

結合社交圈探索與潛在新興趣發現的社會書籤推薦系統設計

廖崇勛

國立高雄第一科技大學資訊管理學系

u9924808@nkfust.edu.tw

黃承龍

國立高雄第一科技大學資訊管理學系

clhuang@nkfust.edu.tw

摘要

過去不少研究著重在於找尋與目標使用者相似的他位使用者做協同推薦，但以相似度為基礎的推薦項目可能與目標使用者原先的收藏同質性較高，本研究以 Diigo 書籤收藏社群網站為實驗資料來源，並以貝氏定理加以改良，同時考量了時間因素與社交網路分析的概念，如此就能從目標使用者較常互動好友延伸出去的人際網路收藏中，依照好友交集程度以及近期興趣偏好，找出目標使用者感興趣的項目。結果顯示，加入社交網路分析的應用，確實可以補強相似度較低的情況下，也能讓使用者得到在社交網路中更多元的收藏項目之推薦。

關鍵詞: 知識分享、天真貝式分類器、推薦系統、社交網路、協同過濾

壹、研究動機與目的

人與人之間的朋友關係也不僅只侷限於現實生活中，從以前的聊天室、交友網站，到現在的Facebook、Plurk、Twitter...等等各種社群網站的出現也讓交友無所限制，我們可以有更多元的方式來認識新朋友。在社群網站中，朋友可以是彼此有共同的興趣，或是同樣的工作環境、同樣的學校，只要彼此同意，任何微小的關聯都可以是網路上彼此成為朋友的理由。也因此在此社交網路中使用者的好友數動則數百或千個好友連結是個很正常的現象，因為在虛擬的網路中建立好友連結比起在現實生活中建立好友關係來的容易。另外研究也指出人們好奇朋友感興趣的項目勝過於推薦系統所推薦的項目(Rashmi et al., 2001)。但又如何從動則數百或千個好友找出真正好友的收藏來推薦給目標使用者呢？

過去研究大多著重在於找尋與目標使用者相似的他位使用者做協同推薦，但這樣方式找出來的推薦項目與目標使用者原先的收藏同質性較高，所以本篇研究著重在於如何在目標使用者數量龐大且複雜的人際網路中不再以相似度為唯一考慮因素，而是從多方面的社交網路分析在動則數百或千個好友連結中去找到與目標使用者高關聯(強度)的朋友，如此的推薦項目不再只侷限在於相似使用者身上，而是從人際網路中可能是非相似的使用者找到目標使用者感興趣的收藏項目，這樣的推薦方式更能讓目標使用者接收到人際網路中更多樣化的知識分享。

貳、文獻探討

一、 社交網路(Social Network)

社交網路的應用在最近幾年越來越受歡迎，像是 Twitter 和 Facebook 就是一種提供使用者交流與分享資訊的社交網路平台。也因為社交網路在近幾年大受歡迎，因此也有越來越多的研究議題被提出來。

在針對大學生所做的調查發現，大學生使用社交網站的主要動機分為兩類：社會連結及資訊分享。社會連結指的是使用者能透過社交網站可以和現在的朋友或以前的朋友做聯繫，資訊分享則是分享身邊的社交圈訊息及其他的訊息事件(Raacke & Bonds-Raacke, 2008)。

(一) 社交網路對資訊分享的影響

資訊分享所代表的是一種溝通的過程，網路的低成本普及化以及社群網站的出現，都改變了資訊分享溝通的方式。學者認為社群網站是一群擁有共同興趣以及特定目標的人，在網路平台中進行知識的交流與資訊的傳達(Koh & Kim, 2004)。因此可以發現人們獲取資訊的方式已經因為這些改變而與過往不同了，例如過去利用搜尋引擎來索取訊息，現今則可能藉由社群網路資訊的傳播，而從好友網路中獲得所需要的資訊。

(二) 社交網路分析

現今社交網站上，平台的開發者致力於開創更多的溝通管道(Kolari et al., 2007)，像是評論(Commentate)、引用(Cite)、社交化(Socialize)...等，讓平台使用者有更多的機會接觸到原本社交網路之外的其它使用者，以形成新的連結網路。也因此社交網路中的使用者好友數動則數百或千個好友連結其實是個很正常的現象，原因在於虛擬的網路中建立好友連結比起在現實生活中建立好友關係來的容易。也因如此，在這樣的社交網路連結中其實存在著許多假連結(False Links)，如何在複雜的人際網路連結中辨別出真正朋友的結連(Real-friend Links)也就是社交網路分析(Social Network Analysis, SNA)的概念。

(三) 社交網路分析的運用

社交網路分析用來嘗試去瞭解這些行為者彼此間的人際關係、特徵並探討這些關係對他人或是組織的影響(Garton et al., 1997)，所謂行為者可以是人(People)、組織(Organizations)或是群組(Groups)，依照研究領域而定 (D. Shen et al., 2006)。社交會網路分析意涵是指運用相關的社交網路分析法指標工具如網路大小、網路密度、網路方向性、網路強度等評估方式來描繪出虛擬人際網路之間，如友誼網路、信任網路、學習網路的網路狀況，或是了解社交網路中的角色是位於核心或是邊陲，來幫助解決社會或是個人所面臨的問題(邱議德, 2003)。

社交網路分析的應用已經相當的成熟也在許多的研究中被提出來討論，例如 社交網路分析被(Schwartz & Wood, 1993)用來分析出現在社交網路之中所形成的群組關係。(Domingos and Richardson, 2001)使用社交網路分析的概念，來計算評估每個獨立的節點在網路之中的重要程度。以及(Adamic & Adar, 2002)從部落格的使用者個人首頁資訊中，用來發現使用者彼此之間的社交互動(Social Unteraction)。從以上的應用中可以發現，社交網路分析可以做為在人際網路中分析出彼此互動性的量化指標，藉此可以找出可能有潛在互動關係的節點或是尚未被發現的潛在好友。而無論何種社交網路皆至少包含的三個基本組成要素:節點(Nodes)即代表網路中的行為者(Actors)、行為者與行為者之間的關係(Relationship)、行為者之間的連接途徑-聯繫(Tie)的強度。

二、 社會性標記(Social Tagging)

社會性標記又稱為協同式標記(Collaborative Tagging)、大眾分類法(Folksonomy)等...，標記指的是利用一個非結構化關鍵字的標籤來做定義，該關鍵字可用來註解網頁文件或是做為圖片的標記，使資源產生具有代表性的意義，目的是為了讓使用者和搜尋引擎更容易找到相關的資訊(O'Reilly, 2005)。當書籤由使用者個人作為索引，進而集結成共同索引的機制時便是稱為社會性標記(Social Tagging)(卜小蝶，2006)。

(一) 社會性標記與其它分類方式差異

這與傳統的分類定義不同的是過去分類方法是由少數該領域的專家相互討論來做決定，社會性標記分類是一種由下而上的大眾分類方式，也就是當某個項目被大多數的人由某特定標籤所標記後，即產生了代表這個項目的新定

義，來取代以往由專家所制定的定義。因此使用者在制定標籤時並不需具備相關的專業知識，而是以自己的喜好來做標記即可，所以可能會因為使用者的專業背景不同、思想上的差異而有不同的標記名稱，這樣的行為可以視為是以群眾智慧(Collective Intelligence)來幫助分類的方式，因為其結果即代表大眾對該項目的共同看法。因此社會性標記並不算是分類(Classification)，而其實是一種歸類(Categorization)的行為(Mathes, 2004)。以下為社會性標籤的研究以及社會性標籤網站的現象探討：

(二) 社會性標籤網站的現象探討

為了比較社會性標記搜尋和搜尋引擎及傳統目標分類法在資訊擷取效能上的差異，將三種搜尋方式運用於社交書籤網站後發現，同時被社會性標記和搜尋引擎都找到的文章，比起單獨只使用搜尋引擎找出的結果較貼近查詢主題，因此社會性標記在社交書籤網站上的應用是有效的，確實可以協助改善搜尋結果的準確度(Jason Morrison, 2008)。另外有學者分析在社交書籤分享網站上使用者的標記行為、標籤使用頻率、標籤種類和數量變化情形，結果顯示標籤數量隨者使用者的時間而呈正向關係，並且每個標籤的使用成長程度不同，因此根據標籤使用頻率可以視為使用者的興趣變化情形(Golder & Huberman, 2006)。另一結果顯示當項書籤被標籤標記至一定數量時，該書籤被各標籤所標記到的比例將呈現穩定狀態，這也表示書籤累積至一定的被標記數量後才能發揮社會性標記所帶來的效用。

三、 社交書籤分享網站 (Social Bookmark Sharing Website)

社交書籤分享網站是社會性標記的應用之一。書籤指的是 URL，在過去 web2.0 尚未蓬勃發展時期，使用者對於欲收藏的網頁資料以存放在瀏覽器的「我的最愛」中為主，而這樣的瀏覽器書籤存放功能不能讓使用者在不同電腦存取書籤，無法與他人共享自己喜愛的書籤，因此有了社會性書籤服務的出現。

書籤網站最主要由資源使用者(User)、標籤(Tag)、(Resource)三部分所組成(Marlow et al., 2006)，提供了使用者組織、儲存、管理與搜尋網路資源的功能，讓現今書籤分享網站的使用者能夠將書籤根據個人的喜好、興趣，讓使用者利用個人化的標籤(Tag)來做分類管理。藉由標籤來標記所收藏的書籤所帶來的好處是，在書籤被標籤之後，能夠讓使用者搜尋網路資訊時，也可同時發掘更多已經被分類好具有相似標籤的其它網路資源，以提供使用者更多的參考(Yanbe et al., 2007)。

(一) Diigo-社交書籤分享網站

Diigo 是國外知名的社交書籤分享網站之一也是本研究實驗資料來源。其以擴展個人記憶，分享群眾知識為理念，改變了將過去將書籤存放在個人瀏覽器我的最愛中的模式，將書籤由使用者使用標籤定義描述後存放於網路平台上與他人共享。此平台主要提供了三種功能：一、儲存(Store)，儲存網頁連結；二、分享(Share)，平台提供讓網路使用者可以針對彼此書籤追蹤、分享、討論的功能；三、發現(Discover)，藉由分享網路使用者的書籤之後進而發現自己所需的新資訊。因此從使用者的書籤收藏行為、標籤使用頻率以及所追隨的好友名單

中已經可以觀察出該位使用者的興趣偏好，也已經有研究針對書籤分享網站加入時間因素的應用，以使用者近期偏好為考量，最後實驗結果顯示考慮時間因素的推薦系統確實擁有更準確的書籤推薦效果(陳威志, 2011)。

四、 推薦系統(Recommender System)

在網際網路蓬勃發展的同時，也意謂著網路上擁有著無數個開放的資源，顯示了網際網路的另一項被討論的議題「資訊超載(Information Overload)」。人們藉由搜尋引擎的輔助來搜尋網路上所需的資源，而推薦系統相對於搜尋引擎來說是一種主動式提供使用者所需資訊的一種概念。推薦系統可以依據使用者興趣偏好、行為或其他需求，推薦使用者所需的潛在資訊、服務或產品(Rashid et al., 2002)。推薦系統的運算過程可以被分三個主要階段:

- (一) 擷取使用者的個人資訊。如興趣、商品購買明細、書籤收藏偏好等。
- (二) 依收集到的使用者資訊進行分析並產生合適推薦項目，如相似度計算。
- (三) 根據使用者對推薦資訊的滿意度，再適當的修改推薦系統。

擷取使用者個人資訊方式中可以被區分為兩種方式，一、主動式:由使用者主動提供個人相關資訊、或由系統要求使用者直接提供個人偏好資訊給予系統做運算，如(O'Donovan & Smyth, 2005)的研究中要求部落格使用者針對其好友做信任度(Trust)的評分，並讓推薦系統從較高信任的好友收藏中找推薦給使用者，主動式的資料收集方式雖然能直接從使用者得到準確的分析資訊，但要求使用者提供資訊的方式較容易造成反感。二、被動式:從使用者的行為中收集所需要的分析資訊，例如從使用者的歷史購買行為來推薦未來可能購買的項目。而在推薦系統的推薦方式中可以分為以下三種方式:內容過濾式推薦系統、協同過濾式推薦系統、混合式推薦。

參、系統架構

本研究目的為:找尋相似度以外的多元推薦項目，因此除了相似度計算以外本研究加入多方面的指標去評估每位使用者與目標使用者的關聯程度，並從目標使用者延伸出去的人際網路中找尋與目標使用者擁有較高關聯度的他位使用者收藏項目，再依照目標使用者的近期興趣與這些收藏項目做貝氏推薦計算。本研究可分為三階段:一、目標使用者的社交網路強度分析階段，二、考量時間因素找出目標使用者的近期偏好標籤階段，三、貝氏計算推薦階段。

一、 社交網路強度分析階段-計算目標使用者與各使用者之間的關聯分數

以往推薦系統大多以標籤或項目做相似度計算找尋相似使用者，而在現有社群平台 Facebook 或 Twitter 中共同好友數也成為了茫茫社交網路中推薦好友的依據之一，因此本研究也將此一概念加入，藉此找出可能與目標使用者不相似但可能是目標使用者感興趣的人或收藏項目，熱門程度的考量則是評估該位使用者的追隨價

值，當熱門程度高時極有能代表著該位使用者是名人或是這位使用者的收藏項目能引起眾人的興趣才吸引較多使用者追隨，各指標公式如下(1)到(3)所示。

$$\text{以標籤為基礎的 Cosine 相似度: } Sim(u, v) = \frac{\sum_{t \in T} (A_{u,t} A_{v,t})}{\sqrt{\sum_{t \in T} A_{u,t}^2} \sqrt{\sum_{t \in T} A_{v,t}^2}} \quad (1)$$

$A_{u,t}$ ：使用者 u 對 Tag_t 的使用頻率； $A_{v,t}$ ：使用者 v 對 Tag_t 的使用頻率

$$\text{共同朋友數的衡量方式: CommonFriend} = \frac{\text{共同好友數}}{\text{目標使用者好友數}} \quad (2)$$

$$\text{熱門程度衡量方式: Popularity} = \text{該使用者的追隨人數(Follower)} \quad (3)$$

公式(1)-(3)各指標數值來自於不同基準的資料，為了將來自各方面的值做權重的計算，避免某一特徵值過大或過小而影響了整體資料的準確性，故進行公式(4)的正規化計算使各屬性值介於[0,1]之間，計算方式為將各指標減去該指標中的最小值除上指標中最大減最小值，最後則將正規化後的指標依公式(5)給予不同權重計算出最後總分，經由排序之後可整理出與目標使用者關聯強度較高的人際網路。

$$X'(i) = \frac{X(i) - \text{Min}(X(i))}{\text{Max}(X(i)) - \text{Min}(X(i))} \quad (4)$$

$\text{Max}(X(i))$: 此屬性中的最大值； $\text{Min}(X(i))$: 此屬性中的最小值； $X'(i)$: 正規化後的屬性值

$$\begin{aligned} \text{關聯分數} = & \left(\text{權重} \times Sim(u, v) \right) + \left(\text{權重} \times CommonFriend \right) \\ & + \left(\text{權重} \times MutualFollow \right) + \left(\text{權重} \times Popularity \right) \end{aligned} \quad (5)$$

網路強度關聯分數計算方式可分為 Step1~3 進行: Step1 執行公式(1)~(3)評估各使用者與目標使用者的網路強度分析、Step2 再將 Step1 所得到的值進行公式(4)正規化計算、Step3 最後再將正規化數值利用公式(5)進行總分計算，以下將以實例展示之。

(一) 實例說明-第一層網路(Level 1)強度分析計算階段:

表 3-1 為各使用者的標籤使用偏好，T_user1 為本實例的目標使用者，而 User2-5 是從 User1 延伸出去的第一層人際網路，User6-17 則是從 User2-5 再延伸出去的第二層人際網路，並以此表做為使用者之間的相似度計算。

表 3-1 使用者標籤偏好表

第一層	Tag1	Tag2	Tag3	Tag4	Tag5	Tag6	Tag7	Tag8	Tag9
T_user1	2		1				1		
User2	1	1		1	1				1
⋮									
User5						1		1	
第二層									
User6	1				2				

⋮	1	1
User17	1	2

Step1. 社交網路強度分析

表 3-2 為各使用者 User2-5 與目標使用者的屬性值，以 User2 為例，表示該位使用者與目標使用者的相似度為 0.365148，而此數值是以公式(1)計算目標使用者與 User2 的標籤相似度而來，並且該位使用者與目標使用者共擁有 30 位好友並有 1 位共同好友、而該位使用者與目標使用者並沒有互相追隨彼此，因此屬性值為 0(互相追隨則為 1)、以及這位使用者目前有 8 個追隨者(Follower)。在此本實例將權重分為三大部份，由相似度佔 70%、交集程度佔 20%(共同好友數佔 10%、回應數佔 10%)、熱門度佔 10%所組成，各比重如下表 3-2 所示。

表 3-2 第一層使用者與目標使用者的社交網路分析

分類	tag 為基礎 70%	交集程度 20%		熱門度 10%
		相似度	共同朋友數 10%	是否互相追隨 10%
U2	0.365148	$\frac{1}{30} = 0.033$	NO(0)	8
U3	0.369274	$\frac{27}{30} = 0.9$	YES(1)	30
U4	0.365148	$\frac{9}{30} = 0.3$	YES(1)	1
U5	0	$\frac{3}{30} = 0.1$	YES(1)	3

Step2. 社交網路分析指標正規化階段

各指標數值來自於不同基準的資料，為了將各指標數值做關聯分數權重加總的計算，避免某一特徵值過大或過小而影響了整體資料的準確性，故在此步驟進行公式(4) 正規化[0,1]的計算，表 3-3 為各欄位指標值正規化後的結果。

表 3-3 正規化的社交網路分析

分類	tag 為基礎 70%	交集程度 20%		熱門度 10%
		相似度	共同朋友數 10%	是否互相追隨 10%
U2	0.989	0	0	0.241
U3	1	1	1	1
U4	0.989	0.308	1	0
U5	0	0.077	1	0.069

Step3. 計算關係分數 (Related Score, RS)

將各正規化分數依照各別權重計算每位使用者與目標使用者的總分， $RS(T_U1,U2)$

=0.654、RS(T_U1,U3)= 0.85、RS(T_U1,U4)=0.693、RS(T_U1,U5)=0.118，計算過程如下所示：

RS(T_U1,U2): $(0.989 \times 0.7) + 0.2 \times (0 \times 0.1 + 0 \times 0.1) + (0.241 \times 0.1) = 0.654$
 Step1~Step3.計算完成後即完成第一層使用者(Level 1)與目標使用者的關聯分數計算，再由分數較高的前 k 位使用者各自延伸至下一層(第二層)的追隨者(Following)，進行第二層使用者與第一層使用者之間的關聯分數計算，同樣取出各別的分數最高者，如圖二。同樣步驟直到第 m 層的計算為止，在此以 k 為 4 位使用者，m 為 2 階層做為本研究實例說明。

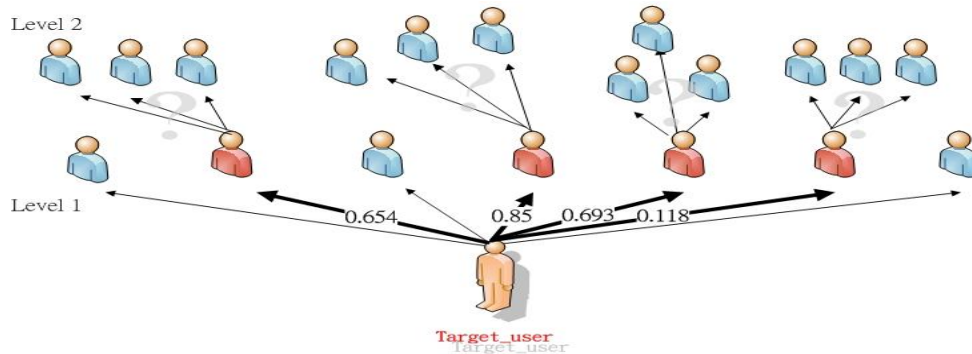


圖 3-2 社交網路分析計算階段圖

(二) 實例說明-第二層網路(Level 2)強度分析計算階段:

由第一層關聯分數計算完畢後可得到 Level1 的前 4 位使用者與目標使用者擁有較高的關聯強度者(圖 3-2 中的紅色使用者)，而再由這些前 4 位分數較高者使用者各別延伸自下一層的 Following(為目標使用者的第二層)，重覆 Step1~Step3 可得第二層與第一層使用者的社交網路強度分數。為了節省頁面空間不再重覆展示第二層 Step1~Step3 計算過程。

假設第二層與第一層之間的前 4 位擁有較高社交網路分數使用者分別為 U6、U7、U8 和 U9，求出結果如表 3-4 所示。本實例為了從第二層使用者的收藏中找尋推薦項目給目標使用者(T_U1)，故在本階段計算目標使用者與第二層使用者的直接關係分數，而計算方式則由已知的「目標使用者與第一層使用者(目標使用者的 Following)之間的分數」以及「第一層與第二層使用者之間的分數」依權重加總推算而來，各層的權重值則是該層的倒數，本實例以(T_U1,U6)= 0.689、(T_U1,U7)=0.885、RS(T_U1,U8)=0.795 和 US(T_U1,U9)=0.256 為前 4 個目標使用者至第二層社交網路較強的連線分數，計算範例如下：

$$RS(T_U1,U6)=\frac{RS(T_U1,U2) \times \frac{1}{\#n_1} + RS(U2,U6) \times \frac{1}{\#n_2}}{\sum \frac{1}{\#n_i}} = \frac{(0.654 \times \frac{1}{1}) + (0.76 \times \frac{1}{2})}{\frac{1}{1} + \frac{1}{2}} = 0.689$$

表 3-4 目標使用者與第二層使用者的關係分數(RS)

		社交網路分析總分			
使用者		RS(T_U1,U2)	RS(T_U1,U3)	RS(T_U1,U4)	RS(T_U1,U5)
階層					

L0-L1	0.654	0.85	0.693	0.118
使用者 階層	RS(U2,U6)	RS(U3,U7)	RS(U4,U8)	RS(U5,U9)
L1-L2	0.76	0.955	1	0.533

二、根據目標使用者的收藏中建構興趣輪廓:讓近期興趣加重

使用者的興趣可能在不同時期會有不同的收藏偏好，為考慮推薦系統的準確度，本研究針對使用者近期某段時間內的標籤使用情形找出近期偏好標籤，在此階段考量到標記頻率(Frequency)以及時間權重(TimeWeight)。使用者對某標籤的使用頻率可視為對此標籤的標好程度，當頻率越高則代表使用者對該標籤有較高的偏好，而使用該標籤的時間點距離現在越近則佔有較高的時間權重，越遠則反之，公式分別如(6)、(7)。最後綜合公式(6)、(7)兩項指標計算出使用者對每個 Tag 在特定期間的標籤偏好分數(Tag Preference Score, TPS)如公式(8)。

$$\text{TimeWeight} = 1 - \frac{\text{距離現在時間的天數}}{\text{觀察天數 } M} \quad (6)$$

$$\text{Frequency} = \frac{\text{Tag 該天出現頻率}}{M \text{ 天內所有 Tag 的使用頻率}} \quad (7)$$

$$\text{標籤偏好分數(TPS)} = \sum(\text{TimeWeight} \times \text{Frequency}) \quad (8)$$

(一) 實例說明:

本研究以目標使用者的最近 30 天為範例找出在這段期間內的偏好標籤，表 3-5 的 Tag1 到 Tag9 呈現目標使用者的每日標籤頻率使用情形，各標籤欄位的使用頻率以公式(7)表示之，最右欄以 30 天為基礎時間權重則是以公式(6)計算出該天在這段期間內所佔的權重。

表 3-5 目標使用者的標籤偏好分數

	Tag1	Tag2	Tag3	Tag4	Tag5	...	Tag9	以 30 天為基礎
9 月 11 日	$\frac{3}{620}$	$\frac{1}{620}$						$1 - \frac{30}{30} = 0$
⋮								
10 月 4 日	$\frac{2}{620}$	$\frac{1}{620}$		$\frac{8}{620}$	$\frac{11}{620}$		$\frac{2}{620}$	$1 - \frac{6}{30} = 0.8$
10 月 9 日 (Today)	$\frac{1}{620}$						$\frac{2}{620}$	$1 - \frac{1}{30} = 0.967$
Total Fre:	9/11-10/10 total tags frequency:620							

依表 3-5 使用公式(8)計算出 Tag1-9，得到前 w 個(本例以 w=3)為近期偏好標籤:Tag2=0.01、Tag4=0.0118、Tag5=0.0303 標籤分數，下式為計算過程示範:

$$\text{Tag1} = \left(\frac{3}{620} \times 0\right) + \left(\frac{2}{620} \times 0.8\right) + \left(\frac{1}{620} \times 0.967\right) = 0.0042$$

三、貝氏推薦公式計算階段

表 3-6、表 3-7 為經由第一階段計算而找出的較高關聯分數第二層使用者，其各別與目標使用者的強度如 User6~9 的括號所示，此階段將根據這些使用者的收藏及標籤使用情形依照目標使用者的近期偏好標籤(Tag2、4、5)做貝氏推薦分數的計算，並將前 x 個分數較高的收藏項目推薦給目標使用者。

表 3-6 第二層關聯使用者的收藏情形:

User \ Item	Item1	Item2	Item3	Item4	Item5
User6(0.689)			✓		✓
User7(0.885)		✓			
User8(0.795)	✓	✓		✓	
User9(0.256)	✓				✓

表 3-7 第二層關係使用者收藏項目的標籤使用情形

	Item1	Item2	Item3	Item4	Item5
Tag1	1		1		
Tag2	1	2			1
Tag3			1		1
Tag4		1	2		
Tag5		1	1		
⋮					
Tag9				1	

本研究改良(Kim et al., 2010)推薦如公式(9)，將目標使用者近期偏好標籤權重與社交網路關聯分數(RS)納入考慮，影響力即為擁有該收藏的使用者與目標使用者的關聯分數，如公式(10)，當分數越高代表影響力越高，即此項目被目標使用者可能有興趣收藏機率較高，而次方項的目標使用者近期偏好標籤分數(TPS)也代表著目標使用者對該標籤的近期偏好程度。

$$P_{u,y} = P(i_y | CTS_w(u)) = \frac{\sum_{u=1}^k R_{u,y}}{\sum_{y=1}^n \sum_{u=1}^k R_{u,y}} \prod_{j=1}^w \frac{1+Q_{j,y}}{m+\sum_{t=1}^m Q_{t,y}} (A_{u,j}+1) \quad (9)$$

$$P_{u,y} = P(i_y | CTS_w(u)) = \frac{\text{有收藏}(i_y)\text{使用者的RS}}{\sum_{\text{User}=1}^U \text{RS} \times U\text{-n的收藏量}} \prod_{j=1}^w \frac{1+Q_{j,y}}{m+\sum_{t=1}^m Q_{t,y}} \text{(TPS)} \quad (10)$$

(一) 實例計算:

依照表 3-9 使用公式(10)中的項目被收藏的機率先計算出 $item_{1\sim5}$ 的機率值 :

$$P(I = item_1) = 0.202 \cdot P(I = item_2) = 0.322 \cdot P(I = item_3) = 0.137 \cdot P(I = item_4) = 0.152 \cdot P(I = item_5) = 0.186$$

，以下為計算過程示範:

$$P(I = item_i) = \frac{\text{有收藏}(i_y)\text{使用者的RS}}{\sum_{User=1}^{U_n} RS \times U_n\text{的收藏量}}$$

$$P(I = item_1) = \frac{0.795+0.256}{0.689 \times 2 + 0.885 \times 1 + 0.795 \times 3 + 0.256 \times 2} = 0.202$$

計算出每項書籤 $item_{1\sim 5}$ 的被收藏機率後，將值代入公式(10)求出最後結果並依照分數高低做為推薦的順位依據，計算過程如下：

$$\begin{aligned} P_{user_1, item_1} &= \\ P(I = item_1) \cdot P(t_5|I = item_1)^{(TPS_{T_5})} \cdot P(t_4|I = item_1)^{(TPS_{T_4})} \cdot P(t_2|I = item_1)^{(TPS_{T_2})} \\ &= 0.202 \times \frac{1+0^{(0.0303)}}{9} \times \frac{1+0^{(0.0118)}}{9} \times \frac{1+1^{(0.0101)}}{9} = 0.181376 \end{aligned}$$

$$\begin{aligned} P_{user_1, item_2} &= \\ P(I = item_2) \cdot P(t_5|I = item_2)^{(TPS_{T_5})} \cdot P(t_4|I = item_2)^{(TPS_{T_4})} \cdot P(t_2|I = item_2)^{(TPS_{T_2})} \\ &= 0.322 \times \frac{1+1^{(0.0303)}}{9} \times \frac{1+1^{(0.0118)}}{9} \times \frac{1+2^{(0.0101)}}{9} = 0.298907 \end{aligned}$$

$$P_{user_1, item_3} = 0.137 \times \frac{1+1^{(0.0303)}}{9} \times \frac{1+2^{(0.0118)}}{9} \times \frac{1+0^{(0.0101)}}{9} = 0.126375$$

$$P_{user_1, item_4} = 0.152 \times \frac{1+0^{(0.0303)}}{9} \times \frac{1+0^{(0.0118)}}{9} \times \frac{1+0^{(0.0101)}}{9} = 0.135529$$

$$P_{user_1, item_5} = 0.186 \times \frac{1+0^{(0.0303)}}{9} \times \frac{1+0^{(0.0118)}}{9} \times \frac{1+1^{(0.0101)}}{9} = 0.16701$$

肆、實驗展示

一、 實驗資料來源與收集方式

本研究實驗資料來源為社交書籤分享網站 Diigo(<http://www.diigo.com/index>)，並為獲取 Diigo 網站上的資料做為資料來源，以 Microsoft Visual Studio 2010 C# 程式語言撰寫了資料擷取的程式，擷取資料可分為三部份：(1.)擷取目標使用者輪廓資料、(2.)擷取目標使用者第一層朋友相關資料、(3.)擷取目標使用者第二層使用者資料。

二、 實驗設計與結果

為驗證本實究所提出的社交網路指標具有參考價值，本實驗隨機抽取 15 位 Diigo 用戶做為目標使用者，而加總這些使用者各自的第一層好友數為 153 人，並找出在這 153 人中自各與目標使用者關聯分數最高者，因此在第一層中有 15 位延伸至下一層(第二層)，而第二層總人數為 316 人，兩層總和為 469 人，進行以下統計驗證。

(一) 共同好友、熱門度與相似度統計分析結果

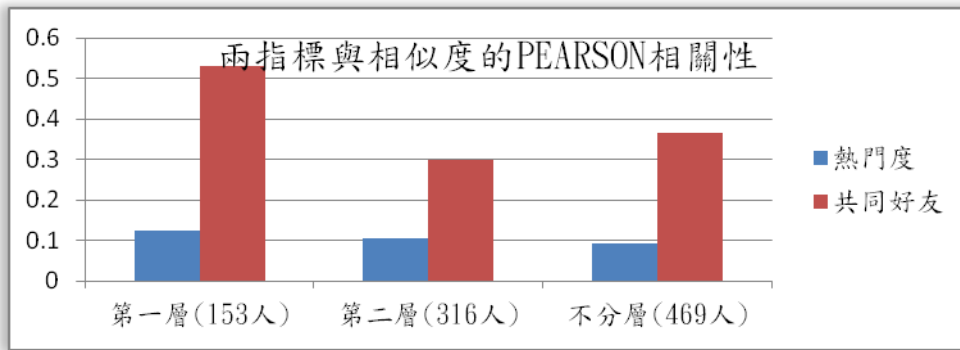


圖 4-1 共同好友與熱門度的相關性比較

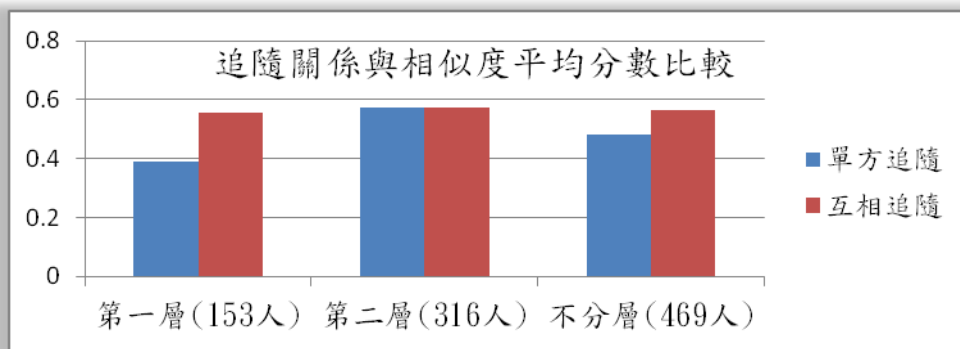


圖 4-2 追隨方式與相似度的差異

熱門度推薦為常被使用的推薦方式，因此本研究將所提出的指標與熱門度做比較。如圖 4-1 所示，經由本研究針對各層與不分層比較，共同好友數比起熱門度，與相似度擁有較高的皮爾森相關性。另外圖 4-2 結果顯示，兩位使用者彼此互相追隨比起只有單向追隨，擁有較高的相似度。

(二) 好友數與共同好友統計分析結果

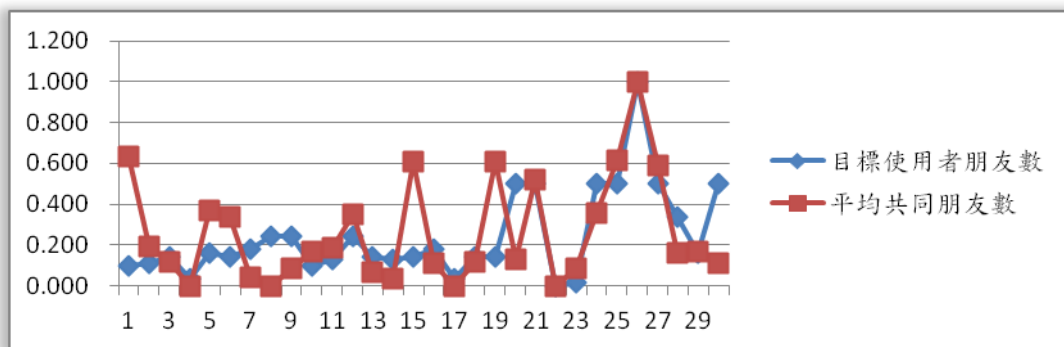


圖 4-3 目標使用者好友數與共同好友關係圖

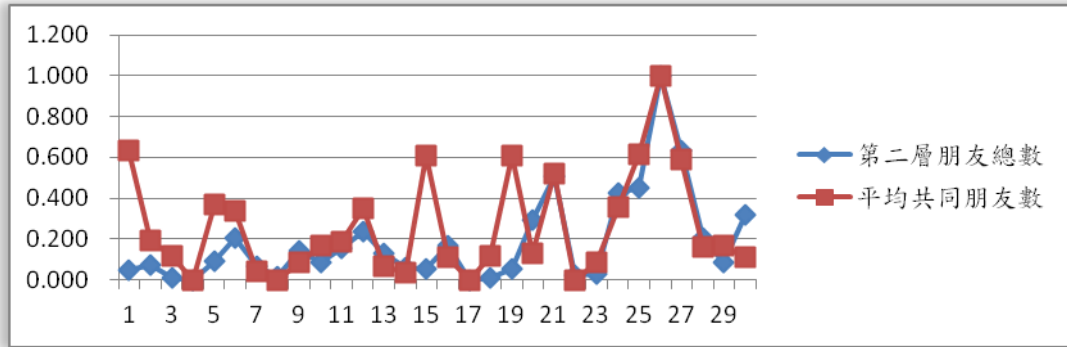


圖 4-4 第二層好友總數與共同好友關係圖

如圖 4-3 所示，本研究針對 30 位使用者好友數與共同好友數進行相關性驗證發現，當使用者擁有較高的好友數時，共同好友數也會較高，相關係數為 0.602。或者另一情況如圖 4-4 所示，當使用者第一層朋友的朋友(第二層)人數越多時，也會提升與目標使用者的共同朋友的數量，相關係數為 0.682。

(三) 書籤推薦結果關鍵字分析

本研究針對 15 位目標使用者進行推薦進行評估，目前本研究進行第一階段的實驗，推薦方式是從目標使用者的第二層 following 之中，挑選關聯分數(公式 5)最高者一人，將其最近期收藏的書籤一篇推薦給目標使用者。

評估方式為三種:(一)、比對欲推薦書籤的擁有人是否已被目標使用者追隨(二)、比對欲推薦書籤的關鍵字，是否與目標使用者的前 10 項常用標籤，或近期使用過的 19 項標籤重疊(三)、比對欲推薦書籤的關鍵字是否與目標使用者近期收藏書籤的關鍵字重疊。

15 位目標使用者經由上述三種評估比對後發現，符合比對方式(一)的有 7 位，符合比對方式(二)的有 3 位，符合比對方式(三)的有 1 位。結果顯示這樣的推薦方式僅有 4 位未與目標使用者沒有關鍵字或標籤的重疊關係。

伍、結論與未來發展

經由上述的分析資料，可以驗證在社交網路中的幾項發現:(一)、共同好友數量與相似度擁有正向關係(二)、好友數量越多，共同朋友數量也會越多(三)、由本研究所提出的關聯分數去做以近期收藏的項目推薦，有 73.3%(15 位中的 11 位)是與目標使用者有興趣重疊的。

以上的發現相對於以往以相似度計算為主的推薦系統有以下優勢:(一)、當目標使用者未收藏任何書籤、標籤時，仍就能依照社交網路分數給予推薦項目(二)、共同好友數量與追隨關係可做為推薦系統的過濾指標，並且好友數量越多，共同好友數越有其參考價值(三)、以近期興趣推薦能符合使用者所需。

未來研究方向將以貝氏推薦公式計算從關聯分數找出的第二層使用者所有收藏項目，被目標使用者收藏的機率，依此給予項目推薦，實驗驗證方式將以 Diigo 使用者為對象，進行實驗評估與第二階段之問卷調查，以評估使用者對於本系統的推薦結果是否

滿意，預期本研究提出的推薦模式，除了考慮目標使用者的收藏的相似度，也能考慮社交網路追隨的多元可能性，可以讓目標使用者接收到人際網路中更多樣的知識收藏。

參考文獻引用

1. 卜小蝶, 淺談社會性標記之意涵與應用, Web2.0 與圖書館研討會論文集, 17-37, 2006
2. 邱議德, 以社會網路分析法評估工作團隊知識創造與分享, 國立中正大學資訊管理學系碩士論文, 2003
3. 陳威志, 以標籤為基礎並結合使用者近期興趣之推薦系統, 國立第一科技大學資訊管理學系碩士論文, 2010
4. O'Donovan, J., & Smyth, B. (2005). "Trust in recommender systems," In Proceedings of the 10th international conference on intelligent user interfaces, 167-174.
5. Kim, H.-N., Ji, A.-T., Ha, I. and Jo, G.-S., 2010 "Collaborative filtering based on collaborative tagging for enhancing the quality of recommendation," *Electronic Commerce Research and Applications*, vol. 9, Issue 1, pp. 73-83.
6. Raacke, J. & Bonds-Raacke, J., (2008). "MySpace and Facebook: applying the uses and gratifications theory to exploring friend-networking sites," *CyberPsychology and Behavior*, 11(2), 169-174.
7. Garton, L., Haythornthwaite C., & Wellman, B., (1997). "Studying online social networks," *Journal of Computer-Mediated Communication*, 3(1).
8. D.Shen, J.T. Sun, Q. Yang, & Z. Chen, (2006). "Latent friend mining from blog data," *Sixth IEEE International Conference on Data Mining (ICDM)*, 552-561.
9. Yanbe, Y., Jatowt, A., Nakamura, S., & Tanaka, K., (2007). "Can social bookmarking enhance search in the web? ," *Proceedings of the 7th ACM/IEEE-CS Joint Conference on Digital Libraries*, 107-116.
10. O'Reilly, T., What is web 2.0.,(2005). "Design patterns and business models for the next generation of software," *Communications & Strategies*, 17-37.
11. Mathes, A.,(2004). "Folksonomies-Cooperative Classification and Communication Through Shared Metadata. Technical Report," LIS590CMC, Computer Mediated Communication, Graduate School of Library and Information Science, University of Illinois, Urbana-Champaign.
12. Kolari, P., Finin, T. & Lyons, K., (2007). "On the structure, properties and utility of
13. P. Jason Morrison, 2008, "Tagging and searching: Search retrieval effectiveness of folksonomies on the World Wide Web," *Information Processing & Management*, Volume 44, Issue 4, 1562-1579.
14. Golder, S.A., & Huberman, B.A., (2006). "The Structure of Collaborative Tagging Systems," *Journal of Information Science*, 32, 198-208.
15. L.A. Adamic and E. Adar., (2002). "A friends and neighbors on the web," Technical report.
16. P. Domingos and M. Richardson., (2001). "Mining the network value of customers,"

SIGKDD, 57-66.

17. M. F. Schwartz and D. C. M. Wood., (1993). "Discovering shared interests using graph analysis," *ACM*, 36(8), 78-89.
18. Rashid, A.M., et al., (2002). "Getting to know you: learning new user preferences in recommender systems," *Proceedings of the IUT 02*,127-134.
19. Rashmi Sinha, Kirsten Swearingen, "Comparing recommendations made by online systems and friends," *Proceedings of the DELOS-NSF Workshop on Personalization and Recommender Systems in Digital Libraries*, Dublin, Ireland, June 2001.

Interest Discovering from Social Networks for Social Bookmark

Recommender Systems

Chung-Hsun, Liao

Dep. of Information Management of National Kaohsiung First University of
Science and Technology
u9924808@nkfust.edu.tw

Cheng-Lung, Huang

Dep. of Information Management of National Kaohsiung First University of
Science and Technology
clhuang@nkfust.edu.tw

Abstract

Researches have focused on collaborative recommendation via finding similar users for the target user. Recently social networks have been widely used in Web2.0. This study proposes an improved recommender system that combines the target user's recent interest, social network (following), and popularity of users (follower). In addition to the cosine similarity between users, this model calculates ratio of common friend, and status of following and follower. This improves the diversity of recommended items. The experimental evaluation reveals that the proposed model is promising.

Keyword: Naive Bayes Classifier 、 Collaborative Filtering 、 Social Network 、 Recommender System