

社群網路資料分析及活動賓客推薦

專題編號： NTPUCSIE-103-002

組員： 李力學 劉炳佑

摘要

近年來社群網路蓬勃發展拉近了人與人之間的距離，即使處於不同國家，不同時間區，人與人仍可互相交流自己的生活經驗，有時也使得人們可以聯繫上許多失聯已久的朋友，成為人們生活中一個重要的連結工具。而除了心情連結、拓展生活圈的功能外，基於通訊上的便利性與資料變更的即時性，也經常取代傳統的聯絡方式，被使用來籌劃和舉辦各種活動。舉辦活動是個複雜繁瑣的工程，除了必須設定時間、地點、內容外，我們特別關注「要邀請哪些賓客」這項問題。對於一個活動而言，有什麼樣特質的賓客才適合參加呢？是和主辦人關係密切嗎？還是個人特質與活動主旨相呼應呢？我們觀察過去的活動得知，同一個主辦人舉辦不同活動的時候邀請的成員也各不相同，過往的推薦中，學者多將推薦方法運用在書籍、電影等處的推薦上，還沒有人考慮對於一任意活動主題，是否可以從分析活動的特性在朋友群中推薦出適合這個活動的賓客。在本次專題中，我們以資料探勘的角度探討這個問題，利用矩陣的方式建立活動與成員間的關聯性，並對矩陣做一連串的運算分析，找出活動的特徵字詞，並以分數計算的方法，計算每個朋友的分數，選出適合參加的賓客，達到我們推薦的目的。

資料蒐集：關於、說讚的內容



方法描述：密度計算與區域切塊

校園/公司類		台北大學	五月天	游泳	擊劍	成功高中
參與人	主辦人					
小明		1.00	1.00	1.00	0.75	0.80
朱朱		1.00	1.00	0.83	0.62	0.60
達達		1.00	1.00	0.78	0.58	0.53
阿德		1.00	0.88	0.67	0.50	0.45
曉華		1.00	0.80	0.67	0.55	0.48
阿花		0.83	0.75	0.67	0.41	0.50

從(1,1)開始 計算有1的密度 並利用三角差法尋找區域切塊斷點

切好區塊後 區塊內包含的字詞就是代表這個活動的重要字詞

推薦階段：分數計算推薦賓客

方法描述：矩陣建立

校園/公司類		學校/公司類詞		興趣類詞		
參與人	主辦人	成功高中	台北大學	五月天	游泳	擊劍
小明		1	0	1	1	0
阿德		0	1	0	0	0
曉華		0	1	1	0	1
朱朱		1	1	0	0	0
達達		1	1	0	1	0
阿花		0	0	1	0	1

若有則為1 沒有就是0

參與人詞

北大金曲比賽		台北大學	建國中學	五月天	跳舞	唱歌	總分
參與人	主辦人						
妮妮		1	0	1	1	1	4
阿呆		1	1	0	0	0	2
小熊		1	0	0	1	0	2
維尼		0	1	1	0	1	3
兔兔		1	0	1	0	1	3
小輝		0	0	1	1	1	3
多尼		1	0	0	0	0	1

實驗結果：準確度評估

方法描述：行列交換

校園/公司類		台北大學	五月天	游泳	擊劍	成功高中
參與人	主辦人					
小明		1	1	1	0	1
阿德		1	0	0	0	0
曉華		1	0	1	1	0
朱朱		1	1	0	0	0
達達		1	1	0	0	0
阿花		0	1	1	1	0

Column計算總數量 由大到小交換

校園/公司類		台北大學	五月天	游泳	擊劍	成功高中
參與人	主辦人					
小明		1	1	1	0	1
朱朱		1	1	0	0	0
達達		1	1	0	0	0
阿德		1	0	0	0	0
曉華		1	0	1	1	0
阿花		0	1	1	1	0

row交換 1的往上 0的往下

