注意一位評分者通常只會評一部的分數，如果某個 $i \in V$ 沒有被這位評分者評到，我們就設 $s_\alpha(i) = 0$ 。

令 $S = \{s_\alpha:V \rightarrow R/\alpha \in A\}$,

則 S 是佈於 R 的 n 維向量空間。一個評分矩陣 $A \in M_{n \times n}(R)$ 就是把兩兩互評的分數記錄下，例如第 i 項勝第 j 項3分，我們有 $a_{ij} = 3$ ，而 $a_{ji} = -3$ 。要從我們的 S 導出一個評分矩陣有很多方式，例如我們可以設 $a_{ij} = \sum_{\alpha} (s_\alpha(i) - s_\alpha(j))/n$ 。

很明顯的，評分矩陣 A 符合 $A^T = -A$ 。我們可令

$$\Phi = \{X \in M_{n \times n}(R) / X^T = -X\}$$

這也是佈於 R 的一個向量空間。

我們的目標就是從實際評分數據中，找出一個函數 $s^*:V \rightarrow R$ ，於是就完成完整排序。要注意在 A 裡面的元素，雖然都可視為一個評分矩陣，但不一定會是平衡且完整的。比方說，我們可能會有 $a_{12} > a_{23} > a_{31}$ 這樣的情況。Xu等人[1]使用Hodge理論，把 Φ 分解成三個向量空間的直和：

$$\Phi = M_G \oplus M_H \oplus M_T^\perp,$$

令

$$M_G = \{X \in M_{n \times n}(R) | x_{ij} = s(i) - s(j), \text{ 其中 } s \text{ 為 } V \text{ 到 } R \text{ 的一個函數}\}$$

正是符合我們要的部份。於是一個矩陣 $A \in \Phi$ 可以唯一分解成 $A = M_I + M_2 + M_3$, 其中的

$M_I \in M_G$ 是由一個真正的排序函數 s^* 導出來的。我們這裡說明 s^* 的算法。

令 $W \in M_{n \times n}(R)$ 為一個矩陣，其中 w_{ij} 是 i 及 j 項目皆有評分的人數總和，而我們令對解線上的 $w_{ii} = 0$ 。再來，我們由 W 導出一個矩陣 $\Delta = (\delta_{ij}) \in M_{n \times n}(R)$,

中間的元素這樣定義：

$$\delta = \begin{cases} \sum_{k=1}^n \omega_{kj}, & \text{當 } i = j \\ -\omega_{kj}, & \text{當 } i \neq j \end{cases} \quad \text{令} \quad Y = \begin{bmatrix} y_1 \\ y_2 \\ \dots \\ y_n \end{bmatrix}.$$

$$\text{其中 } y_i = \sum_{k=1}^n a_{ik} \cdot w_{ik}$$

我們可以很快的求出 $s^*:V \rightarrow R$,且

$$\begin{bmatrix} s^*(1) \\ s^*(2) \\ \dots \\ s^*(n) \end{bmatrix} = -\Delta^+ \cdot Y,$$

其中 Δ^+ 是 Δ 的擬反矩陣(pseudo inverse)。

詳細理論及 Φ 的Hodge分解中其他兩個空間 M_H 及 M_T 的意義,請參考[1,2]。

Results and discussion

這種Hodge 排序的手法，可以讓我們把不平衡、不完整的排序數據，做一個完整的排序。我們實作了以下三個例子。

1. 銀行匯率比較

不論是出國、投資等等都需要換外幣，如何找到最優惠的匯率是大家所關心的。對於相同的貨幣我們能輕易比較出哪一家銀行有相對的優惠，但並不是每一家銀行都有各國貨幣可以換，因此我們能針對有相同貨幣的銀行排名，卻無法直接將所有銀行綜合去排名。

我們選擇20家銀行的4月份平均匯率，將之轉換成1～5的評分，以每一個貨幣所評比分數得出S向量，兩兩相比的分差平均得出A矩陣，被評分次數得出W矩陣，接著使用Hodge排序理論的計算方法得出結論如表(一)。

2. 課程排名

對大學生來說，每學期初選擇課程時，總是讓我們煩惱，五花八門的課程讓我們不知從何選起。如果我們能為課程排名，學生在選課時就能有參考的依據。

首先我們選擇50門政治大學通識課程，以問卷形式請學生針對自己修習過的課程做1～5分的評分，由於不可能每位學生都修習過所有課程，不難想像資料中一定會有遺漏值的存在。於是我們應用Hodge排序理論，以每一位學生所評比分數得出S向量，兩兩相比的分差平均得出A矩陣，被評分次數得出W矩陣，最後得出結論如表(二)。

3. App排名

伴隨著智慧型手機的普及，App的數量與日俱增，總有許多功能相仿的App讓我們難以抉擇，雖然可以參考官方上App的評價，可是方法是取平均數去決定的，我們認為這樣的方法固然方便，卻可能失真，因此我們使用Hodge排序理論去排名。

相同的，我們以問卷形式請使用者針對自己使用過的App做1～5分的評分，以每一位使用者所評比分數得出S向量，兩兩相比的分差平均得出A矩陣，被評分次數得出W矩陣，最後計算出結果如表(三)。

銀行	分數
新光銀行	2.057856
大眾銀行	2.057856
安泰銀行	1.462618
中國信託	1.254234
元大銀行	0.880399
台新銀行	0.526313
匯豐銀行	0.400284
兆豐銀行	0.400284
臺北富邦銀行	0.265613
上海銀行	-0.095912

課程	分數
通往宇宙的神奇路	1.259885
動物與人類社會	0.669220
野生動物保育行銷	0.459516
醫學的美麗魔法棒	0.455906
電影與國際關係	0.413420
同志生命美學	0.391713
探索遺傳特性	0.359771
心理與生活	0.330076
生活中的經濟學	0.290509
藝術欣賞與創作	0.281950

APP	分數
台北等公車	0.548011
Google Maps	0.535863
Chrome	0.526279
LINE	0.486914
Instagram	0.268137
Facebook	0.232211
Google Drive	0.222110
Adobe Acrobat DC	0.213204
Mo PTT	0.128730
英漢字典	0.128719

Conclusions

本研究嘗試利用兩兩評價中的比較結果轉換成全體排名，藉此把不能做直接評分的資料去做出排序，並且修正個人對於給分的高低偏差，進而得到一個較有參考價值的總體排序。然而，我們的調查對象主要是以學生為主，在抽取樣本上可能會有所偏差，未來研究將收集更廣泛的樣本，以求更真實的排名順序。

References

- [1] Xiaoye Jiang, Lek-Heng Lim, Yuan Yao, and Yinyu Ye. Statistical ranking and combinatorial Hodgetheory. Math. Program., 127(1, Ser. B):203–244, 2011.
[2] Robert Kelly Sizemore. Hodgerank: Applying combinatorial hodge theory to sports ranking. Master’sthesis, Wake Forest University, 2013.