

實務專題計畫摘要報告 推薦系統

專題編號：103-CSIE-S008

執行期限：102 年第 1 學期至 103 年第 1 學期

指導教授：尤信程

專題參與人員：100820336 李昀昊

一、摘要

此次專題是為了幫助同學進一步了解推薦系統的內容以及基本原理，而用 C++ 語言撰寫的簡單推薦程序。

先從網路現有資料庫獲取信任資料，再由程式讀入，經過程式推薦演算法的計算，可以得知某用戶可能會喜好的目標，由不同的資料庫可以得到可能喜好的音樂或是可能認識的朋友等等...

關鍵詞：recommender system 、trustlet 、 ratings 。

二、緣由與目的

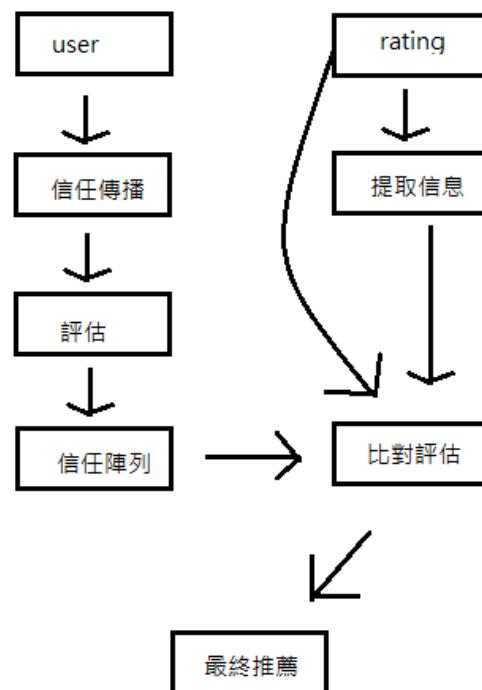
隨著科技的進步，網路上的數據越來越多，在搜尋想要的資料後出來的往往不是自己想要的，而推薦系統就是希望能藉由使用者的喜好資訊來提供可能是使用者也喜歡的事物，現今大社群網站 facebook 、或是音樂購物網站 youtube 、 amazon 都有提供完善的推薦系統，例如：可能認識的好友、建議影片及建議購買物品等...

這次專題使用大學所學的知識來寫出一個簡單的推薦系統，將大型購物網站資料庫取出的使用者喜好記錄讀入程式中，再經嘗試後得到最好的演算法來取得 rating 比較高的值當作是推薦目標，然後回傳給使用者以達成推薦步驟。

由 facebook 的讚好、粉絲專業、共銅好友我們就可以知道推薦系統的實用性，雖然比不上其大型系統的完整架構以及演算法的複雜度，但這次專題能方面同學認識它基本的架構以及用法，並以此為基礎向上完成更完善的推薦系統。

三、研究報告內容

(一) 推薦系統主要架構



user 經由輸入或點擊喜好的人或影片之訊息，經過使用者到推薦系統的傳播路徑(信任傳播)，再經由評估後存入信任陣列(資料庫)中，而另一方面，推薦系統部分也提供等級，例如喜歡此影片的人也為其他哪些影片提供幾星評價，經由提取後與使用者剛剛點擊的資料進行演算法比對，進而推薦出最後多種結果。

(二) 專題使用的資料庫

由 Epinions 網站擷取出來的文字資料庫 (Epinions 為國外大型購物網站，有介紹與販賣各種不同的事物，例如各種電影，音

樂等，也有提供用戶信任系統，可以將用戶喜好的對象加以記錄，而提供可能也會喜歡的對象)，其格式是由"使用者 ID" "信任對象 ID" "信任值 1"三串字元組成，表示使用者信任信任對象。而其資料量約為四十八萬筆。

檔案(F) 編輯(E) 格式(O) 檢視(V) 說明(H)

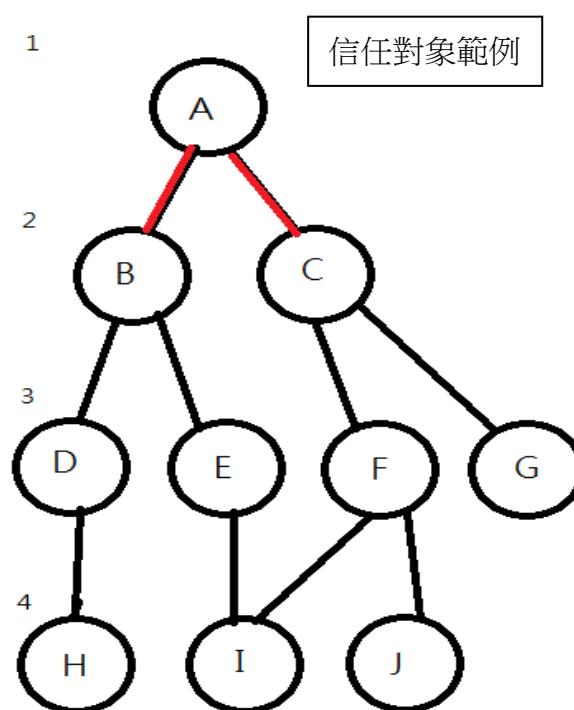
```
3624 664 1 3624 2199 1 3624 575 1 3624 8820
1 3624 2047 1 3624 289 1 3624 571 1 3624
3605 1 3624 653 1 3624 988 1 3624 286 1 3624
616 1 3624 3253 1 3624 2123 1 3624 452 1
3624 517 1 3624 1845 1 3624 1035 1 3624 2687
1 3624 516 1 17674 20747 1 17674 3481 1
17674 2551 1 13554 37814 1 13554 1747 1
```

而喜好音樂資料則是由"使用者 ID" "喜好音樂編號" "加權值"組成，資料量約為七十萬筆。

檔案(F) 編輯(E) 格式(O) 檢視(V) 說明(H)

```
|1 158 4 1 159 5 1 160 2 1 161 5 1 162 5 1 163 2
1 70 5 1 71 5 1 72 5 1 7 3 1 73 4 1 74 2 1 75 3
4 3 4 335 4 4 336 5 4 337 4 4 338 4 4 339 5 4 34
8 4 7 459 4 7 460 4 7 461 4 7 462 4 7 463 3 7 46
6 4 7 587 4 7 588 5 7 589 4 7 590 4 7 591 4 7 59
```

三)取樣標準



A	-----
B	001, 015
C	001, 002, 009
D	005, 014
E	016, 019, 100
F	001, 005, 009
G	002, 008, 015
H	003, 009
I	001
J	-----

每個對象信任的例子

這次推薦系統採用簡易的加權法，假設 A 只信任 B 和 C，則我們將 B 與 C 信任的物件 001,001,002,009,015 加權 x1 分(其中 001 總共為兩分)，再將其之下 D E F G 信任的物件 x0.5 分，再將更深層的 H I J 信任的物件 x0.25 分，最終比對分數高低後取比較高的幾項為推薦對象。由於加權關係，經過專題實作後發現搜尋至第三層以後的推薦結果都不易變動，所以為了加速執行效率，在第三層後採用隨機選取 1000 信任人避免因龐大資料指數成長導致效能低落。

四、參考文獻

- [1] WEIWEI YUAN, DONGHAI GUAN, and SUNGYOUNG LEE, "Recommender System," FTRA AIM 2013, pp. 161~162, February 21~23
- [2] Recommender_system
http://en.wikipedia.org/wiki/Recommender_system
- [3] Datasets
<http://www.trustlet.org/wiki/Datasets>