

東 海 大 學

工業工程與經營資訊學系

碩士論文

利用社群網路來補強推薦系統

研 究 生：林佳慧
指 導 教 授：王偉華 教授

中 華 民 國 一 〇 四 年 七 月

Enhancing Recommender Systems by Social Network

By
Chia-Hui Lin

Advisor : Prof. Wei-Hua Andrew Wang

A Thesis
Submitted to the Institute of Industrial Engineering and Enterprise
Information at Tunghai University
in Partial Fulfillment of the Requirements
for the Degree of Master of Science
in
Industrial Engineering and Enterprise Information

June 2015
Taichung , Taiwan

利用社群網絡來補強推薦系統

學生：林佳慧

指導教授：王偉華教授

東海大學工業工程與經營資訊學系

摘 要

目前社群網路的崛起，人與人之間的訊息更容易被連結起來，我們透過社會影響力將資訊集結起來成有用的資訊，社會網路推薦系統。在傳統的推薦系統中，社群網路推薦系統很少被應用與討論在此領域。本研究試圖運用朋友圈與朋友的影響力來補強推薦系統資料稀疏與精準度的問題。

我們提取了在線社群網絡資料，第一部分提出從小資料探討到大資料的觀點，從雜訊較少的區塊開始探討，倘落不足，再適度加入訊息與雜訊。第二部分根據朋友與項目之間的相關係數，演化出兩種推薦方式：目標使用者本身的意見與朋友圈的意見，中間我們提出了參數做為最適當的參考推薦比例，使結果更加個人化。

實驗分析結果顯示，朋友的意見與目標使用者本身的意見在適當的比例下接受推薦，可以降低誤差且提高我們的推薦預測值精準度；且透過我們資料的分群演算與朋友圈的連結，推薦系統的問題資料稀疏與冷啟動的問題便可獲得改善。

關鍵字詞：推薦系統、協同式過濾推薦系統、社群網路、個人化

Enhancing Recommender Systems by Social Network

Student : Chia-Hui Lin

Advisor : Prof. Wei-Hua Andrew Wang

Department of Industrial Engineering and Enterprise Information
Tunghai University

ABSTRACT

The sharp rise of social networks recently let people be more easily linked together by using social network, we collected information and transformed them into a useful message, becoming the social recommender system. The social recommender system has rarely been considered and applied in traditional recommender systems. In this paper we attempt to apply friendship and friend influence to enhance the data sparseness problem and accuracy problem in recommender systems.

We extract data from a real online social networks. At first, we discuss small data to large data, we began to explore small data that is less noise blocks. If insufficient, we would add more information and noise. The second, the correlation coefficient between the friends and items, evolved two recommender measures: target user's own opinions and friend's opinions. Then, we proposed the most appropriate reference parameter to make the recommenders more personal.

Experimental analysis revealed combining the target user's own opinions and friend's opinions, with appropriate reference parameter, could improve prediction error. Using our grouping approach and data connected from friendships, it could improve the data sparseness problem and cold start in the recommender system.

**Keywords : Recommender Systems, Collaborative-Filtering, Social Network,
Personalization**

誌謝

能完成這學業，有很多幫助我的貴人與親朋好友的出現，我真心的感謝你們在路途上的幫助與鼓勵，謝謝！

第一，要感謝的當然是我的指導教授，不僅在知識上給予我們提點在作人的道德觀與態度可是也不容輕忽。剛開始要了解我們老師的話中話時，需要回家消化一陣子後，才能體悟老師要教導我們的事情，非常有趣！

過程中，一度想放棄，但是有偉華老師與同學的鼓勵下，終於完成了不可能的任務，對我來說這是人生中另一筆戰績，戰勝恐懼與自己。很感謝老師在這兩年來，陪伴著我們爬完這座峰頂，重要的是登頂的過程，是否已盡力與堅持下去，這會是我們未來需要有的態度。

第二，要感謝的是口委們，召集人盧希鵬老師、口試委員薛文蔚老師、口試委員謝宛霖老師，用不同的觀點給予我建議與指導，讓我的論文可以更加完善。

第三，要感謝的是研究所的同學們，以勒、偉祥與珈禾，過程中我們互相討論內容與替對方煩惱對方論文的部分，讓我學到照顧別人也是一件很重要的事情，雖然並不是各自領域的專家，但是經由討論與思辨，我們可以一起找問題且一起解決問題，也讓我成長不少。

最後，要感謝的是我的家人，自從大學畢業後，支持著我在外面兩年發展自己的興趣，跳舞，之後也支持我回來念研究所，一路上無私的包容與關愛，我真的很幸福也很幸運。

林佳慧 謹誌於

東海大學工業工程與經營資訊學系

中華民國 104 年 7 月

目錄

摘要	i
ABSTRACT	ii
誌謝	iii
目錄	iv
表目錄	vi
圖目錄	vii
第一章 緒論	1
1.1 研究動機	1
1.2 研究目的問題與假設	5
1.3 論文研究架構	6
第二章 文獻探討	7
2.1 推薦系統	7
2.2 信任網路(Trust Network)	12
2.3 社群網路(Social Networks)	13
第三章 研究模型與方法	15
3.1 研究資料範圍	15
3.2 研究問題	15
3.3 研究工具	16
3.4 研究的機制建立	16
3.5 整合 Item-based 與 Combining User-based and Item-based collaborative filtering	33
第四章 實證分析	34
4.1 實驗系統架構與實驗設計	34
4.2 資料分析與整理	35
4.3 實驗結果與分析	47
4.4 小結	49
第五章 結論與建議	50

5.1 結論.....	50
5.2 未來發展方向.....	50
參考文獻.....	53
附錄一.....	55
附錄 1.1 (Part A)迴歸分析結果.....	55
附錄 1.2 (Part B，第一組)迴歸分析結果.....	56
附錄 1.3 (Part B，第二組)迴歸分析結果.....	57
附錄 1.4 (Part C)迴歸分析結果.....	58

表目錄

表 2.1 內容式推薦系統與協同式過濾推薦系統比較表.....	11
表 3.1 研究資料.....	15
表 3.2 朋友圈分群矩陣之變項名稱說明.....	16
表 3.3 演算法之變項符號說明.....	18
表 3.4 範例 1 相關係數.....	25
表 3.5 範例 2 相關係數.....	29
表 3.6 範例 3 相關係數.....	31
表 4.1 使用者與 λ 分類規則.....	35
表 4.2(Part A) MAE & λ 敏感度分析.....	38
表 4.3 Part B and Part C 分類規則.....	41
表 4.4 (Part B) MAE & λ 敏感度分析.....	44
表 4.5 (Part C) MAE & λ 敏感度分析.....	46

圖目錄

圖 1.1 US Mobile Shoppers and Penetration (2011-2016)	2
圖 1.2 Shopping list of Amazon	3
圖 1.3 Yahoo!拍賣的評價方式	3
圖 2.1 Epinions.com 使用者評價數對應.....	9
圖 3.1 截取有用資料.....	15
圖 3.2 擴增矩陣的方式.....	17
圖 3.3 情境介紹.....	20
圖 3.4 預算估計值流程圖.....	21
圖 3.5 The user-item matrix	22
圖 3.6 item-item similarity	24
圖 3.7 Item-based Collaborative Filtering 範例.....	25
圖 3.8 user-user similarity	27
圖 3.9 User-based Collaborative Filtering 範例	28
圖 3.10 Combining user-based and Item-based Collaborative Filtering.....	30
圖 3.11 Combining User-based and Item-based collaborative filtering.....	31
圖 4.1 實驗系統的架構.....	34
圖 4.2 18 位目標使用者 MAE & λ 的變化圖	38
圖 4.3 (Part A) MAE & λ 迴歸分析圖	39
圖 4.4 13 位目標使用者 MAE & λ 的變化圖	43
圖 4.5 (Part B, 第一組) MAE & λ 迴歸分析圖	44
圖 4.6 (Part B, 第二組) MAE & λ 迴歸分析圖	45
圖 4.7 4 位目標使用者 MAE & λ 的變化圖	46
圖 4.8 (Part C) MAE & λ 迴歸分析圖	47
圖 4.9 user-based, item-based & P(self & friend)三個演算法比較圖	48
圖 4.10 Level and MAE	48
圖 4.11 Level and Missing ratio	49

第一章 緒論

1.1 研究動機

隨著網際網路及資訊的蓬勃發展，網路使用人口大幅成長，各項應用日趨成熟的情況下，網際網路突破地域性與時間性之特性，瞬間完成的各式交易，創造龐大的市場經濟，更深入了個人生活中，舉凡食、衣、住、行與娛樂，使用者生活所需之要務，樣樣都可以透過網際滿足需求，儼然成為了現代人生活中不可或缺的重要部分！

商業活動受網際網路影響，許多公司將各式服務合併至平台，從傳統以銷售商品為導向的商業活動，轉變為以滿足顧客需求為導向的商業行為。提供各式各樣的個人化服務，進而促成電子商務蓬勃的發展。哈里斯民意調查(Harris Interactive)¹有 89%接受調查的使用者使用網路來做出購物決策，消費者越來越依賴網路上所提供的產品資訊，在進行購物決策之前，會先上網查詢相關資訊，幫助其作出購買的決策或者直接在網路上進行購物行為。線上購物已逐漸成為現代人購物方式的新選擇，經由行動裝置(例如，手機、平板電腦、桌上型電腦...等)在網路上直接搜尋產品資訊(產品本身、價格或是相關資訊)且直接交易購物。

經由美國紐約的網路市場調查公司 eMarketer 顯示圖 1.1，每年透過行動裝置在網路上進行消費的百分比逐年增加，近三年已增加 16.9%的行動裝置用戶進行網路消費總消費金額為 41.1 百萬美元，預計明年行動裝置線上購物消費再度成長至 84.60%，總消費金額高達到 174.9 百萬美元。

¹哈里斯民意調查(Harris Interactive) 是全球領先的市場研究公司之一，充分利用科技研究，技術和商業智慧提供商業信息與見解。

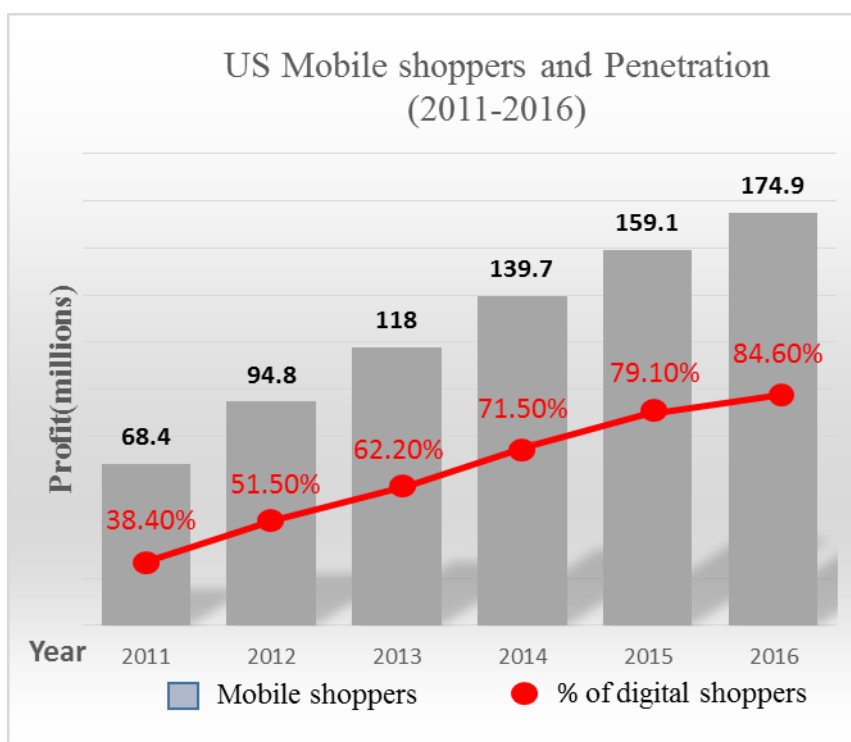


圖 1.1 US Mobile Shoppers and Penetration (2011-2016)

資料來源:eMarketer

然而，在電子商務規模的持續擴大與資訊的爆炸情況下，消費者需要花費許多心力從眾多產品種類與產品資訊中，找到真正符合需求的產品資訊，這使消費者容易陷入資訊超載(Information Overload)的問題中。Gupta and Lenhmann (2003)從消費者購買觀點，分析消費者對於通路轉換的行為，消費者購買決策過程中，哪些操作特性與影響因素使得消費者轉換了通路？研究中發現，資訊搜尋成本與評估資訊耗費心力的考量，顯著的影響了消費者對於線上購物的意願，資訊網路的發達雖然降低了搜尋的成本，卻增加了消費者評估資訊的心力，再次的形成另一種形式的「超載」！

網站如何能有效地提供切合使用者需要的資訊，正面臨著考驗，而推薦系統正是因應這個問題而產生的一個解決方案。推薦系統可以幫助使用者克服網路資訊過載問題，並改善使用者的決策制定(Haubl and Trifts, 2003)。而推薦系統典型的例子是 Amazon.com，當使用者購買時，會顯示購買同一本書的其他使用者也同時購買了哪些書，而這些欲推薦的項目將會出現在使用者購物清單中，使使用者更容易且更快速的選擇適合自己的產品。

這樣子的推薦系統是利用使用者的相似度(similarity)，再以購物習慣的

關聯性結合來作為推薦的基礎，圖 1.2。



圖 1.2 Shopping list of Amazon

資料來源：Amazon

近觀本國的例子，Yahoo！奇摩拍賣，則是結合買賣雙方互動的評價制度，透過每一次的買賣雙方的交易完成後留下一個評分記錄，作為日後購買/銷售商品時的參考依據。透過買賣雙方的交易成功紀錄所累積的信譽能力，作為店家與買家的之間的信賴橋梁，提供消費者購物決策參考的依據。

評價 (214)
拍賣商品

214 + 0 - 0 = 214

[詳細資料](#)
 [詳細資料](#)
 [詳細資料](#)

正面評價百分比：100.00% *計算方式：正評 / (正評+負評)

此會員已隱藏 11 筆 買商品評價

- 正面評價： 214 個 (兩年內共交易 145 項商品)
- 輕鬆付正面評價： 0 個 (兩年內共交易 0 項商品)
- 負面評價： 0 個 (兩年內共交易 0 項商品)
- 普通評價： 2 個 (兩年內共交易 1 項商品)

圖 1.3 Yahoo!拍賣的評價方式

資料來源：Yahoo!拍賣

近年來，社群網路的蜂起，我們發現消費者喜歡透過朋友分享購物經驗 (He and Chu, 2010; Sun, Han, Huang, Wang, Zeng, Wang and Yan, 2015)，來考量對於不熟悉的產品或者猶豫不決的商品，進行購物的決策。在現實生活中，我們對於產品的多種選擇有困惑的時候，會傾向朋友的意見與經驗分享，從中獲得直接與較可信任的經驗分享評價，比起經由前面談到的

以未熟悉的使用者給予的推薦評分來得相信與提高採納度。從社群網路 (Social Network) 補強推薦系統的文獻，顯示可改善傳統推薦系統的缺點，並擁有較佳的推薦效果。

研究觀點：

隨著科技的進步，大量計算所有的資料不會是問題，但是當資料量太大，所帶來的信號與雜訊也會增加，漸漸的會影響信號品質的問題。本研究要探討的是如圖 1.4 質與量在到達最適點前，A 點的部分。我們可以看到 A 點質與量的關係逐漸趨緩，逐漸到了頂點，過了頂點後，如果持續蒐集資料，未必對資料的質有幫助。我們用少量的資料逐漸搜集較多的資料，找出最好的質，也就是用最有相關的資料推估出最好的推薦值。

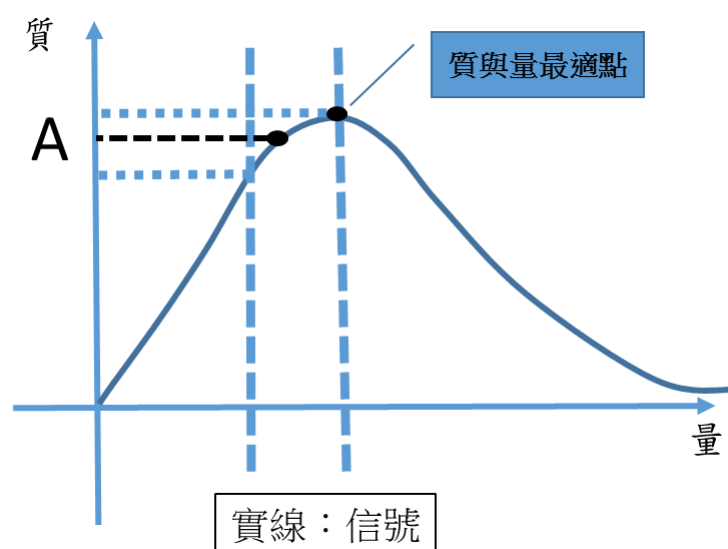


圖 1.4 質與量

資料來源：本研究整理

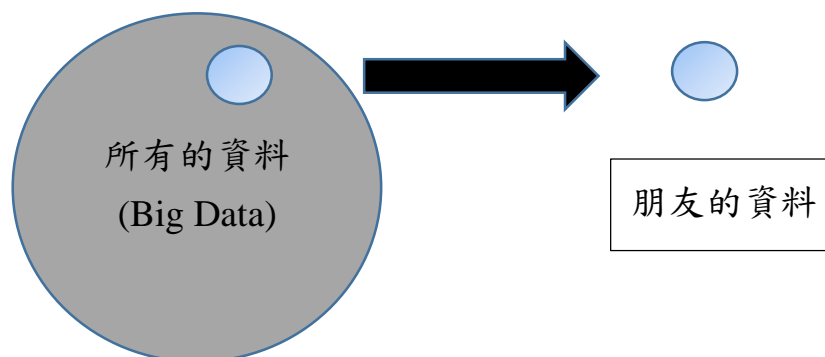


圖 1.5 截取有用資料

資料來源：本研究整理

我們考慮從雜訊較小的區塊開始建立模型，倘訊息不足，再適度加入訊息與雜訊的作為，以朋友圈的朋友為主要處理對象範圍，我們不以資料庫全部的使用者作為參考資料，透過人的關係篩選出我們所關心的資料，我們在篩選使用者，如圖 1.5。相信朋友的意見會與目標使用者的意見更相近，在我的生活中沒有實際關係的資料我們則不考慮。

綜合前述的探討，我們跳脫過去推薦系統以往不斷發明精進的演算法與求取最精確的推薦系統與評估推薦值，本研究試圖以新的觀點來探討推薦系統資與量的關係；將透過社群網路來補強推薦系統外，再以使用者本身過去的消費習慣記錄的結合，綜合推薦系統更具個人化與有效性。

1.2 研究目的問題與假設

本研究預期藉由使用者過去對相似項目的評價推薦紀錄及社群網路的社會影響(social influence)之間對項目的評價推薦紀錄加以整合(He and Chu, 2010)，過往文獻談及到利用使用者之間的相似行為估算出推薦結果(Resnick, Iacovou, Suchak, Bergstrom and Riedl, 1994; Breese, Heckerman and Kadie, 1998)，近年也有將使用者之間的信任值加入計算(O'Donovan and Smyth, 2005; Massa and Bhattacharjee, 2004; 黃河銓、李奕縉和黃昱凱, 2012; 黃河銓和王群元, 2010)，使用者本身項目的意見卻未一同納入考量。本研究將使用者的估算推薦值一同考慮，且使用傳統的協同推薦系統演算法針對下列問題進行探討與分析：

1. 透過朋友圈的資料來推估推薦值，且提出質與量的觀點，從小資料到擴增資料量的過程。
2. 使用者過去對相似目標項目的評價與社會影響(social influence)的評價推薦之間，所佔的比重需要多少程度時，所推估出來的推薦值準確度會是最適當？
3. 在巨量評分資料中，當資料不完整的狀況下，本研究對 miss rating 的問題補強提出可行的建議。

本研究演算之基本假設條件為：

- (1) 研究分析目標使用者為採納朋友意見的部分群體，對朋友意見

沒有影響者皆不在本研究討論。

- (2) 朋友關係的建立在本研究皆為相同喜好與品味相似的群體。
- (3) 在非第一層朋友圈的朋友，我們以目標使用者有共同評分過的項目數量做為使用者進入計算的門檻限制。

1.3 論文研究架構

本研究將分為五章，章節內容簡述如下：

第一章、緒論

本文章主要說明研究背景與動機、研究目的與問題與論文結構。

第二章、文獻探討

針對與本研究相關的推薦系統、目前文獻談到的演算法與因素(Term)，進行探討與整理。

第三章、研究模型與方法

說明本研究的研究模型、研究假說、變數定義與操作化、研究設計與方法。

第四章、實證分析

針對本研究提出之研究模型，進行資料分析，以及對應之研究假說檢定，最後匯總說明研究發現。

第五章、結論與建議

說明本研究的研究結論、對實務界建議、研究貢獻、研究限制以及未來之研究方向。

第二章 文獻探討

2.1 推薦系統

推薦系統 (Recommender System) 是一種使用資訊過濾 (information filtering) 的系統，可以幫助使用者過濾資訊並推薦合適的資訊給使用者。與傳統的資訊檢索系統不同的部分在於，使用者無須主動下關鍵字，系統只要藉由瀏覽紀錄，也能夠主動推薦結果 (Das, Datar, Garg and Rajaram, 2007)。在電子商務領域中，推薦系統符合使用者需求的書籍、音樂、電影和餐廳等產品，增加交叉銷售的機會 (Schafer, Konstan and Riedl, 1999)。傳統推薦機制可分為三種：(1) 內容式過濾 (content-based filtering, CN)、(2) 協同式過濾 (collaborative-filtering, CF)、與 (3) 混合式 (hybrid-based)：

2.1.1 內容式推薦系統

內容式推薦系統所使用的方法是根據使用者過去提供的 Profile，找出使用者喜好的項目特徵並進行內容分析，將其使用者喜好特徵內容從 Profile 擷取出來，並找出項目之間相同的特徵，作為推薦的根據。以書籍推薦系統為例，系統擷取使用者過去的閱讀記錄，分析該使用者常看的書籍當中，有哪些作者、作品類別、以及作品經常出現的字詞，對應系統內相似的項目特徵並給予推薦。News Weeder (Lang, K., 1995)，經由使用者提供的 Profile (使用者對於系統所提供的項目特徵感興趣的內容)，再與推薦系統裡的項目特徵集分類且進行配對。

2.1.2 協同式過濾推薦系統

協同式過濾系統所使用的方法是根據使用者過去使用的項目評分記錄，找出一群評價行為最相似的使用者群，並透過相似的使用者群所評分的項目來預測使用者對項目的喜好程度。MovieLen 電影推薦系統利用使用者的過去評價，計算使用者之間的相關係數後，由演算法估算使用者可能被推薦的分數 (Schafer, Konstan and Riedl, 1999)。

協同過濾推薦系統分為兩類，(Wang, Vries and Reinders, 2008) 分別為 Memory-Based Methods and Model-Based Methods，我們將在 2.1.2.2 與 2.1.2.3 將會談到，兩者都是透過 user-item matrix 的資料庫，運用演算法與模型的

學習來估算推薦值，根據我們獲得的 dataset 有不同方式的呈現，面對不同的推薦方式而有不同的演算法。

2.1.2.1 關於資料的收集與協同式過濾推薦系統的限制

1. 資料收集的方式

透過網路收集使用者的資料，分為顯性跟隱性的部分，顯性的資料，一般的網站系統使用評分的方式或是給予評價，使用者對於產品的喜好程度給予評分，直接回饋給系統的資料(例如，MovieLens, Yelp and My Music *et al.*)簡單來說，使用者與系統互動的方式傳回系統。隱性的部分，是根據使用者的行為模式由系統代替使用者完成評價，不需要使用者直接打分數或輸入評價資料，透過使用者瀏覽行為(例如，瀏覽時間、點閱的頻率、點擊了哪些相關的產品)或者過去消費行為的週期性(點閱的周期時間)，直接輸入到資料庫當中進行分析(黃河銓等人，2012)。

2. 協同式過濾推薦系統的限制

(1) 資料稀疏(Data Sparseness Problem)

協同過濾式推薦系統利用使用者給予的產品或物件評價作為推薦依據，但是使用者接觸的物件且給予評價的數量大多只佔系統非常小之比例。Bobadilla, Ortega, Hernando and Gutiérrez(2013)與 Massa and Bhattacharjee(2004)在協同式過濾推薦系統的例子中，在兩個公開的典型資料數據 Eachmovie and MovieLens，顯示資料稀疏比例高達 97.6% and 95.8%。在此情況下，系統如果根據使用者購買相同或相似的產品給予推薦時，將難以找到良好的使用者來進行推薦且推薦結果相關的顯著性沒有很高。透過皮爾森相關係數(Pearson correlation)尋找相似者的演算法將會造成雜訊且不可靠。Sinular Value Decomposition(SVD)以被提出減少矩陣維度技術以減少資料稀疏的問題。

Massa and Bhattacharjee(2004)說明協同式過濾推薦系統的資料稀疏與冷啟動的問題，(圖 2.1)說明協同式過濾推薦系統的資料

稀疏與冷啟動的問題，在 Epinions.com²的資料稀疏更高達 99.99135%。平均每一位使用者評價項目數量為 13.49，中位數為 4，此資料庫內容使用者 49,290 位、商品項目 139,738 和所收集的評分紀錄總數為 664,824。

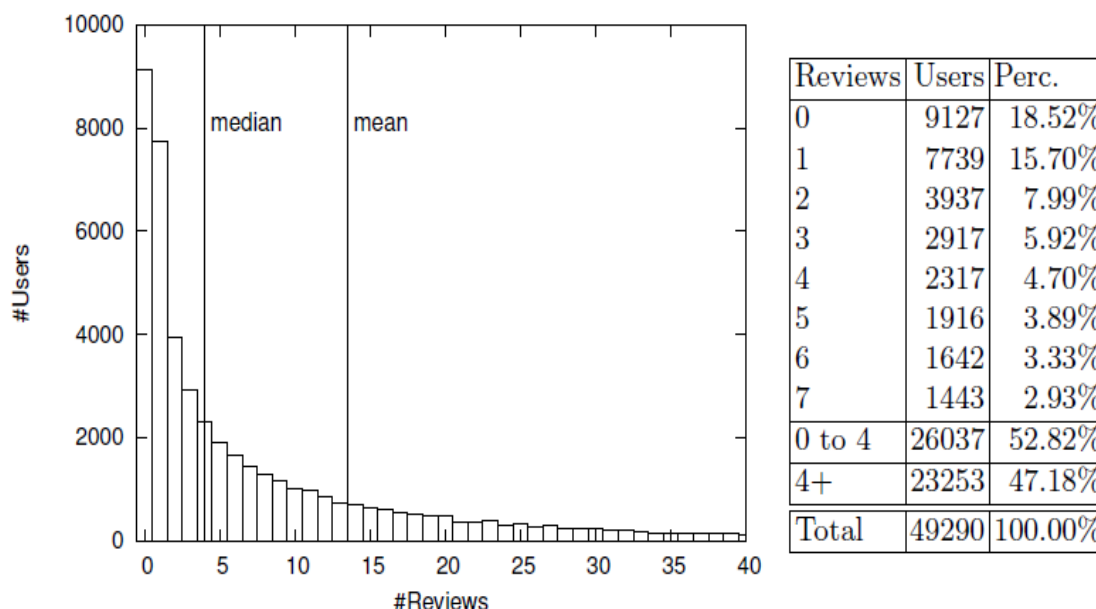


圖 2.1 Epinions.com 使用者評價數對應

資料來源：Paolo Massa and Bobby Bhattacharje, 2004

(2) 冷啟動(Cold-start)

當一個新使用者進入協同式過濾推薦系統時，系統沒有新使用者的過去評價紀錄資料，所以無法給予新使用者任何評價；或者一個新的項目進入系統中，當使用者需要新項目的推薦值時，新項目沒有任何評分紀錄，系統將無法給予任何的推薦。Massa and Bhattacharjee(2004)定義新使用者進入系統評分紀錄未達滿 5 個 item 以上，我們稱為“冷啟動使用者”(cold start users)。呈如(圖 2.1)冷啟動使用者為 52.82%，超過資料庫使用者的一半。

推薦系統是否能夠給予良好的推薦值，依循著系統背後的資料庫是否完整與充足，針對推薦系統的限制與不同推薦

²Epinions.com 是一個大眾消費者點評網站，成立於 1999 年。訪客可以瀏覽各種各樣物品的評論且可以選擇信任的使用者加入自己的信認使用者列表中，幫助他們做出正確的購物選擇；參考 <http://www.epinions.com>

系統的需求與特性，我們伴隨著不同演算模型與因子組合，例如，信譽、信任、社群網路等，我們在 2.3 與 2.4 相繼說明) 而延伸出不同的推薦系統的文獻。

2.1.2.2 Memory-Based Methods

協同過濾推薦技術，利用使用者可能喜歡參考來自與他興趣或喜好項似的使用者的建議，找出與此使用者行為與喜好最為相似的一群使用者，經由使用過去的交易紀錄或其他的特徵來預測使用者對於事物的喜好程度，進而估算出推薦值給使用者。最初提出協同過濾式推薦系統的學者為 Resnick's 標準預測模型(Paul Resnick *et al.*, 1994)，透過皮爾森相關係數(Pearson correlation)、向量相似度(Vector Similarity)計算使用者之間的相似度，前 top-n 有對目標項目(Target-Item)評分的使用者為納入 Resnick's 標準預測模型的範圍，預估推薦值目標項目的推薦值。

前述推薦模式提到的是以 user-based 為主要的相似對象作為推薦的依據，以 user-based 的協同過濾式推薦演算法隨著使用者數量的增多，計算的時間就會變長。相似者有評過的項目在 2.1.2.1 談到資料稀疏的問題，使用者平均有給予評分的項數甚少，系統無法有完整的推薦值給予目標使用者(target user)。

Sarwar, Karypis, Konstan and Riedl(2001)提出利用項目之間的相似度進行推薦值的估算，以項目為基礎的協同過濾方法有一個基本的假設，「能夠引起使用者興趣的項目，必定與其之前評分高的項目相似」，透過計算項目之間的相似性來代替使用者之間的相似性。以使用者此以 item-based 為主要的相似對象作為推薦的依據。另外也是以 item-based 為主要推薦的依據，Amazon.com，以相似項目為主要分群的依據，再將同一群的使用者利用項目之間的關聯性，給予推薦預測值(Linden, Smith, and York, 2003)

隨後 Jun Wang *et al.*(2008)提出了將 User-based and Item-based 一同討論，且結合了此兩種的相似度為並加入 Resnick's 標準預測模型運算。

2.1.2.3 Model-Based Methods

Netflix Prize 的懸賞比賽³，提出矩陣因子分解模型(matrix factorization models)，加入使用者隱性資料與使用者的 profile，使其成為一組潛在的向量因子，讓系統辨識因子群找相似的项目並給與推薦值，此方法提高推薦系統的精準度與系統運算的效率。(He and Chu, 2010; Wang, Vries and Reinders, 2008; Breese, Heckerman and Kadie, 1998)以概率的方式計算目標使用者對目標项目的機率，給與目標使用者其它每個可能評分值的條件機率，然後再以最大概率值的評分作為預測值(Koren, Bell and Volinsky, 2009)。

2.1.2.4 內容式推薦系統與協同式過濾推薦系統

表 2.1 內容式推薦系統與協同式過濾推薦系統比較表

Technique	Pluses	Minuses
Collaborative filtering (CF)	<p>A. Can identify corss-genre niches.</p> <p>B. Domain knowledge not needed.</p> <p>C. Adaptive:quality improves over time.</p> <p>D. Implicit feedback sufficient.</p>	<p>I. New user ramp-up problem</p> <p>J. New item ramp-up problem</p> <p>K. “Gray sheep” problem</p> <p>L. Quality dependent on large historicaldata set.</p> <p>M. Stability vs. plasticity problem</p>
Content-based (CN)	B, C, D	I, L, M

資料來源：(Robin Burke, 2002)

過濾推薦系統除了記錄使用者项目的評分外，對於產品特徵本身的屬性並無額外描述與分析，乃協同式過濾推薦系統的缺點。以書籍推薦的例子，常見的做法是利用資訊檢索、k 個鄰近的鄰居演算或是概念關聯 (concept association)的方式將書籍(項目特徵)加以分類，將不同的書籍類別分類，且對分類的內容進行相似度分析，已知的使用者偏好與分類間的對

³DVD 租賃通司 Netflix 於 2006 年公開這項競賽，如果能改進協同過濾式推薦演算法比 Netfix 自己的推薦系統提升 10%就能得到 100 萬美元的獎金;參考 [Http://www.netflixprize.com](http://www.netflixprize.com)

應關係加以推薦。

主要這種做法的目的就是學習使用者與物件特定屬性間的相關性來進行推薦，這樣的方式可以改進協同式過濾推薦系統中對於項目無須進行分析的漏洞，但是在物件關聯的程度，以及在使用者相似上的表現還是有改進的空間。故因此有不少文獻談提到兩種方式混合在一同討論我們稱“混合式協同推薦系統(Hybrid recommender system)”。

2.1.3 混合式推薦系統

混合式推薦系統結合內容式推薦系統與協同推薦系統的優點，可針對問題特性，結合不同推薦系統的優點以彌補各推薦系統在特性問題處理上的盲點與不足之處。(黃純敏，林重佑與黃進瑞，2013); 黃河銓與王群元，2010)以電影的推薦系統為例子，由內容式推薦系統將電影的類型與特徵區分使用者類型，當使用者要求推薦目標項目時，利用使用者所提供的 Profile，經由系統查詢過去使用者對不同電影特徵(電影的類型相似度、電影特徵值相似度)的評分喜好紀錄且估算推薦值，假設沒有此項紀錄，系統將運用協同式過濾推薦系統計算相似使用者評分過的紀錄，並經由電影的類型相似度、電影特徵值相似度估算目標使用者的推薦值。

2.2 信任網路(Trust Network)

Wang, Vries and Reinders(2008)網路購物上出現了假帳號製造假評價，賣家出於自己的利益，讓假帳號將自家商品的評價提高，使使用者對購物系統給予的評價機制失去了信心。除了系統內部的精準度提高與系統運算效率提高之外，還有哪些特質可以幫助使用者在購物的決策且提高對推薦系統的信任呢？

「信任乃是指某一群體在預期對方會表現合乎自己利益的基礎下，不管有無能力監督或控制對方的行為，願意承擔受傷害的不確定性」(Mayer, Davis and Schoorman, 1995)將人與人之間的信任度視為一種關係值，關係值越高表示信任度越高，並以信任推演的方式來進行推薦(Golbeck, 2005)，此種推薦方式可以降低資料稀疏(Data Sparseness Problem)和冷啟動(Cold-Start)的問題，當使用者之間沒有共同評價關係時故無法獲得相關係數，無法進

一步獲得推薦值，但是透過信任推演來取得關係值，並透過關係值增加推薦機會與成功率。

O'Donovan and Smyth (2005)提出計算使用者之間的 profile level trust and item level trust 成為使用者之間的信任值。Massa and Bhattacharjee (2004)透過經驗分析 epinions.com，使用者建立信任使用者列表，隨著時間的推移便可建立使用者之間的信任網路，資料稀疏與冷啟動可經由信任網路的組合得到改善。主要推薦方法先分別計算信任值(評估信任值程度)與使用者相似程度，即整合信任網路與傳統的協同式過濾推薦演算法來取得推薦值。

黃河銓，李奕縉與黃昱凱(2012)認為雖然信任網路(trust network)可以解決協同式過濾推薦系統的限制，由於系統和使用者的回饋尚顯不足，便將使用者回饋的因素結合信任網路使推薦更個人化。Arazy, Kumar and Shapira (2009)則透過消費者的回應加入商品的信譽與消費者的消費模型以傳統的協同式過濾推薦演算法來估算推薦值。

2.3 社群網路(Social Networks)

Domingos and Richardson (2001)從商業的角度，發現隱藏顧客不在於金字塔頂端較有錢的顧客群，而是利用消費者網絡價值，消費者周遭的朋友有非常大的影響在主導消費的購物決策上。

過去研究推薦系統的資料是以相似使用者與相似的項目作為推估推薦值的資料，過程中系統需要運算大量的資料，可能達到千萬筆過去使用者的評分紀錄，造成系統負擔與時間的耗費，故有人提出鄰近的 top-n 的 user-item matrix 作為收集估算值的範圍。前面提到經由信任值使資料稀疏與冷啟動的問題獲得改善，近年來社群網絡的串起，朋友之間的互動與關係很容易的可以取得，利用這樣的資料運用在推薦系統，近年來開始有文獻在討論。

以朋友的過去評分紀錄作為推薦的範圍，不僅縮小了資料的範圍且人與人之間的影响力直接反應出來，朋友不僅有相似的品味且喜歡的消費項目習慣也會相似，當我們對於不熟悉的商品與無法做決策的時候，我們往往會尋求親朋好友的意見，相信朋友的意見會比過去相似使用者的推薦更

適合作為我們下決策的考量討論在朋友之間不同喜好的朋友需要有程度上的區分，經由朋友中有相似喜好為主要分群依據，再以矩陣因子分解的方式計算推薦值(He and Chu, 2010; Sun, *et al.*, 2015)。

第三章 研究模型與方法

3.1 研究資料範圍

來自豆瓣(Douban)電影數據資料庫資(Ma, Zhou, Liu, Lyu, and King, 2011)，含有 129,490 個使用者、58,541 個項目與 16,830,839 評價的巨大資料庫，實際上每戶使用者所評價的項目卻占總資料的少數，見表 3.1。

表 3.1 研究資料

使用者	129,490
項目	58,541
評分	16,860,839
朋友關係	1,692,952

為了提高系統在尋找資料的效率，本研究截取朋友的資料為有效資料做計算，過濾其他使用者的資料，將運用最少資料做演算，圖 3.1。

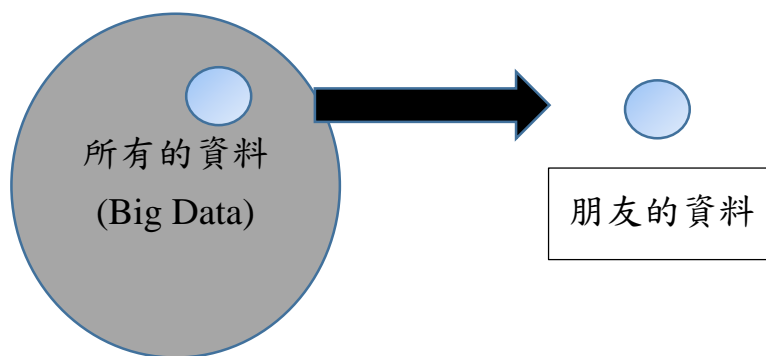


圖 3.1 截取有用資料

資料來源：本研究整理

由於朋友之間有較高的相似度與評分項目間的相似度較高，可避免一般使用者在推薦上的雜訊干擾；當遇到資料稀疏時，我們則運用社群網絡來連結不同朋友圈的資訊來解決在推薦系統上資料稀疏的問題。

3.2 研究問題

1. 透過朋友圈的資料來推估推薦值，且提出質與量的觀點，從小資料到擴增資料量的過程，提出質與量的最適點建議。
2. 使用者過去對相似目標項目的評價與社會影響(social influence)的評價推薦之間，所佔的比重需要多少程度時，所推估出來的推薦值準確度會

是最適當？

3. 在巨量評分資料中，當資料不完整的狀況下，本研究對 miss rating 的問題補強提出可行的建議。

3.3 研究工具

本研究原始資料為文字檔格式字串，利用 SQLite 打開巨量資料與檢視查閱資料，再由 Python2.7.9 編寫運算邏輯並用 SQL 指令將資料匯入運算，Python 是以物件導向的語言，有助於在資料庫的數學運算、系統管理到網站開發，以 Notepad++ 為編寫程式整合介面的工具，將不同的資料集合整合在字串中運算，再輸出 Excel 2013 後進行資料的分析，最後以 Python 數據分析視覺化模擬呈現分析結果。

3.4 研究的機制建立

1. Exploration phase

從圖 3.1 我們從原始資料擷取有用的資料，朋友的資料來作推薦。由於第一層朋友圈所能提供的推薦項目有限，故須藉由不同層的朋友圈補強第一層朋友圈無法推薦的項目，以此類推。

朋友關係的建立以實際資料朋友名單為主，隨著朋友層的擴大我們能給目標使用者的推薦項目也會增多。在擴增下一層朋友圈時，本研究提取概念(O'Donovan *et al.*, 2005)信任的概念，數據上的信任，本研究在非第一層朋友圈的朋友，我們以目標使用者有共同評分過的項目數量做為使用者進入計算的門檻限制，例如：在系統將 level2 朋友納入之前，我們設定門檻值，須與目標使用者有共同項目達到 25 個以上才納入矩陣朋友圈，此為本研究設定之參數，太小，矩陣中與目標使用者的關聯性不大，如果太大則沒有使用者進入系統，此乃本研究對信任的定義。

根據小世界理論(Milgram, 1967)，Magram 提出只要透過平均約 6 個朋友的聯繫便可以連接到所要找尋的人。我們以朋友做為連接點，且連結的層數設為 6，相信在 level 6 我們便可以找到使用者所有預想知道的推薦項目。依照朋友圈分群建置矩陣說明由表 3.2 與圖 3.2 作說明：

表 3.2 朋友圈分群矩陣之變項名稱說明

矩陣位置	變項名稱	變項名稱定義
擴增矩陣列位	第一層朋友圈($\text{friend}^{\text{level}1}$)	目標使用者的朋友。
	第二層朋友圈($\text{friend}^{\text{level}2}$)	目標使用者第一層朋友圈的朋友。
	第三層朋友圈($\text{friend}^{\text{level}3}$)	目標使用者第二層朋友圈的朋友。
	第四層朋友圈($\text{friend}^{\text{level}4}$)	目標使用者第三層朋友圈的朋友。
	第五層朋友圈($\text{friend}^{\text{level}5}$)	目標使用者第四層朋友圈的朋友。
	第六層朋友圈($\text{friend}^{\text{level}6}$)	目標使用者第五層朋友圈的朋友。
擴增矩陣欄位	第一層($\text{item}^{\text{level}1}$)	目標使用者與第一層朋友圈所有評分過的項目。
	第二層($\text{item}^{\text{level}2}$)	目標使用者、第一層和第二層朋友圈所有評分過的項目。
	第三層($\text{item}^{\text{level}3}$)	目標使用者、第一層、第二層和第三層朋友圈所有評分過的項目。
	第四層($\text{item}^{\text{level}4}$)	目標使用者、第一層、第二層、第三層朋友圈和第四層朋友圈所有評分過的項目。
	第五層($\text{item}^{\text{level}5}$)	目標使用者、第一層、第二層、第三層朋友圈、第四層和第五層朋友圈所有評分過的項目。
	第六層($\text{item}^{\text{level}6}$)	目標使用者、第一層、第二層、第三層朋友圈、第四層、第五層和第六層朋友圈所有評分過的項目。

資料來源：本研究整理

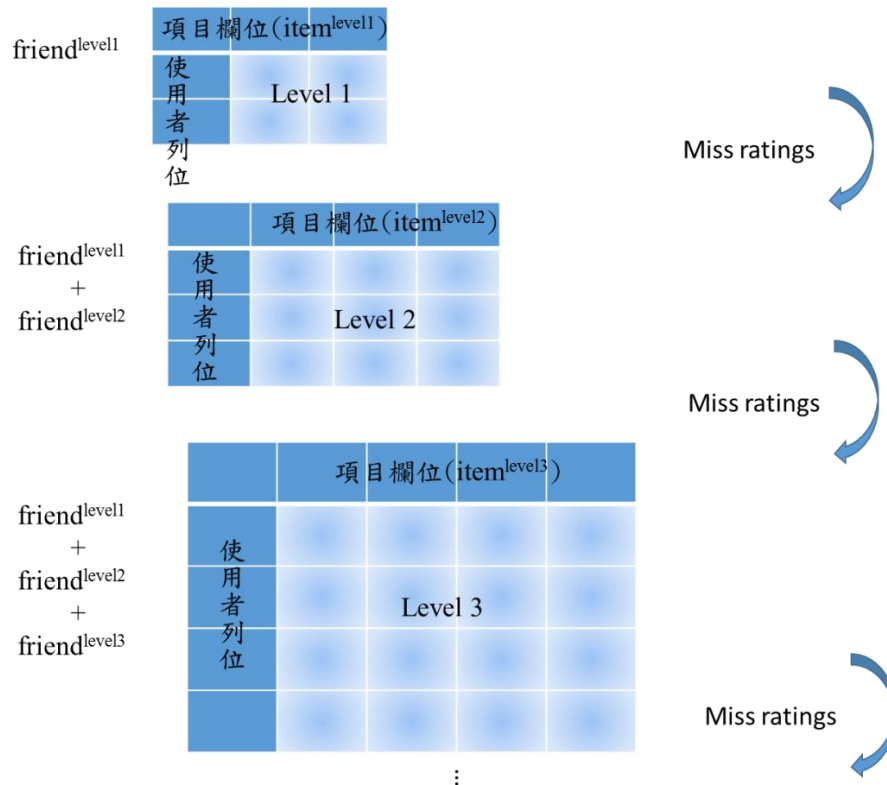


圖 3.2 擴增矩陣的方式

資料來源：本研究整理

2. Exploitation phase

經由朋友圈分群建置矩陣後，我們計算目標使用者與其朋友各自的相關係數後，再推估目標使用者待推薦的目標項目，主要演算法有 Item-based Collaborative Filtering (Sarwar *et al.*, 2001)、User-based Collaborative Filtering (Paul Resnick *et al.*, 1994)與 Combining Item-based and User-based(Wang, *et al.*, 2008)的方法，構成兩種推薦值分別為：經由目標使用者過去評分紀錄所推算的推薦值與朋友過去評分紀錄所推算的推薦值，最後我們加入參數， λ ，來測試與整合個人化的推薦值。我們所運用的符號說明如表 3.3：

表 3.3 演算法之變項符號說明

類別	變項符號	變項內容與條件	變項說明
基本符號說明	λ	$[0,1]$	目標使用者 a 採納自己過去評分習慣的意見比例。
	$1 - \lambda$	$[0,1]$	目標使用者 a 採納朋友的意見比例。
	u_a	$a=1, \dots, m$	目標使用者 a(target user)，a 有 1 到 m 位。
	u_c	$c=1, \dots, m$ 且 $c \neq a$	使用者 c，c 有 1 到 m 位且 c 不等於 a。
	i_b	$b=1, \dots, n$	目標項目 b(target item)，目標項目有 1 到 n 個。
	i_d	$d=1, \dots, n$ 且 $d \neq b$	項目 d，項目有 1 到 n 個且 d 不等於 b。
	$R_{m \times n}^p$	$R_{m \times n}^p = [r_{1,1}, \dots, r_{a,b}, \dots, r_{c,d}, \dots, r_{m,n}]$	第 p 層矩陣 R 中，所有使用者對所有項目的評分集合。
	$m = [u_1, u_2, u_3, \dots, u_m]^T$	第 p 層矩陣 R 中，列位為所有的使用者集合，使用者有 1 到 m 位。T 為轉置符號。	
	$n = [i_1, i_2, i_3, \dots, i_n]$	第 p 層矩陣 R 中，欄位為所有的項目集合，項目有 1 到 n 項。	

(續下頁)

(承上頁)

類別	變項符號	變項內容與條件	變項說明
Item-based Collaborative Filtering	$Sim_{item}(b, d)$	[-1,1]	目標項目 b 與項目 d 之間的相關係數。
	i	組數	目標項目 b 與項目 d，有同時被同一位使用者評分的組數。
	$F_i(i_b)$	$\left\{ i_b \mid \begin{array}{l} i_b \in itemlevel^p, \\ rank\ sim_{item}(b, d) < n' \end{array} \right\}$, $p=1, \dots, 6$	目標項目 b 須屬於目標使用者 a 的 $itemlevel^p$ 範圍。
	$P_{item-based}(a, b)$	[0,5]	經由 item-based 推估目標使用者 a 在項目 b 的評分，評分分數為 1 到 5 分。
User-based Collaborative Filtering	$Sim_{user}(a, c)$	[-1.1]	目標使用者 a 與使用者 c 之間的相關係數。
	j	組數	目標使用者 a 與使用者 c 有共同對相同的項目評分組數。
	$F_u(u_a)$	$F_u(u_a) = \left\{ u_c \mid \begin{array}{l} u_c \in friendlevel^p, \\ rank\ sim_{user}(a, c) < m' \end{array} \right\}$, $p=1, \dots, 6$	使用者 c 須屬於目標使用者 a 的 $friendlevel^p$ 範圍。
	$P_{user-based}(a, b)$	[0,5]	經由 user-based 推估目標使用者 a 在項目 b 的評分，評分分數為 1 到 5 分。
Combining User-based and Item-based Collaborative Filtering	$W_{user, item(a, b)}(c, d)$		使用者 c 對項目 d 的評分，與目標使用者 a 對目標項目 b 的評分之間的綜合相關係數。
	$F_{u, i}(u_a, i_b)$	$\left\{ r_{c, d} \mid \begin{array}{l} u_c \in F_u(u_a), i_d \in F_i(i_b), \\ c \neq a, b \neq d, W_{user, item(a, b)}(c, d) < k \end{array} \right\}$	$r_{c, d}$ 須在第 p 層中，且其綜合相關係數 $W_{user, item(a, b)}(c, d)$ 排行前第 k 項排列。
	$P_{friend}(a, b)$	[0,5]	經由 item-based 與 user-based，推估目標使用者 a 在項目 b 的評分，評分分數為 1 到 5 分。

資料來源：本研究整理

本研究以 Pearson correlation coefficient 與傳統協同式過濾推薦的演算

法為主要基礎(Paul Resnick *et al.*, 1994)，並且加入朋友圈的資料與 λ ，在原演算法的演算中。3.4.1 小節為情境與模型介紹。3.4.2 小節說明以 Item-based Collaborative Filtering 為基礎的相關係數與推薦值估算。3.4.3 小節說明以 User-based Collaborative Filtering 為基礎的相關係數與推薦值的估算。3.4.4 實際建構一個結合目標使用者與朋友圈過去評價紀錄的估算推薦系統。

3.4.1 情境與模型介紹

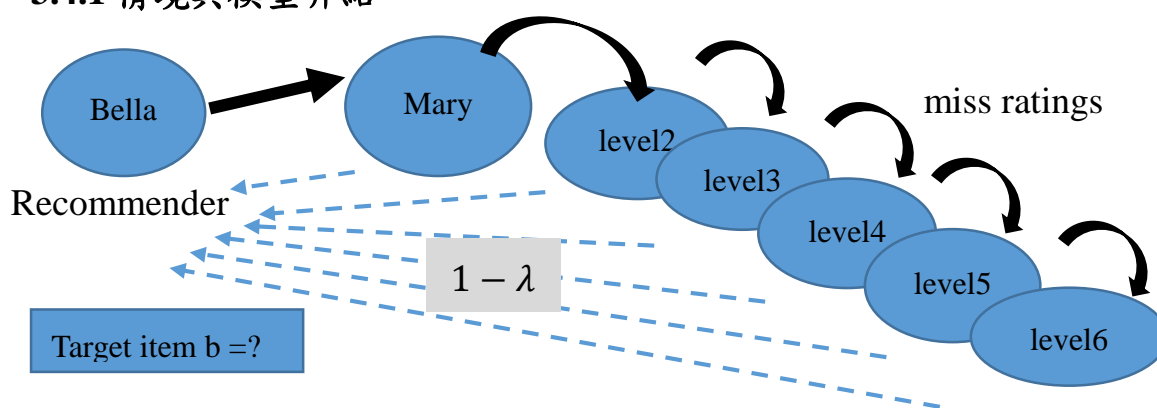


圖 3.3 情境介紹

資料來源：本研究整理

情境介紹，Mary 為 Bella 第一層(level1)其中的一位朋友，我們相信 Mary 喜歡的喜好與品味會與 Bella 相似，當 Bella 想知道 Target item b 的評價時，會詢問品味相近的 Mary 意見作為依據，如果 Mary 也沒有看過 Target item b，Mary 先會以自己看過類似項目作為意見分享，Bella 再根據自己過往對相似 Target item b 的評分，綜合其意見後再進行考量；如果 Bella 與 Mary 都沒有相似的项目經驗，則 Mary 會詢問她比較信任的朋友(level2)，如果 level2 的朋友也沒有看過 Target item b 與沒有對相似 Target item b 的项目評價過，則會再往下一層朋友去詢問。本論文以朋友為主要的相似使用者來探討，且中間我們加入了參數 λ 作為使用者本身考量的權重調整與朋友給予的建議比重 $1-\lambda$ 。在過程中，朋友圈全部的評分資料尚且不足給予目標使用者時，將透過其他朋友圈的朋友給予支援，我們相信朋友的朋友(level2)距離目標使用者的品味與喜好，會有一定程度的雷同，故第二層的朋友圈來補強第一層朋友圈的資訊稀疏問題，因此我們將尋找較信任的第一層朋友的朋友圈(level2)來給予推薦，若第二層朋友圈無法給予推薦目標項目將由第三層朋友圈給予完成，依此類推。

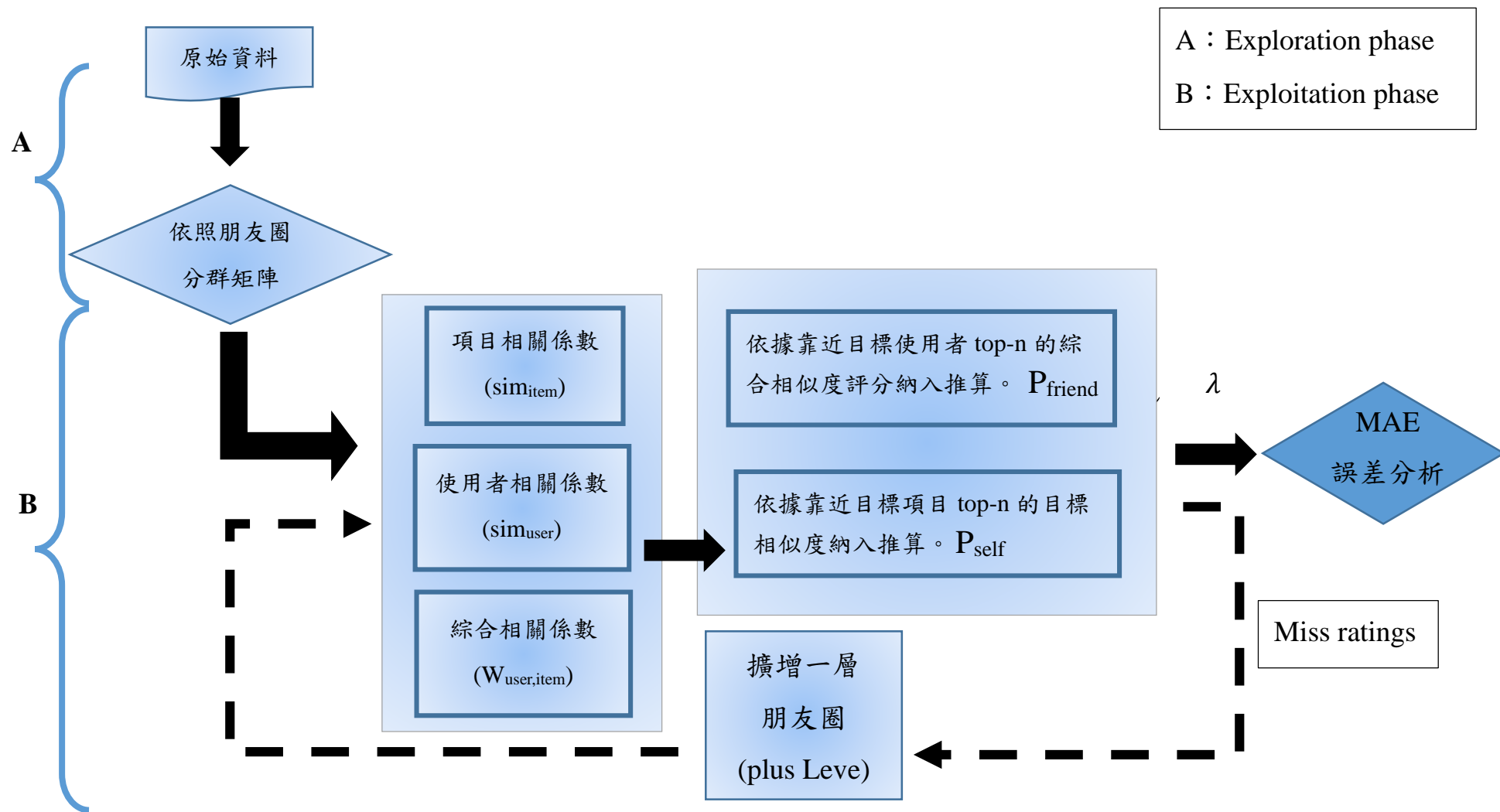


圖 3.4 預算估計值流程圖

資料來源：本研究整理

模型介紹，典型的過濾式協同推薦系統輸入的資料為 M 位 user 對於 N 項的 item 給予不同的評價分數(例，1 分到 5 分)而成為一個 M*N 的 user-item matrix。由於資料稀疏問題，不是每一位 user 給予每個 item 評價，本研究將資料維度降至 n 欄，也就是說，目前的 user-item matrix 縮減至 m*n， $n \ll N$ ，在這 n 個 item 中，至少都會有一個 user 給予評價，不會有 miss ratings 的欄位出現。

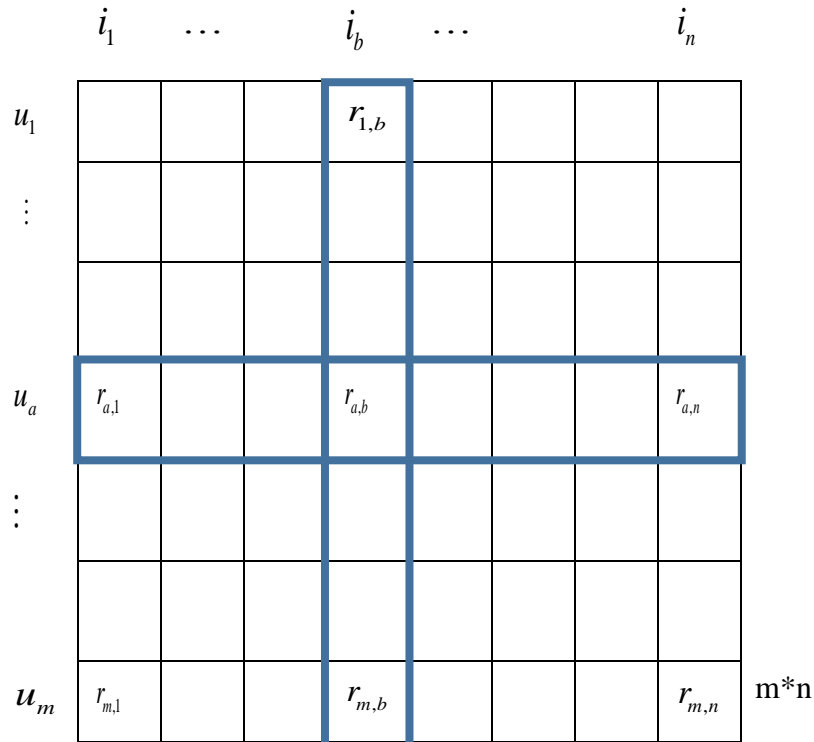


圖 3.5 The user-item matrix

資料來源：本研究整理

在本研究的模型介紹資料是以 target user 與 level1 朋友圈共 m 位使用者，target user 與 level1 朋友圈給予項目評分的欄數為 n，所評分過後的 user-item matrix 大小為 m*n (圖 3.5)。

$$R_{m*n}^1 = \left\{ R_{m*n}^1 \mid R_{m*n}^1 \in level1 \right\} \quad (3-1)$$

$$m = [u_1, u_2, u_3, \dots, u_m]^T$$

$$u_a = [r_{a,1}, \dots, r_{a,n}], a = 1, \dots, m$$

公式(3-1)為 R_{m*n}^1 矩陣中的列位，為 level1 朋友圈的各使用者 m 位，使用者對項目 1 到 n 的評價狀況。

$$R^1_{m*n} = \{R^1_{m*n} | R^1_{m*n} \in level1\}$$

$$n = [i_1, i_2, i_3, \dots, i_n]^T$$

$$i_b = [r_{1,b}, \dots, r_{mb}]^T, b = 1, \dots, n$$
(3-2)

公式(3-2)為 R^1_{m*n} 矩陣中的欄位，為項目 b 被使用者 1 到使用者 m 評價的狀況。

3.4.2 Item-based Collaborative Filtering

1. 情境說明：

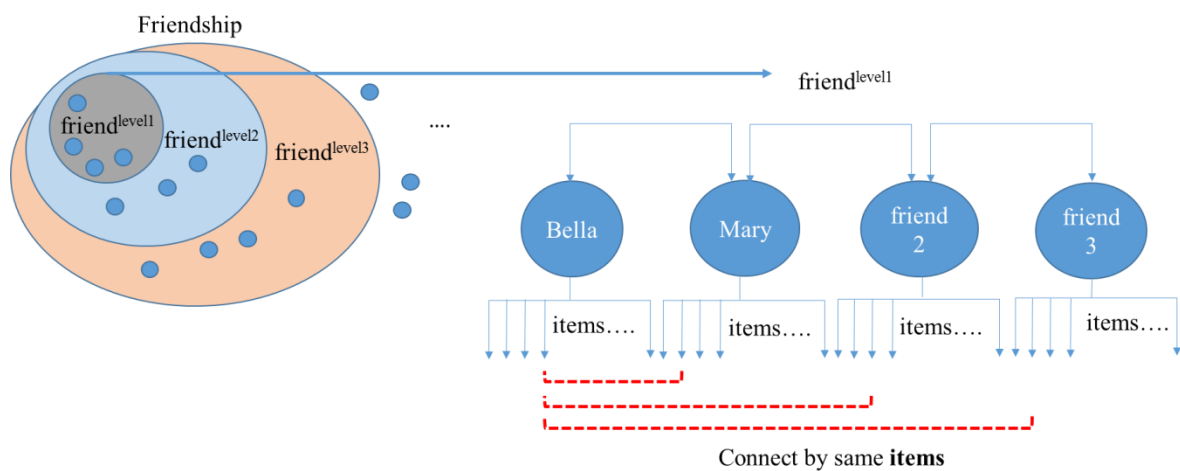


圖 3.7 Item-based 情境說明圖

資料來源：本研究整理

我們先以朋友圈 level1 的項目作介紹(圖 3.7)，此 $item^{level1}$ ，為我們第一層朋友圈與目標使用者所有評分的项目，Bella 需要透過自己過去評分的習慣對目標項目推估可能給的推薦值，(因)我們先找尋朋友同時對目標項目評分與其它項目評分的組數，推估出項目之間的關聯，此關聯的程度以 Pearson correlation coefficient 作為關聯的係數。(果)再以目標使用者對類似目標項目的項目評分紀錄對應該項目的關聯係數，推估出推薦值。

2. 模型說明：

	i_1	i_2	...	i_b	...	i_d	i_{n-1}	i_n	
u_1									
\vdots									
u_a	$r_{a,1}$	-		$r_{a,b}$		$r_{a,d}$		$r_{a,n-1}$	
\vdots									
u_c	$r_{c,1}$	$r_{c,2}$		$r_{c,b}$		$r_{c,d}$		$r_{c,n}$	
\vdots									
u_m									$m*n$

圖 3.6 item-item similarity

($i=2$ ，有兩位使用者同時給予項目 b and d 評價)

資料來源：本研究整理

在推估 Item-based Collaborative Filtering 預測值之前，必須先測量項目之間的相關係數 $Sim_{item}(b,d)$ ，圖 3.6 項目 b 與項目 d，由使用者同時給予評分此兩項目評分的使用者組數 i ，代入公式(3-3)，測量出項目 b 與項目 d 的項目相關係數。 $F_i(i_b)$ 為 level1 的使用者 c 與目標使用者 a 的相關係數的集合，公式(3-4)。

$$Sim_{item}(b,d) = \frac{\sum_i (r_{i,b} - \bar{r}_b) \times (r_{i,d} - \bar{r}_d)}{\sqrt{\sum_i (r_{i,b} - \bar{r}_b)^2 \sum_i (r_{i,d} - \bar{r}_d)^2}} \quad (3-3)$$

$$F_i(i_b) = \{i_b \mid i_b \in itemlevel^p, rank\ sim_{item}(b,d) < n'\} \quad (3-4)$$

項目 d 屬於 $F_i(i_b)$ 集合中，須與目標項目的相似度需在 top-n' 才納入估算項目推薦演算。當目標使用者有對相似度高的 d 項目評分時，將其項目 b 與 d 之間的相關係數，代入公式(3-5)，估算出推薦值。當目標使用者沒有，在 $F_i(i_b)$ 集合中，沒有相似目標項目的評分資料，則

項目 b 為 miss rating。

$$P_{item-based(a,b)} = \bar{r}_d + \frac{\sum_{sim_{item}(b,d) \in F_i(i_b)} sim_{item}(b,d)(r_{a,d} - \bar{r}_d)}{\sum_{sim_{item}(b,d) \in F_i(i_b)} sim_{item}(b,d)} \quad (3-5)$$

(範例 1)

	i_1	i_2	...	i_b	...	i_{n-2}	i_{n-1}	i_n	
u_1		4		3		3	4		
⋮									
u_a		2		?		5	5		
u_c		2		2		4	4		
⋮							3		
u_m		4		3		2	3		$m*n$

圖 3.8 Item-based Collaborative Filtering 範例

資料來源：本研究整理

表 3.4 範例 1 相關係數

資料來源：本研究整理

\bar{i}_b	2.66	$sim_{item}(b,2)$	1
\bar{i}_d	3	$sim_{item}(b,n-1)$	-0.5

$$Sim_{item}(b,d) = \frac{\sum_{i_b \in F_i(i_b)} (r_{i,b} - \bar{r}_b) \times (r_{i,d} - \bar{r}_d)}{\sqrt{\sum_{i_b \in F_i(i_b)} (r_{i,b} - \bar{r}_b)^2 \sum_{i_b \in F_i(i_b)} (r_{i,d} - \bar{r}_d)^2}} \quad (3-6)$$

$$\begin{aligned}
&= \frac{(3-2.66)(3-3) + (2-2.66)(4-3) + (3-2.66)(2-3)}{\sqrt{((3-2.66)^2 + (2-2.66)^2 + (3-2.66)^2) * ((3-3)^2 + (4-3)^2 + (2-3)^2)}} \\
&= \frac{-1}{\sqrt{0.6668 * 2}} \\
&= -0.866
\end{aligned}$$

計算目標項目之間的相關係數，分母為項目之間的共變異數和分子為項目之間的標準差進行評估，根據上式計算目標項目 b 與項目 n-2 之間的相關係數為-0.866，公式(3-6)，我們預先計算好目標項目 b 與項目 2 相關係數與目標項目 n-1 的相關係數，表 3.5。接下來代入推估目標使用者 a 對目標項目 i_b 的預測值，公式(3-7)。

$$\begin{aligned}
\hat{P}_{item-based(a,b)} &= \bar{r}_b + \frac{\sum_{sim_{item}(b,d) \in F_i(i_b)} sim_{item}(b,d)(r_{a,d} - \bar{r}_d)}{\sum_{sim_{item}(b,d) \in F_i(i_b)} sim_{item}(b,d)} \quad (3-7) \\
&= 2.66 + \frac{(-0.866) * 1.5 + (-0.5)(1.2) + 1 * (-1)}{(-0.866) + (-0.5) + (-1)} \\
&= 2.66 + \frac{(-2.8990)}{(-2.366)} \\
&= 3.8853
\end{aligned}$$

3.4.3 User-based Collaborative Filtering

1. 情境說明：

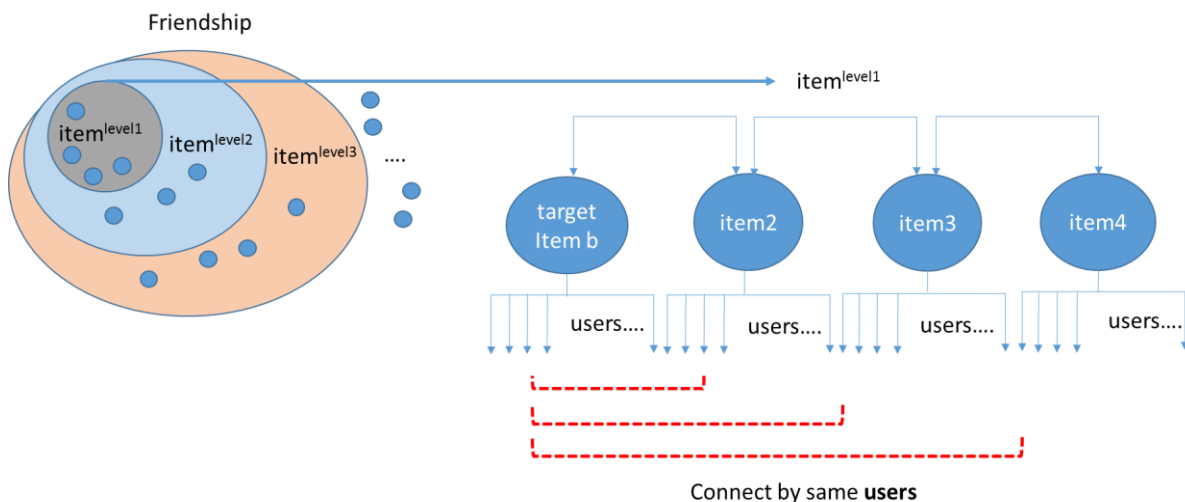


圖 3.10 User-based 情境說明圖

資料來源：本研究整理

我們先以朋友圈 level1 作介紹(圖 3.10)，透過資料我們可以找出

Bella 與其朋友的資料，再以 Bella 的朋友對目標項目評分的紀錄作為推薦的依據，(因)我們透過 Bella 與朋友之間同時對 item 評分的組數，推估人與人之間的關聯係數，以有對相同項目作評分的项目為關係連結的方式，此關聯的程度以 Pearson correlation coefficient 作為關聯的強度，(果)之後再以朋友對目標項目評分對應與 Bella 的關聯強度(相關係數)再給予推算推薦值。

2. 模型說明：

在推估 User-based Collaborative Filtering 預測值之前，必須先測量使用者之間的相關係數 $Sim_{user}(a,c)$ ，經由 Pearson correlation coefficient 的公式(3-8)來推估相關係數。圖 3.8 將使用者 a 與使用者 c 共同評分的項目 j 的組數，測量出使用者 a 與使用者 c 之間的使用者相關係數。在集合 $F_u(u_a)$ 中，我們擷取與目標使用者最相近 top-n 的使用者 c，來推估推薦值,公式(3-9)。

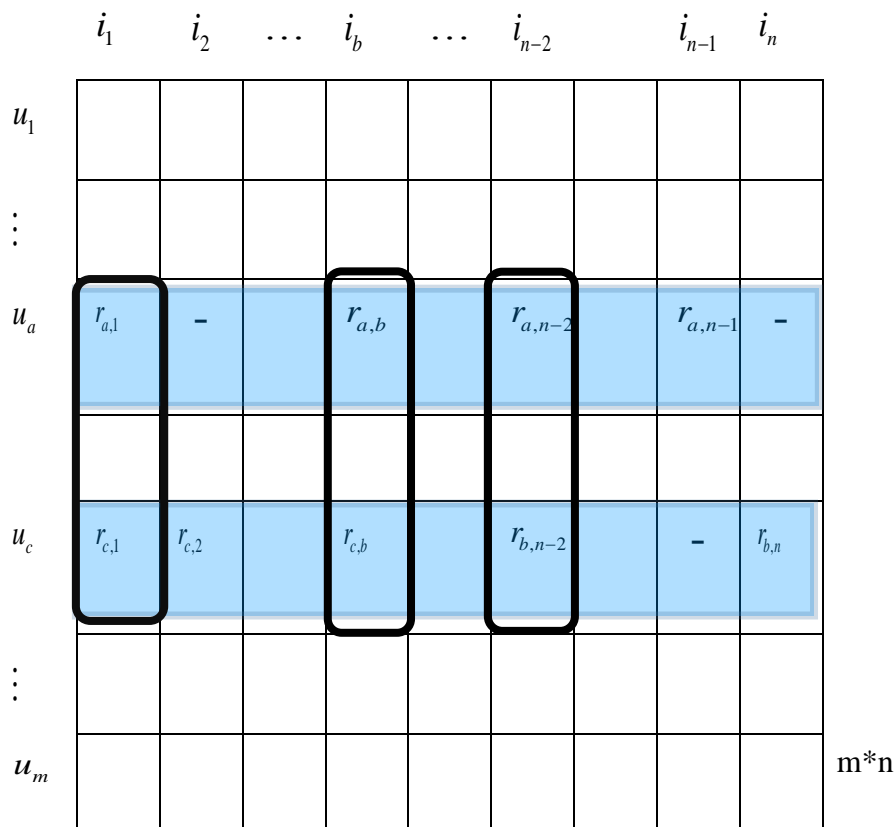


圖 3.9 user-user similarity

(j=3，使用者之間共同給予項目 1, b and n-2 評價)

資料來源：本研究整理

$$Sim_{user}(a,c) = \frac{\sum_j (r_{a,j} - \bar{r}_a) \times (r_{c,j} - \bar{r}_c)}{\sqrt{\sum_j (r_{a,j} - \bar{r}_a)^2 \times \sum_j (r_{c,j} - \bar{r}_c)^2}} \quad (3-8)$$

$$F_u(u_a) = \{u_c | u_c \in friendlevel^1, rank\ sim_{user}(a,c) < m\} \quad (3-9)$$

使用者 c 屬於 $F_u(u_a)$ 的集合，需與目標使用者的相似度在 top- m 時，納入估算推薦值中，公式(3-10)，將每位相似目標使用者的使用者 c 乘上相關係數加總後，除上所使用者相關係數。使用者 c 有評分項目 b 時，則給予使用者 a 為估算推薦值的分數，當所有使用者 c 沒有給予項目 b 評價時，則項目 b 為 miss rating。

$$P_{user-based}(a,b) = \bar{r}_a + \frac{\sum_{sim_{user}(a,c) \in F_u(u_a)} sim_{user}(a,c)(r_{c,b} - \bar{r}_c)}{\sum_{sim_{user}(a,c) \in F_u(u_a)} sim_{user}(a,c)} \quad (3-10)$$

(範例 2)

	i_1	i_2	...	i_b	...	i_{n-2}	i_{n-1}	i_n	
u_1	5	4	5	3		3			
⋮									
u_a	5	?		4		5	2	-	
u_c	5	3		3		4	-	3	
⋮									
u_m	2	3		3		2			$m * n$

圖 3.11 User-based Collaborative Filtering 範例

資料來源：本研究整理

表 3.5 範例 2 相關係數

資料來源：本研究整理

$sim_{user}(a,1)$	-0.5
$sim_{user}(a,c)$	1

$$\begin{aligned}
 Sim_{user}(a,c) &= \frac{\sum_{u_c \in F_u(u_a)} (r_{a,j} - \bar{r}_a) \times (r_{c,j} - \bar{r}_c)}{\sqrt{\sum_{u_c \in F_u(u_a)} (r_{a,j} - \bar{r}_a)^2 \times \sum_{u_c \in F_u(u_a)} (r_{c,j} - \bar{r}_c)^2}} \quad (3-11) \\
 &= \frac{(5-4.6)(5-4) + (4-4.6)(3-4) + (5-4.6)(4-4)}{\sqrt{((5-4.6)^2 + (4-4.6)^2 + (5-4.6)^2) * ((3-4)^2 + (5-4)^2 + (4-4)^2)}} \\
 &= \frac{1}{1.1547} \\
 &= 0.8660
 \end{aligned}$$

計算使用者之間的相關係數，公式(3-11)，分母為使用者之間的共變異數和分子為使用者之間的標準差進行評估，根據上式計算使用者 a 與使用者 c 之間的相關係數為 0.866，我們預先計算使用者 a 與使用者 1 相關係數，表 3.4。接下來代入推估使用者 a 對目標項目 i_2 的 User-based Collaborative Filtering 的預測值，公式(3-12)。

$$\begin{aligned}
 \hat{P}_{user-based(a,i_2)} &= \bar{r}_a + \frac{\sum_{sim_{user}(a,c) \in F_u(u_a)} sim_{user}(a,c)(r_{c,b} - \bar{r}_c)}{\sum_{sim_{user}(a,c) \in F_u(u_a)} sim_{user}(a,c)} \quad (3-12) \\
 &= 4 + \frac{sim(a,c) * (4-4) + sim(a,1) * (3-3.6) + sim(a,m) * (3-2.5)}{0.86 + 0.5 + (-1)} \\
 &= 4 + \frac{0.86 * (0) + 0.5 * (-0.6) + (-1) * (0.5)}{0.86 + 0.5 + (-1)} \\
 &= 4 + \frac{(-0.8)}{0.36} \\
 &= 4 - 2.2222 \\
 &= 1.7778
 \end{aligned}$$

3.4.4 Combining User-based and Item-based collaborative filtering

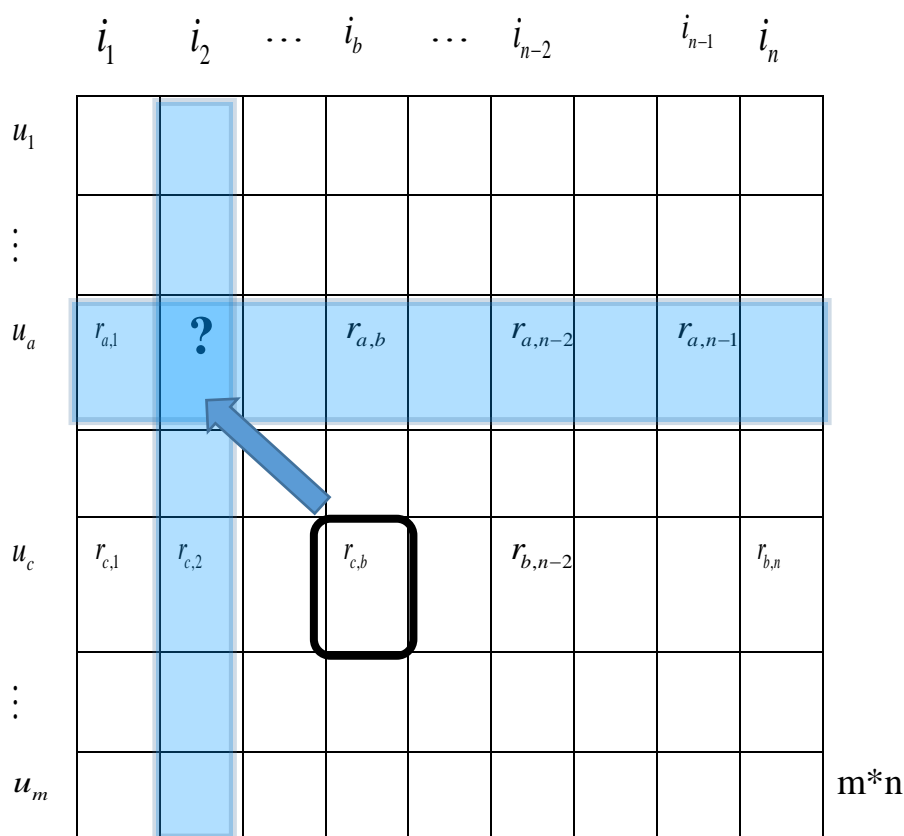


圖 3.12 Combining user-based and Item-based Collaborative Filtering

資料來源：本研究整理

透過結合 user-based 與 item-based 的概念，我們將前面提到的 $sim_{user}(a, c)$ 與 $sim_{item}(b, d)$ 結合在一起，公式(3-13)，我們利用調和平均數(Harmonic Mean) 將兩個相關係數結合(Wang *et al*, 2008)，將使用者 c 對相似項目 b 評分納入推薦估算的資料，公式(3-14)。

$$W_{user,item}(c, b) = \frac{2(sim_{user}(a, c))(sim_{item}(b, d))}{sim_{user}(a, c) + sim_{item}(b, d)} \quad (3-13)$$

$$F_{u,i}(u_a, i_b) = \left\{ \begin{array}{l} r_{c,d} \mid u_c \in F_u(u_a), i_d \in F_i(i_b), \\ c \neq a, b \neq d, rank W_{user,item}(c, d) < k \end{array} \right\} \quad (3-14)$$

在 level1 朋友圈的矩陣中，有眾多朋友對相似 b 項目的項目評分，我

們採取 $F_{u,i}(u_a, i_b)$ 前 top-k 的名單納入推薦估計， $W_{user,item}(c,b)$ 值在 level1 排行前幾名我們才納入計算，如果將全部的 $W_{user,item}(c,b)$ 都採納，則會有增加誤差的可能。

$$P_{friend(a,b)} = \frac{\sum_{w(c,b) \in F_{u,i}(u_a, i_b)} w(c,d) \times (r_{c,d} - (\bar{r}_c - \bar{r}_a) - (\bar{r}_d - \bar{r}_b))}{\sum_{w(c,b) \in F_{u,i}(u_a, i_b)} w(c,d)} \quad (3-15)$$

我們將使用者 c 在目標項目 d 的評分扣除使用者之間的差異與項目之間的差異後，乘上綜合相關係數累加，再除上前 top-n 的綜合相關係數，來計算我們朋友推估出來的推薦值，公式(3-15)。

(範例 3)

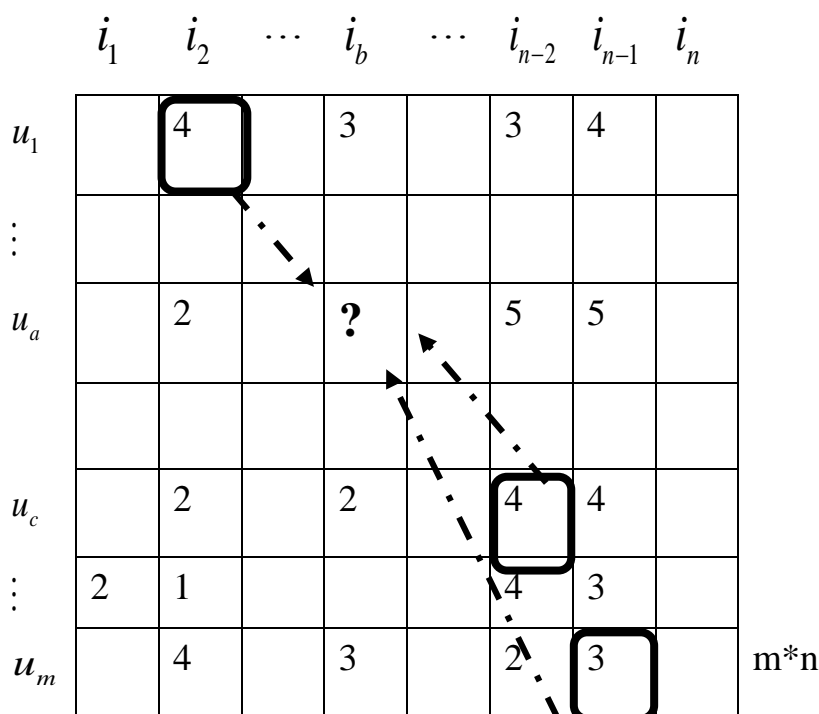


圖 3.13 Combining user-based and item-based collaborative filtering 範例

資料來源：本研究整理

表 3.6 範例 3 相關係數

$sim_{user}(1, a)$	-0.5	$sim_{item}(b, 2)$	1
$sim_{user}(a, c)$	1	$sim_{item}(b, n-2)$	-0.86603
$sim_{user}(a, m)$	-0.866	$sim_{item}(b, n-1)$	-0.5

資料來源：本研究整理

範例 3 我們以 3 位的目標使用者的朋友 1、c 與 m 為計算使用者，並以計算使用者評分的項目分別為 (u_1, i_2) 、 (u_c, i_{n-2}) 與 (u_m, i_{n-1}) 為計算項目，表 3.5 為預先計算的相關係數，我們運用調和平均數，公式(3-16)計算其綜合相關係數的值。

$$\begin{aligned} W_{user,item}(u_1, i_2) &= \frac{2(sim_{user}(1, a))(sim_{item}(b, 2))}{sim_{user}(1, a) + sim_{item}(b, 2)} \\ &= \frac{2(-0.5 \times 1)}{(-0.5) + 1} = -2 \end{aligned} \quad (3-16-1)$$

$$\begin{aligned} W_{user,item}(u_c, i_{n-2}) &= \frac{2(sim_{user}(a, c))(sim_{item}(b, n-2))}{sim_{user}(a, c) + sim_{item}(b, n-2)} \\ &= \frac{2(1 \times (-0.86603))}{1 + (-0.8663)} = \frac{-1.7321}{0.1337} = -12.955 \end{aligned} \quad (3-16-2)$$

$$\begin{aligned} W_{user,item}(u_m, i_{n-1}) &= \frac{2(sim_{user}(a, m))(sim_{item}(b, n-1))}{sim_{user}(a, m) + sim_{item}(b, n-1)} \\ &= \frac{2((-0.866) \times (-0.5))}{(-0.866) + (-0.5)} = \frac{0.866}{-1.366} = -0.634 \end{aligned} \quad (3-16-3)$$

我們選取最靠近 top-k 的使用者與項目後，減去使用者之間的平均差與項目之間的平均差，各自乘上綜合相關係數加總後，再除以相關係數的總和，公式(3-17)。

$$\begin{aligned} P_{friend(a,b)} &= \frac{\sum_{w(c,d) \in F_{u_i}(u_a, i_b)} w(c,d) \times (r_{c,d} - (\bar{r}_c - \bar{r}_a) - (\bar{r}_d - \bar{r}_b))}{\sum_{w(c,d) \in F_{u_i}(u_a, i_b)} w(c,d)} \\ &= \frac{(-2)(4 - (4 - 3.5) - (2.6 - 2.66)) + (-12.955)(2.06) + (-0.634)(0.86)}{(-2) + (-12.955) + (-0.634)} \\ &= \frac{-34.3525}{-15.589} = 2.2036 \end{aligned} \quad (3-17)$$

3.5 整合 Item-based 與 Combining User-based and Item-based collaborative filtering

本研究將 level1 朋友圈給的建議與目標使用者本身的經驗結合在一起，並加入參數 λ 的權重調整，隨著參數的變化探討其目標使用者採納 Item-based(目標使用者的意見)與 Combining user-based and Item-based collaborative filtering(朋友的意見)之間的權重為推估的內容，公式(3-18)。

$$P_{self, friend(a,b)} = (\lambda) \left(\bar{r}_a + \frac{\sum_{sim_{item}(b,d) \in F_i(i_b)} sim_{item}(b,d)(r_{a,d} - \bar{r}_d)}{\sum_{sim_{item}(b,d) \in F_i(i_b)} sim_{item}(b,d)} \right) + (1-\lambda) \left(\frac{\sum_{w_{(c,b)} \in F_{u,i}(u_a, i_b)} w_{user,item}(c,d) \times (r_{c,d} - (\bar{r}_c - \bar{r}_a) - (\bar{r}_d - \bar{r}_b))}{\sum_{w_{(c,b)} \in F_{u,i}(u_a, i_b)} w(c,d)} \right) \quad (3-18)$$

在藉由目標使用者依據自己評分經驗，我們選取項目相關係數 $sim_{(item)}(b,d)$ ，最靠近目標項目 b 的相似項目為對目標使用者的推估評分，在加上 level1 朋友圈對類似目標項目的評分狀況綜合考量， $W_{user,item}(c,d)$ 以第一層朋友圈最靠近目標使用者的目標項目來做為計算，藉此得到靠近目標使用者對目標項目的綜合評分狀況。 $P_{self, friend(a,b)}$ 整合目標使用者對目標項目的建議及朋友圈對類似項目的綜合意見。

當遇到 level1 朋友圈的相似資料無法提供給目標使用者推估時，我們將利用 level2 朋友圈所給予的評價作為參考依據。當第一層朋友圈無法對目標使用者產生 miss ratings 時，我們擴大矩陣來補強第一層朋友圈 miss ratings 的不足，我們從 level1 朋友圈擴增至 level2 朋友圈，矩陣的列位使用者(目標使用者的朋友)也會增加，欄位也會隨著增加，再進行使用者的人與人之間的相關係數分析與項目之間的相關係數分析，可參考圖 3.4 預算估計值流程流程圖，當 level2 朋友圈對目標使用者產生 miss ratings 時，我們則增加 level3 朋友圈的評分記錄到 level1 與 level2 的朋友圈矩陣中，擴增更多資料來推估目標使用者待推薦的項目，以此類推。

第四章 實證分析

本研究以社群網絡來補強推薦系統的不足，針對朋友圈對使用者的影響與使用者本身的意見之間的程度比例，提出適合的推薦評估值，以協助使用者進行決策。

4.1 實驗系統架構與實驗設計

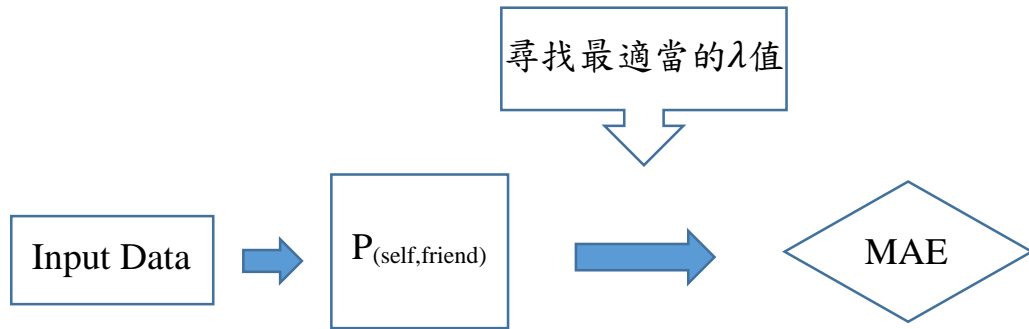


圖 4.1 實驗系統的架構

資料來源：本研究整理

我們將每一位目標使用者與其連結的朋友圈資料，經由第 3 章的方法後，我們得到預估值，再由 Mean Absolute Error(MAE)計算實際值與推測值之間的誤差值。

檢驗方法：

我們採用 MAE 來測量預測結果的品質：

- $R_{a,b}$ 為目標使用者 a 對目標項目 b 的實際評分。
- $P_{self,friend(a,b)}$ 為綜合朋友的意見推估的預測值。
- T 為提出檢測樣本的數量。
- i 為目標使用者的範圍。
- j 為目標項目提出的樣本範圍。

$$MAE = \frac{1}{T} \sum_{i,j} |P_{self,friend(a,b)} - R_{a,b}| \quad (4-1)$$

4.2 節我們針對不同的使用者 MAE 與 λ 的變化，將線性變化有具相同特性者給予分群的結果，4.3 個別使用者在適當的 λ 情況下，用不同的演算

法測量其 MAE 是否有改善。

4.2 資料分析與整理

根據我們實驗結果，發現每一位目標使用者對 P_{self} 與 P_{friend} 的採納比例因人而異，但是其中確有相似的特性存在，本小節將會展現實驗的結果且根據實驗的結果分群。隨機抽取 35 位使用者分析實驗結果且根據 MAE & λ 的線性趨勢將其分為 3 個群體，分類規則為表 4.1：

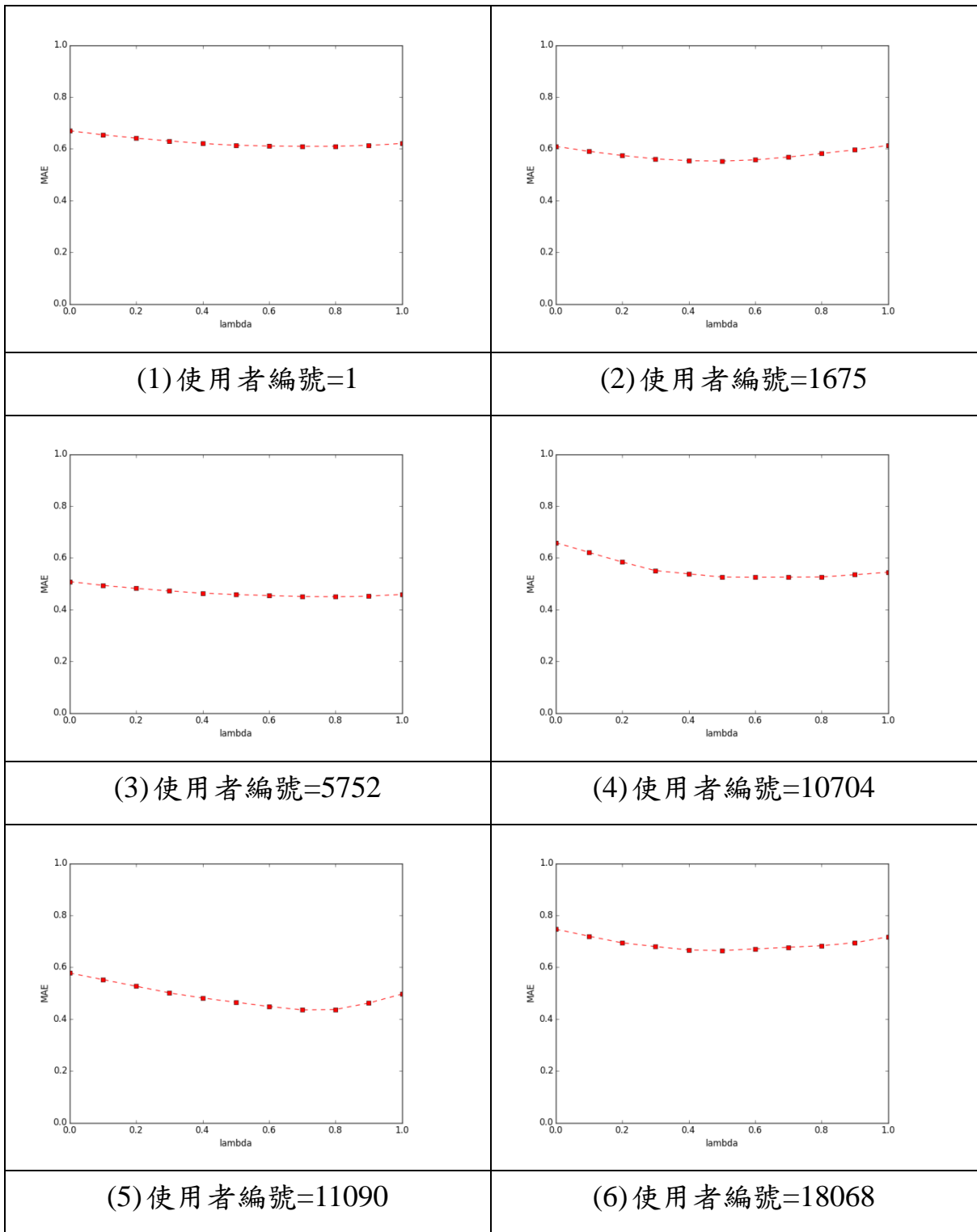
表 4.1 使用者與 λ 分類規則

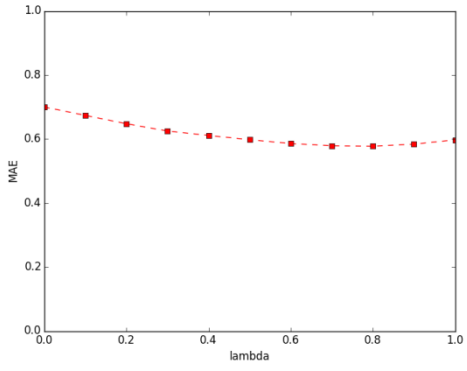
資料來源：本研究整理

λ	沒有顯著的 λ 值且 MAE 全距為 0.15 區間內	$\lambda = 0.5$	Part A	
	有顯著的 λ 值且 MAE 全距為 0.15 區間外	$\lambda > 0.5$	Part B	
		$\lambda \geq 0$ or $\lambda \leq 1$	判斷拋物線開口方向	Part B
		$\lambda < 0.5$	Part C	
		$\lambda \geq 0$ or $\lambda \leq 1$	判斷拋物線開口方向	Part C

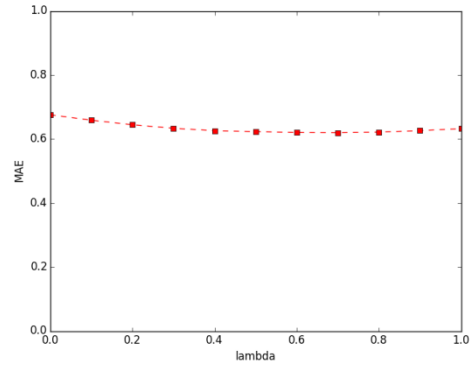
4.2.1 Part A

我們一開始以使用者是否對 λ 有明顯影響作為區分的法則，預先將樣本分為兩部分，即對 λ 有敏感度與沒有敏感度的使用者。故我們以離散資料最簡易的測度值，全距，為門檻值作為區分的指標，對控制因子 λ 的調控我們以 0.15 為界線，MAE 的之間差距如果太小，我們認為沒有顯著性的改善，如果大於 0.15 則控制因子 λ 對 MAE 具有影響力。Part A 使用者特性為對朋友的意見沒有顯著影響的人，如下圖 4.2：

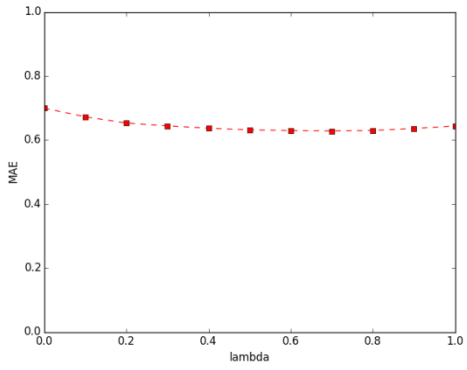




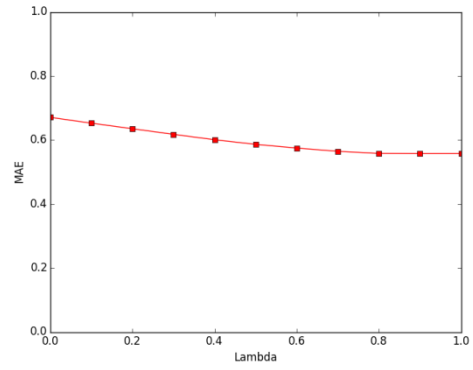
(7) 使用者編號=19577



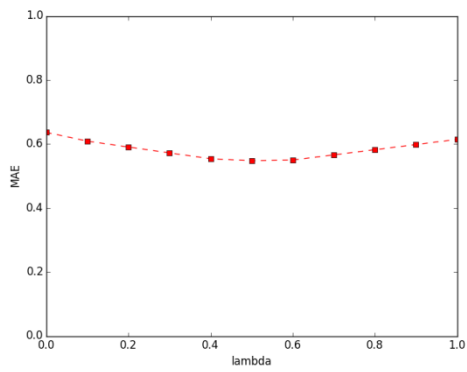
(8) 使用者編號=21489



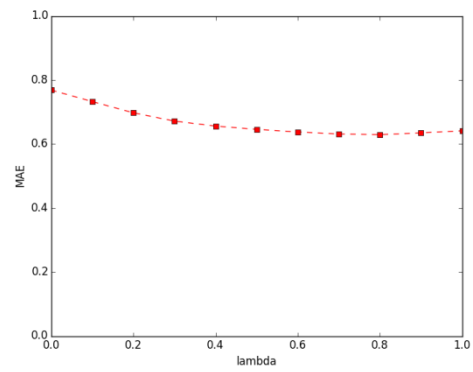
(9) 使用者編號=28982



(10) 使用者編號=28821



(11) 使用者編號=65665



(12) 使用者編號=68241

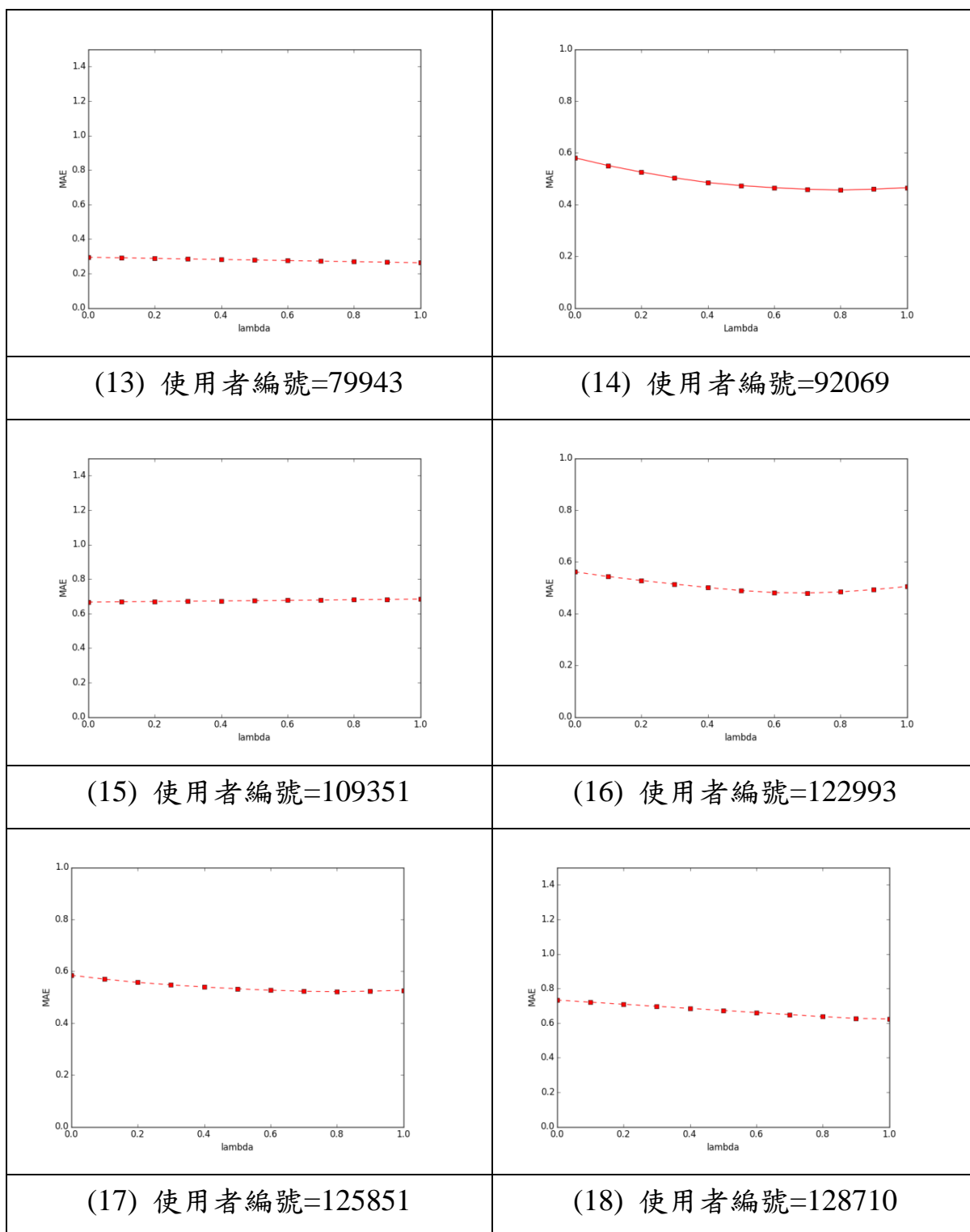


圖 4.2 18 位目標使用者 MAE & λ 的變化圖

資料來源：本研究整理

此分群的條件門檻值為自行設置的參數，不確定是否符合分群真正的條件，故我們將 Part A 作 S/N(信號雜訊比，signal-to-noise)測試此組 MAE & λ 之間的敏感度變異情況，檢驗是否具有同一特性，當調整 λ 的變化時，

觀察其誤差敏感度，如表 4.2：

表 4.2 (Part A) MAE & λ 敏感度分析

敏感度 分析	個數 (A part)	平均數 (mean)	標準差 (std)	S/N 比= -10*LOG(std ² /mean ²)
0	18	0.5516257	0.250427	6.85926737
0.1	18	0.5436033	0.1867794	9.279063246
0.2	18	0.5437754	0.1869067	9.275892421
0.3	18	0.5439508	0.1870245	9.27322306
0.4	18	0.5439508	0.1871292	9.268360384
0.5	18	0.5439508	0.1872197	9.264160002
0.6	18	0.5444368	0.1872941	9.26846687
0.7	18	0.5445576	0.1873494	9.267831937
0.8	18	0.5446439	0.1873858	9.267519103
0.9	18	0.5446983	0.1874073	9.267391065
1	18	0.5447321	0.1874206	9.267313118

資料來源：本研究整理

我們從表 4.2 得知 S/N 值皆為相同，只有在 λ 等於 0.0 時，S/N 值約為 6.86 變異性較大， λ 等於 0.1 到 1.0 之間的變化皆為相同，即變異程度皆相同，再回到圖 4.2，觀察其線性變化，此組的 MAE 與我們設定的 λ 沒有明顯的關係變化，我們可以說此 Part A 的目標使用者族群對 λ 值具同一種性質敏感度。判斷此 18 組使用者趨勢走向的情況，下圖為迴歸分析來解釋 18 組，自變項 λ 由 0 到 1 與依變項 MAE 由 0 到 1 之間的模型走向，如圖 4.3：

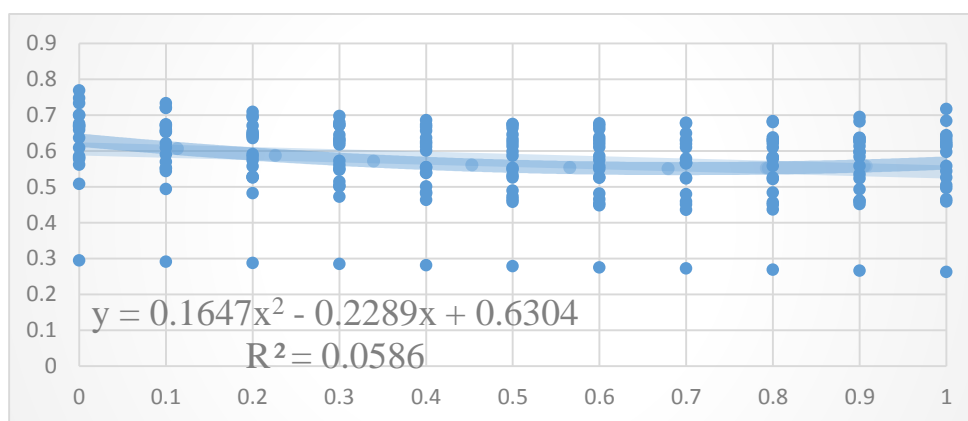


圖 4.3 (Part A) MAE & λ 迴歸分析圖

資料來源：本研究整理

經由附錄 1.1 迴歸分析結果我們的分析模型具有顯著性，我們可以解釋 Part A 的目標使用者的 MAE & λ 之間的趨勢走向具有一致性，從圖 4.3 我們看到 MAE & λ 之間沒有顯著變化，我們可以說此群體對於朋友與自身本身的推薦結果比重，沒有明顯的變化。

4.2.2 Part B

當全距 >0.15 時所篩選出來的使用者，對於 MAE 與 λ 之間有大於 0.15 的差距，且隨著 λ 值越大 MAE 也會隨之遞減，歸類在第二群體。

有顯著的 λ 值座落位置與全距為 0.15 區間外的使用者，由於我們的資料屬於離散資料，須從這些資料中找到其規則，表 4.3，於是我們繪出這些離散點的圖後，找出這群離散資料的最適當曲線方程式，稱為曲線配適法 (Curve Fitting)，找到方程式後我們再透過微分得到相關係數判斷其圖形的特性，我們找出線性的二次函數後，判定其開口位置與 λ 座落的位置：

1. $\lambda > 1$ 時， $a > 0$ 則線性為遞減趨勢：Part B。
2. $\lambda < 0$ 時， $a > 0$ 則線性為遞增趨勢：Part C。
3. $\lambda > 1$ 時， $a < 0$ 線性為遞增趨勢：Part C。
4. $\lambda < 0$ 時， $a < 0$ 線性為遞減趨勢：Part B。

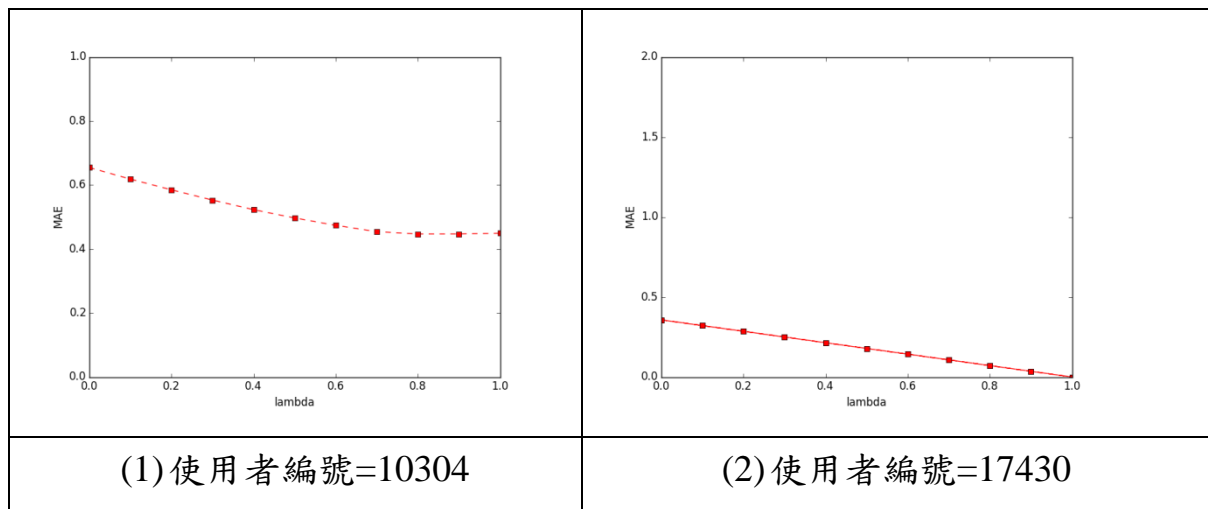
經由上述的規則，我們可以分成兩類， λ 介於 0 到 1 之間，我們以 0.5 為界線，大於 0.5 的使用者我們歸類在 Part B，也就是說接受自己意見比重較多的使用者。當 λ 在大於 1 與小於 0 時，則判斷其拋物線開口方向與 λ 座落位置，再給予分群，例：userid=17430，開口位置為向下，頂點位置 λ 座落在 -0.0609，則在 MAE & λ 的範圍，為拋物線右半曲線模式， λ 增加 MAE 有遞減的趨勢。

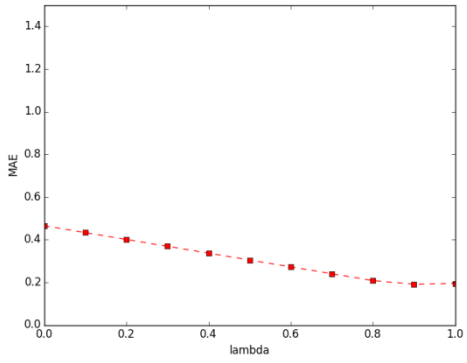
表 4.3 Part B and Part C 分類

組別	目標使用者	全距	λ 值	開口位置
B	10304	0.207582	0.978	0.4512
B	17430	0.357146	-0.0609	-5.086
B	21503	0.274196	1.979	0.198
B	22814	0.415352	1.241	0.6005
B	29656	0.180451	0.759	0.6265
B	35121	0.342768	1.644	0.3072
B	115936	0.829187	0.809	2.724
B	116350	0.192366	1.395	0.2263
B	119522	0.410117	-0.0414	-9.903
B	121011	0.351111	0.804	1.052
B	126227	0.186114	0.763	0.6764
B	103583	0.300197	1.489	0.3169
B	123785	0.97997	0.507	3.807
C	41454	0.193283	-0.04114	0.3847
C	32111	0.238279	0.003	0.4613
C	100798	0.30544	0.265	1.072
C	22054	0.55352	0.0455	-12.1654

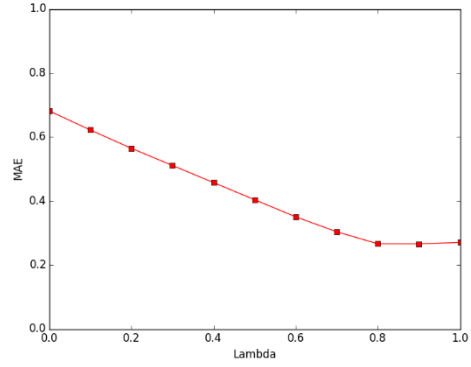
資料來源：本研究整理

雖然我們給予的 λ 值為百分比，區間應為 0 到 1，由於我們是經由 11 個資料點模擬出二次函數，再以係數判斷圖形的趨勢，故 λ 值會有超出 0 到 1 的範圍，我們再以拋物線開口的方向進一步判斷線性的趨勢後再分群。Part B 使用者特性為，採納本身意見成分較多的人，MAE 的走向為為遞減或者在適當的 λ 值大於 0.5 的使用者，其 MAE 與 λ 之間的變化圖如下：

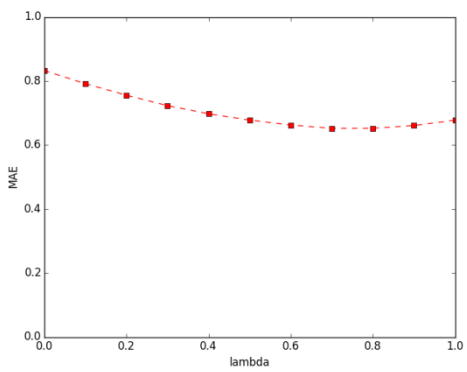




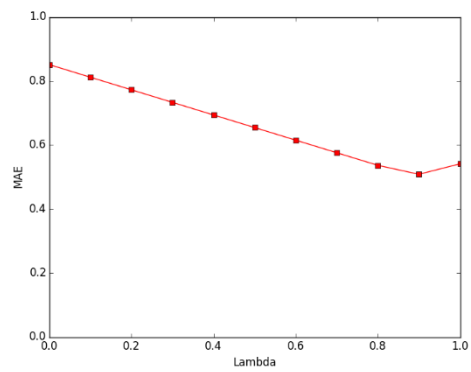
(3) 使用者編號=21503



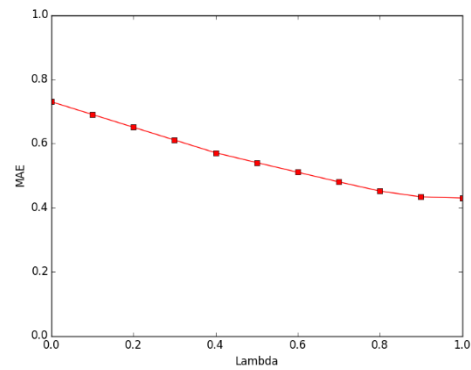
(4) 使用者編號=22814



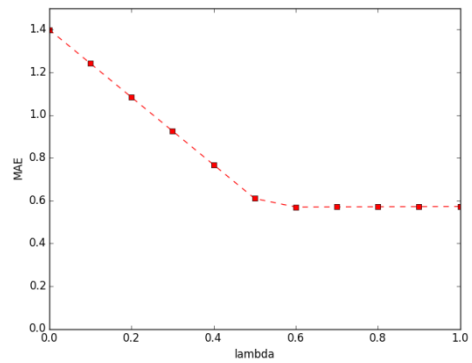
(5) 使用者編號=29656



(6) 使用者編號=35121



(7) 使用者編號=103583



(8) 使用者編號=115936

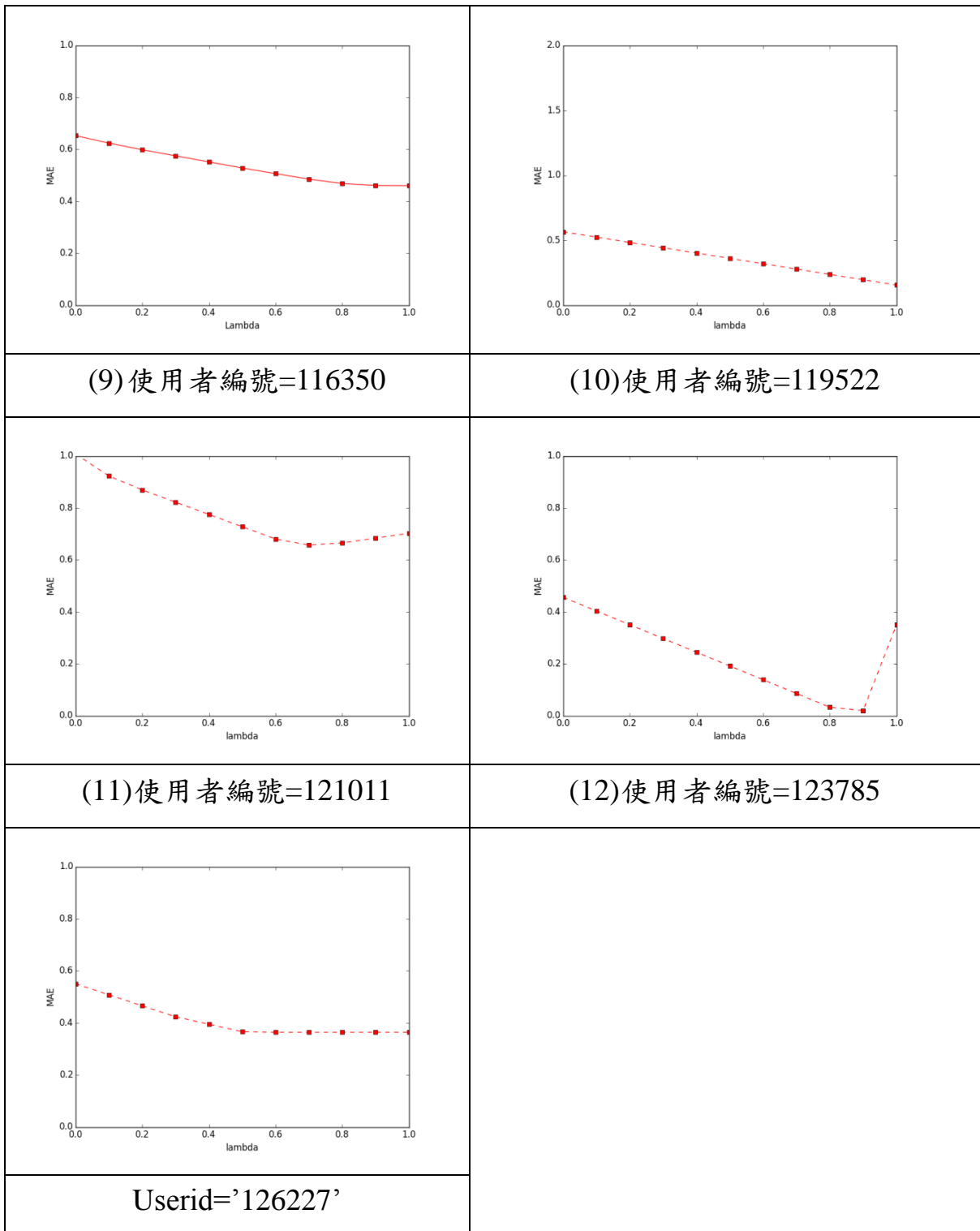


圖 4.4 13 位目標使用者 MAE & λ 的變化圖

資料來源：本研究整理

從資料觀看，可將群體分為兩種模式，明顯的 λ 值座落位置與 MAE 隨著 λ 增加而遞減。看似群體對 MAE 的變化為相似，我們在表 4.4 的 MAE & λ 敏感度分析，來分析此群體的變異。

我們從下表 4.4 得知分析結果 S/N 值皆為近似，S/N 值的變異性在 λ 等於 0.0 到 1.0 之間的變化皆為相同，即變異程度皆雷同。在圖 4.4 觀察其線性變化，此組的 MAE 與我們設定的 λ 有顯著反應，13 位使用者具有同一種特性，趨勢線為遞減的情況。

表 4.4 (Part B) MAE & λ 敏感度分析

敏感度分析	個數 (A part)	平均數 (mean)	標準差 (std)	S/N 比= -10*LOG(std ² /mean ²)
0.0	13	0.5503387	0.2587455	6.55514399
0.1	13	0.542911	0.2618697	6.332867375
0.2	13	0.5424697	0.2618819	6.325399388
0.3	13	0.5420169	0.2619651	6.315388481
0.4	13	0.5409912	0.2588072	6.404276754
0.5	13	0.5446591	0.258987	6.456936809
0.6	13	0.5476149	0.2621618	6.398115505
0.7	13	0.5472807	0.2651522	6.294297185
0.8	13	0.547006	0.2624707	6.378225253
0.9	13	0.5467539	0.2626067	6.369720245
1.0	13	0.5465017	0.2627425	6.361224766

資料來源：本研究整理

我們將 Part B 群體分為兩種類別，故我們用兩種迴歸模型解釋其趨勢走向，下圖為迴歸分析來解釋 13 組使用者的線性走向，我們以 λ 值介於 0 到 1 之間為第一組， λ 值大於 1 與小於 0 的第二組作分析：

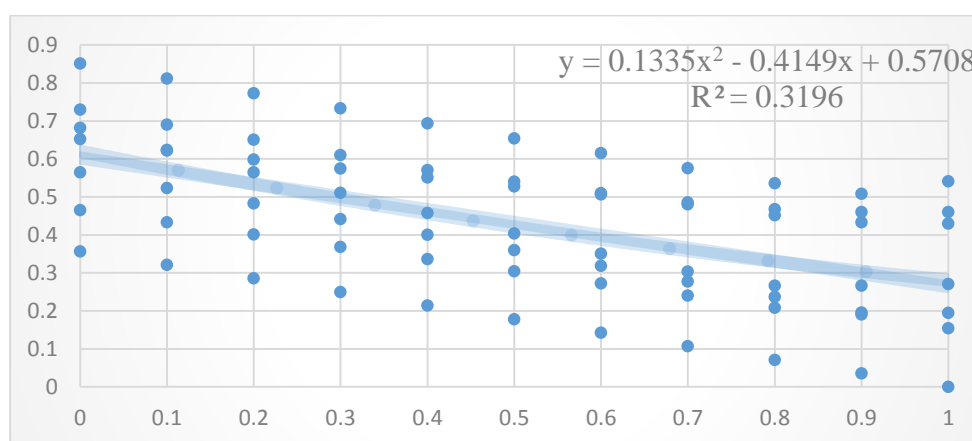


圖 4.5 (Part B，第一組) MAE & λ 迴歸分析圖

資料來源：本研究整理

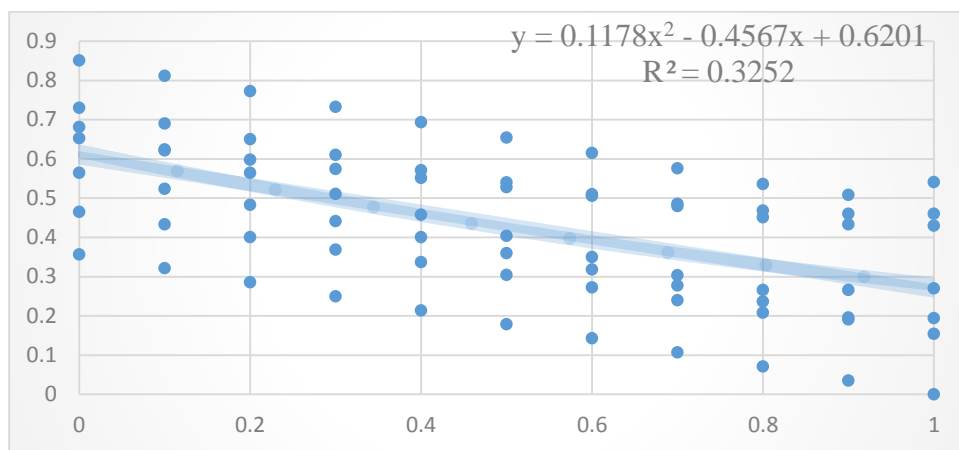


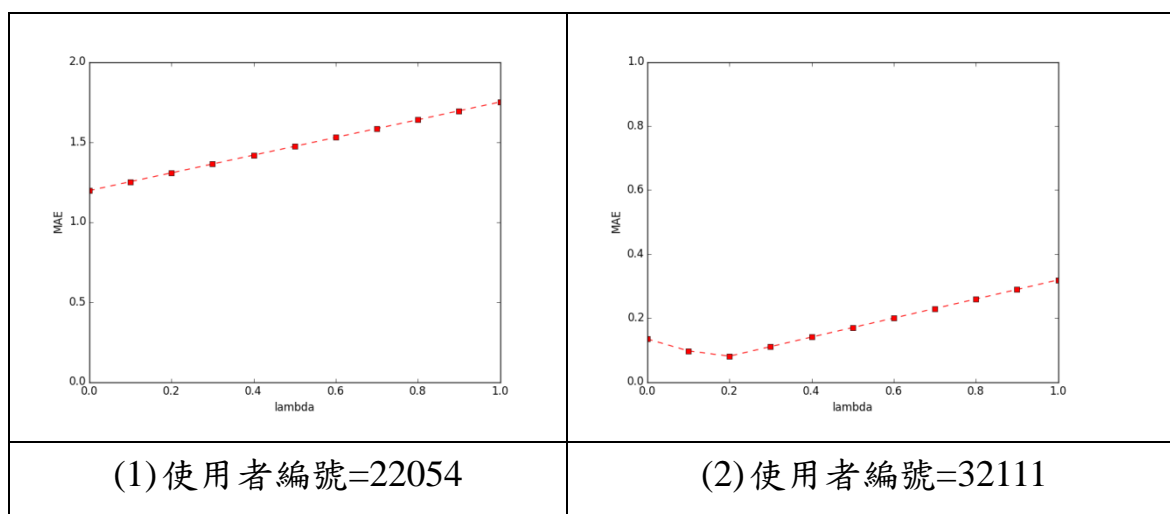
圖 4.6 (Part B，第二組) MAE & λ 迴歸分析圖

資料來源：本研究整理

由附錄 1.2 與附錄 1.3 顯示其變異數分析結果顯示第一組具有顯著性，第一組使用者特性各自有明顯的 λ 座落點，其線性迴歸趨勢如圖 4.5 判斷過程可得 MAE 隨 λ 增加而遞減的趨勢。也代表此使用者對自己本身的意見較有影響力。第二組的情況，如圖 4.6，其變異顯示不顯著，我們檢視圖 4.4，第二組的使用者 7 位，其 MAE 線性趨勢隨著 λ 增加而遞減的趨勢，此使用者對自己本身的意見較有影響力。我們可以說此 Part B 的目標使用者對自己的意見採納比重並有明顯的影響力。

4.2.3 Part C

MAE 最大值與最小值之間的差距大於 0.15，且隨著 λ 值越大 MAE 也會隨之遞增，或者 λ 小於 0.5 的使用者我們歸類在 Part C。Part C 使用者特性為，採納朋友意見成分較多的人或者易受朋友影響的人，其 MAE 與 λ 之間的變化圖如下：



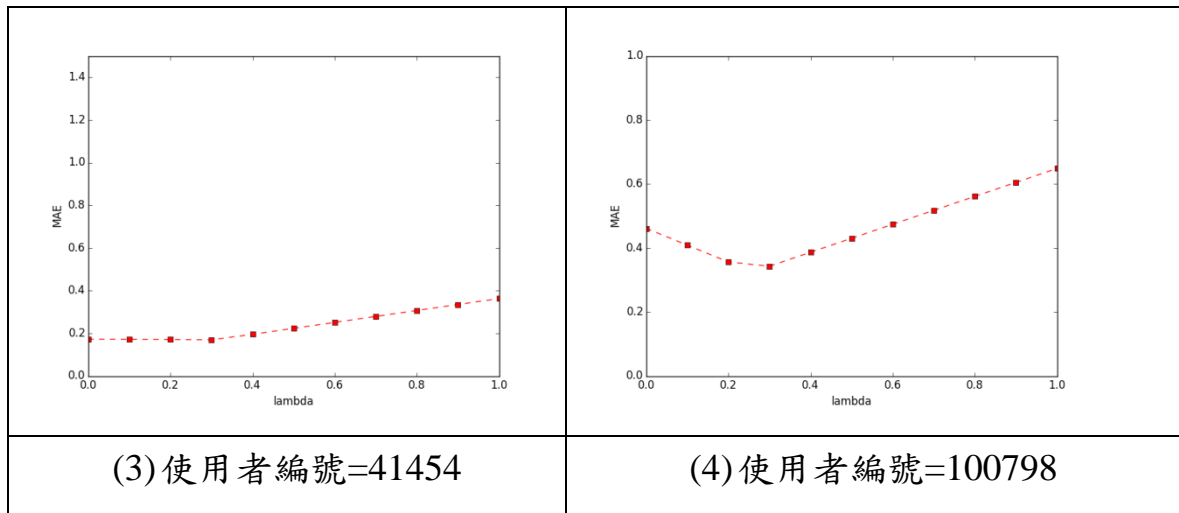


圖 4.7 4 位目標使用者 MAE & λ 的變化圖

資料來源：本研究整理

我們從表 4.5 得知 S/N 值隨著 λ 值增加有漸增的趨勢，但是變異程度大小為在可接受範圍，故我們可以判斷這些使用者對於 MAE 與 λ 的敏感度有相同的變異性。再進一步了解其群組的線性趨勢，我們用迴歸分析來說明此群體 MAE 與 λ 值特性的顯著性。

我們參見附錄 1.4 顯著性為 0.125948，發現此群體與前面的群體較高一些，由於我們的樣本數較少且 Userid='22054' 其 MAE 介於 1 到 2 之間高於其它三位使用者，而產生的偏異，但是從 Userid='22054' 的 MAE & λ 圖顯示，其為朋友意見較有影響力的使用者。從圖 4.8 我們可以看出線性趨勢走向隨 λ 增加 MAE 有遞增的情況，我們可以說 Part C 的目標使用者對朋友的意見採納比較具有明顯的影響力。

表 4.5 (Part C) MAE & λ 敏感度分析

資料來源：本研究整理

敏感度分析	個數 (A part)	平均數 (mean)	標準差 (std)	S/N 比= -10*LOG(std^2/mean^2)
0	4	0.5792905	0.3307997	4.866626666
0.1	4	0.5733685	0.3266915	4.885918671
0.2	4	0.5666356	0.3217938	4.914525558
0.3	4	0.5593871	0.3157881	4.966335567
0.4	4	0.5520505	0.3082319	5.062025822
0.5	4	0.5446258	0.2989717	5.20936142
0.6	4	0.537113	0.2878059	5.419319577

敏感度分析	個數 (A part)	平均數 (mean)	標準差 (std)	S/N 比= -10*LOG(std^2/mean^2)
0.7	4	0.529512	0.2744625	5.707855783
0.8	4	0.521823	0.258563	6.099136249
0.9	4	0.5140459	0.2395538	6.631975633
1	4	0.5061806	0.2165677	7.374238781

資料來源：本研究整理

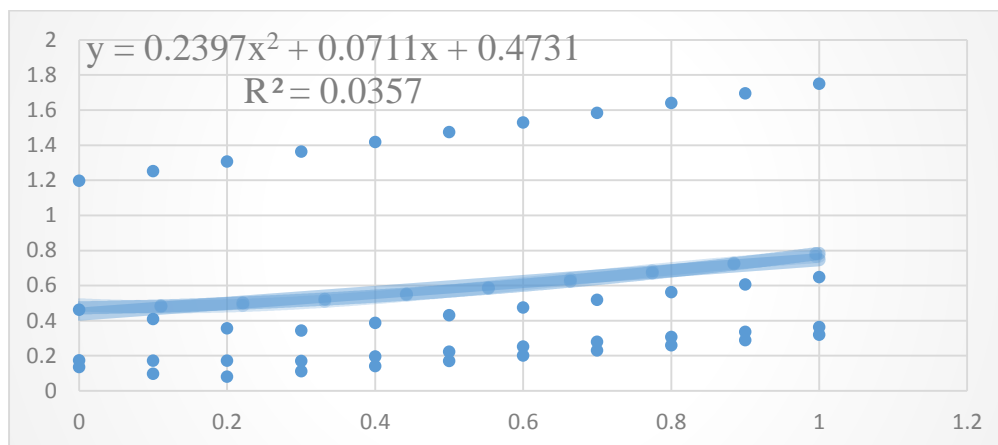


圖 4.8 (Part C) MAE & λ 迴歸分析圖

資料來源：本研究整理

4.3 實驗結果與分析

每位使用者在第一層朋友圈所得到的 λ 與 MAE 之間的變化，我們得到最適當的 λ ，再將每一位使用者取出其值後，試圖比較使用者經由適當的 λ 在我們的演算法中 $P_{(self,friend)}$ ，其 MAE 是否有得到較好的推估值。

4.3.1 比較 User-based、Item-based and Self & Friend-based

我們隨機抽取 35 位目標使用者，分別用三個不同的演算法推估推薦值並檢測 MAE，圖 4.9，發現使用者在適當的 λ 將會有最好的推薦結果，此結果，顯示朋友的意見與目標使用者本身的意見中得到良好的比重時，我們可以改善預測推薦值。

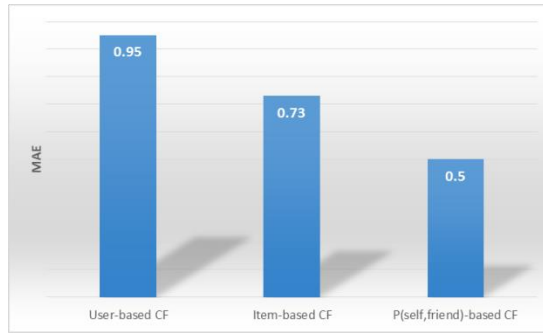


圖 4.9 user-based , item-based & P(self & friend)三個演算法比較圖

資料來源：本研究整理

4.3.2 MAE & Level

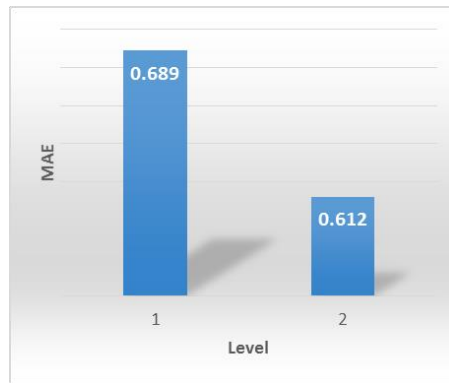


圖 4.10 MAE and Level

資料來源：本研究整理

在巨量資料庫中，我們運用較少的朋友圈資訊計算推薦值，當第一層朋友圈無法有效地給予推薦值時，我們將矩陣擴大至第二層朋友圈(放大資料量，3.4 節說明)，將在第二層預估出來的推薦值測量其誤差值。

實驗顯示在第 2 層時，MAE 的誤差值將會逐漸下降，本研究目前測試到第 2 層，當矩陣擴大我們得到更多的資料進行推估，實驗終點為須擴大到多少朋友層的時候，我們的 MAE 將會到停滯點，也是我們要證實在巨量資料庫中，選擇部分朋友圈的資料作為推估值的資料，我們選擇多少層朋友圈時，就足夠運用此資料量給予推薦值且 MAE 在我們可以接受的範圍。

4.3.3 Missing ratio & Level

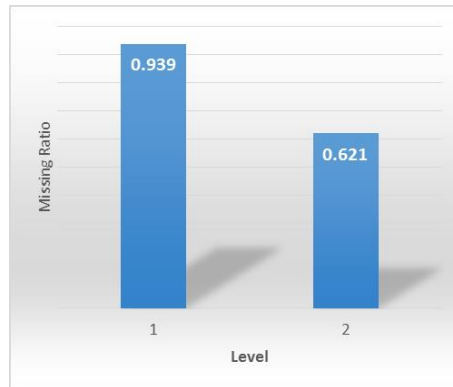


圖 4.11 Missing ratio & Level

資料來源：本研究整理

由上圖顯示 missing ratio 由 93% 降至 62%，我們擴增矩陣來推估推薦值，成功降低了 missing ratio，missing ratio 為每一層朋友圈未能給予目標使用者所評分的目標項目除上目標使用者已評分的數量之百分比，第一層朋友圈完成 7% 目標使用者待推薦值的項目，而第二層朋友圈完成 31% 目標使用者待推薦值的項目，當擴增的朋友圈越大越能降低 missing ratio，但是相對也要考量系統運算的負擔能力是否可以接受。

4.4 小結

我們將 3 個群組分群後找到其最佳的 λ 值，35 個抽樣中，MAE & λ 的線性趨勢圖將其分群，有 18 位對於 λ 是沒有敏感度的使用者、13 位對於自己本身意見較有影響力的使用者與 4 位對於朋友的意見較有影響力的使用者。有 51% 的 λ 值對於使用者是沒有影響力的，近一半的使用者受 λ 值影響，37% 是對自己本身意見影響較顯著的與 12% 是對朋友給的意見影響力較顯著的。

第五章結論與建議

我們提出朋友的意見與使用者本身的意見綜合一起討論的方法，抽樣的 35 位使用者，我們歸類為 3 個群體，分別為對 λ 值沒有影響的使用者、受 λ 影響且朋友的意見影響力較大、受 λ 影響且本身的意見影響較大，我們從這些使用者找出最佳 λ 值，計算其預測值，並從不同的演算法中測試，我們的方法確實有較佳的推薦效果。

5.1 結論

我們藉由 social influence 來改善我們推薦的能力，藉由人與人之間的連結，將有效的資料給予估算。藉由 social network 的補強，我們直接尋找與我們最相似與品味最雷同的人來給予推薦值。當第一層朋友圈無法足夠的資料給予推薦值時，則有 miss ratings 的問題，經由 friendship 的關係，我們擴大資料矩陣，尋找第二層朋友圈來補強資料的不足，也幫助我們降低 missing ratio。隨著朋友圈的擴大，資料量也較充足，我們可以得到更好的 MAE，這當中也伴隨著運算矩陣的增大，這時就要考量系統運算的資源是否足夠的處理速度。對於本研究的貢獻如下：

1. 在巨量資料庫中，透過朋友圈的資料來推估推薦值，且提出質與量的觀點，從小資料到擴增大資料的過程，有助於減少運算的效率與蒐集資料的時間。
2. 提出朋友的意見與使用者本身意見之間的比重，在正確的比重下，可有效的改善推薦能力。
3. 建議適用於社群網路資料的推薦機制。
4. 補強推薦系統的問題：資料稀疏。

5.2 未來發展方向

1. 門檻值的設置

透過朋友圈擴充的矩陣，我們設置每一層朋友圈的朋友須達到與目標使用者有共同的評分項目 25 項為門檻值，再納入朋友圈矩陣計算。門檻值的設置是將所納入的矩陣有良好的大小提供系統做運算，太小，在該矩陣中進入的朋友數量過多進而使系統運算效率減低；太大，會

有部分朋友無法進入矩陣中。我們建議可以進一步探討擴充下一層朋友圈的同時，朋友進入矩陣的門檻值設置。

2. 少部分群體為受朋友影響的族群

建議在使用本推薦系統的前提下，目標使用者皆對朋友的意見有敏感度，在 4.7 圖我們可以看出 35 位隨機抽樣使用者中，有 12% 是對朋友的意見較明顯的反應者，所佔的比率較小。我們建議在使用本研究的推薦方法時，所參與的目標使用者事先調查是否對朋友意見有敏感度，可第一時間過濾無需(使用者對朋友的意見沒有敏感度)計算的使用者。

3. 朋友關係的建立在本研究皆有相同喜好與品味的群體

我們對目標使用者的朋友圈未進一步的作分群與關聯式的探索，在本研究同一層朋友圈我們假設朋友皆有相同的喜好與群體，在實際生活中，朋友圈雖為朋友關係，但是確有不同的喜好與品味。建議可以在蒐集使用者的資料(例，已是朋友關係且加入相同的喜好社群或者有實際朋友關係身分的資料)或者是發現其關聯規則(例，Rough Set)，將同一層朋友圈的使用者做進一步不同的處理預測值給目標使用者。

4. 本研究受限於時間與資源下，仍有不足之處。本研究主要限制如下：

(1) 時間

由於研究時間有限，對於不同層次的 MAE & Level 與 missing ratio 我們希望能夠分析至第六層，以驗證根據小世界理論(Milgram,1967)理論，透過平均 6 個朋友的聯繫便可連接到你要尋找到的人。也就是可以解決 missing ratio 降至最低，是否在第六個朋友時，目標使用者可以找到所有的 target items。

(2) 是否真正具有購買目標

本研究情境是給予目標使用者預估推薦值，然而是否真正的購買行為無法在本研究中得知，對一個真正的資源決策系統真正好的推薦系統除了要引起使用者的興趣，最重要是經由推薦後產生真正購買的行為。

5. λ 值的計算與發展

我們運用使用者評分資料與朋友圈的資料作為演算的依據，我們透過朋友圈的連結來補強資料稀疏的問題，朋友的資料在本研究中只有 0 與 1 的關係，在真實生活中，有不同維度的觀點(例如，身分、行業、年齡等。)或參考 (Elaboration Likelihood Model, ELM)，談到人會不會採取別人的建議，會牽扯到中央路徑與邊陲路徑，中央路徑是邏輯思考，邊陲路徑是情感上的思考，會牽扯到專業度與重要度，專業度靠自我邏輯思考，相關我們對這件事情的專業度與重要度(Lee, K. C. and Kwon, S., 2008)，基本上本研究模型在有限的資料中，需驗證人格特質上的判別是很難回答的問題。建議在未來研究上，可以用此理論來解釋 λ 的定義，且會取決於此推薦的情境(例，餐廳、音樂與電影等。)而每個人對不同情境的專業度與重要度，有不同程度的推薦方式，如此多維度的觀點是否可以改善 51% 對 λ 值沒有影響的使用者，則是可繼續研究探討的。

6. 推薦值檢測方式

在對於檢測推薦值的精準度，以工程結構的推薦與真實生活中的滿意度的測試結合，利用 validity 的方式測試是否準確，如此一來更讓人相信我們推論的推薦值是準的。(例：準確率要多高的情況下，可以被信以為真。)後續可為電子商物的公司評估推薦系統的機制，推薦準確率的改善中間的成本效益是否適宜的評估測試。

參考文獻

中文文獻

1. 黃河銓, 李奕縉, 黃昱凱(2012)。整合信任網路與回饋機制之個人化餐廳推薦系統。 *資訊與管理科學*, 5, 65-78。
2. 黃河銓, 王群元(2010)。以社會網路與電影本體為架構之電影推薦系統。 *資訊與管理科學*, 12, 595-620。
3. 黃純敏, 林重佑, 黃進瑞 (2013)。結合學習向量量化與協同過濾之交換混合式過濾電影推薦架構。 *資訊管理學報*, 20, 423-448。

英文文獻

1. Arazy, O., Kumar, N., and Shapira, B. (2009). Improving Social Recommender Systems. *IT Professional*, 11(4), 38-44.
2. Bobadilla, J., Ortega, F., Hernando, A., and Gutiérrez, A. (2013). Recommender systems survey. *Knowledge-Based Systems*, 46, 109-132.
3. Breese, J. S., Heckerman, D., and Kadie, C. (1998). Empirical Analysis of Predictive Algorithms for Collaborative Filtering. In *Proceedings of the Fourteenth conference on Uncertainty in artificial intelligence*.
4. Das, A. S., Datar, M., Garg, A., & Rajaram, S. (2007). Google News Personalization: Scalable Online Collaborative Filtering. *Proceedings of the 16th international conference on World Wide Web*, P271-280. doi:10.1145/1242572.1242610
5. Domingos, P. and Richardson, M. (2001). Mining the network value of customers. *Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining*, 59-66. doi:10.1145/502512.502525
6. Golbeck, J. (2005). Personalizing applications through integration of inferred trust values in semantic web-based social networks. In *Proceedings of the semantic network analysis workshop at the 4th international semantic web conference*.
7. Gupta, S., and Lehmann, D. R. (2003). Customers as Assets. *Journal of Interactive Marketing*, 17(1), 9-24.
8. Häubl, G., and Trifts, V. (2003). Consumer Decision Making in Online Shopping Environments: The Effects of Interactive Decision Aids. *Marketing Science*, 19(1), 4-21.
9. He, J., and Chu, W. W. (2010). A Social Network-Based Recommender System. In Nasrullah, M., Xu, J. J., Hicks, D. L., and Chen, H. (Eds.), *Data Mining for Social Network Data (Vol. 12, 47-74)*. New York, NY: Springer Press.
10. Koren, Y., Bell, R., and Volinsky, C., (2009). Matrix Factorization Techniques for Recommender Systems. *IEEE Computer Society*, 42(8), 30-37.
11. Lang, K., (1995). NewsWeeder: Learning to Filter Netnews. *Proceedings of the 12th*

- International Machine Learning Conference*, 331-339. doi:10.1016/B978-1-55860-377-6.50048-7
12. Lee, K. C. and Kwon, S. (2008). Online Shopping Recommendation Mechanism and its Influence on Consumer Decisions and Behaviors: A Causal Map Approach. *Expert Systems with Applications*, 35 (4), 1567–1574.
 13. Linden, G., Smith, B., and York, J. (2003). Amazon Recommendations Item-to-Item Collaborative Filtering. *IEEE internet computing*, 7(1), 76-80.
 14. Ma, H., Zhou, D., Liu, C., Lyu, M. R., and King, I. (2011) Recommender Systems with Social Regularization. *Proceedings of the fourth ACM international conference on Web search and data mining*, 287-296. doi:10.1145/1935826.1935877
 15. Massa, P. and Bhattacharjee, B. (2004). Using Trust in Recommender Systems: An Experimental Analysis. *Proceedings of 2nd International Conference on Trust Managment*. 221-235. doi: 10.1007/978-3-540-24747-0_17
 16. Milgram, S. (1967). The Small World Problem. *Psychology Today*, 2, 425-443.
 17. Mayer R. C., Davis J. H. and Schoorman, F. D. (1995). An Integrative Model of Organizational Trust. [Review of the book *The Academy of Management*, by J. S. Brown & P. Duguid] doi:10.5465/AMR.1995.9508080335
 18. O'Donovan, J. and Smyth, B. (2005). Trust in recommender systems. *Proceedings of the 10th international conference on Intelligent user interfaces*, 167-174. doi: 10.1145/1040830.1040870
 19. Resnick, P., Iacovou, N., Suchak, M., Bergstrom, P., and Riedl, J. (1994). GroupLens: An Open Architecture for Collaborative Filtering of Netnews. *Proceedings of the 1994 ACM conference on Computer supported cooperative work*, 175-186. doi:10.1145/192844.192905.
 20. Sun, Z., Han, L., Huang, W., Wang, X., Zeng, X., Wang, M., and Yan, H., (2015). Recommender systems based on social networks. *Journal of Systems and Software*, 99, 109-119.
 21. Sarwar, B., Karypis, G., Konstan, J., and Riedl, J. (2001). Item-based Collaborative Filtering Recommendation Algorithms. *Proceedings of the 10th international conference on World Wide Web*, 285-295. doi:10.1145/371920.372071
 22. Schafer, J. B., Konstan, J., and Riedl, J. (1999). Recommender Systems in E-Commerce. *Proceedings of the 1st ACM conference on Electronic commerce*, 159-166. doi:10.1145/336992.337035
 23. Wang, J., Vries, A. D., and Reinders M. J. T. (2008). Unified Relevance Models for Rating Prediction in Collaborative Filtering. *ACM Transactions on Information Systems*, 26(3), 1-42.
 24. Wang, J. C. and Chiu, C. C. (2008). Recommending Trusted Online Auction Sellers Using Social Network Analysis. *Expert Systems with Applications*, 34(3), 1666-1679.

附錄一

附錄 1.1 (Part A)迴歸分析結果

迴歸統計	
R 的倍數	0.190915
R 平方	0.036449
調整的 R 平方	0.031507
標準誤	0.101854
觀察值個數	197

ANOVA

	自由度	SS	MS	F	顯著值
迴歸	1	0.076523	0.076523303	7.376349	0.007202
殘差	195	2.022958	0.010374143		
總和	196	2.099481			

	係數	標準誤	t 統計	P-值	下限 95%	上限 95%	下限 95.0%	上限 95.0%
截距	0.604579	0.013663	44.24895371	1.2E-103	0.577633	0.631526	0.577633	0.631526
0	-0.06257	0.023037	-2.715943568	0.007202	-0.108	-0.01713	-0.108	-0.01713

附錄 1.2 (Part B, 第一組)迴歸分析結果

迴歸統計	
R 的倍數	0.355602
R 平方	0.126452
調整的 R 平方	0.112803
標準誤	0.245361
觀察值個數	66

ANOVA

	自由度	SS	MS	F	顯著值
迴歸	1	0.557739	0.557739	9.264475	0.003388
殘差	64	3.852921	0.060202		
總和	65	4.41066			

	係數	標準誤	t 統計	P-值	下限 95%	上限 95%	下限 95.0%	上限 95.0%
截距	0.730722	0.056502	12.93258	1.63E-19	0.617846	0.843599	0.617846	0.843599
X 變數 1	-0.2907	0.095507	-3.04376	0.003388	-0.4815	-0.0999	-0.4815	-0.0999

附錄 1.3 (Part B, 第二組)迴歸分析結果

迴歸統計	
R 的倍數	0.56756
R 平方	0.322124
調整的 R 平方	0.313086
標準誤	0.157538
觀察值個數	77

ANOVA

	自由度	SS	MS	F	顯著值
迴歸	1	0.884512	0.884512	35.6397	7.34E-08
殘差	75	1.861361	0.024818		
總和	76	2.745873			

	係數	標準誤	t 統計	P-值	下限 95%	上限 95%	下限 95.0%	上限 95.0%
截距	0.602459	0.033587	17.93718	7.55E-29	0.53555	0.669368	0.53555	0.669368
X 變數 1	-0.33893	0.056773	-5.9699	7.34E-08	-0.45202	-0.22583	-0.45202	-0.22583

附錄 1.4 (Part C)迴歸分析結果

迴歸統計	
R 的倍數	0.23701
R 平方	0.056174
調整的 R 平方	0.033154
標準誤	0.527167
觀察值個數	43

ANOVA

	自由度	SS	MS	F	顯著值
迴歸	1	0.678145	0.678145	2.4402	0.125948
殘差	41	11.39412	0.277905		
總和	42	12.07227			

	係數	標準誤	t 統計	P-值	下限 95%	上限 95%	下限 95.0%	上限 95.0%
截距	0.371468	0.154973	2.39699	0.021172	0.058494	0.684441	0.058494	0.684441
	0	0.404522	1.562114	0.125948	-0.11845	0.927497	-0.11845	0.927497

口試問題整理

➤ 薛文蔚老師

1. 問：已經假設朋友是購物習慣相同的人，為何會有 user-based 且又再計算其相似度呢？

學生：本論文主要是考慮從雜訊較小的區塊開始建立模型，倘訊息不足，再適度加入訊息與雜訊的作為，以朋友圈的朋友為主要處理對象範圍，前面假設是由於部分文獻(He and Chu, 2010; Sun, *et al.*, 2015)談即雖為同一層朋友圈的朋友，但是仍有興趣不相同的情況，而造成在推薦值推算時，運算的相關係數會有所不同。本研究假設朋友圈為相同的情況，但是在運算推薦值時，過程是以使用者之間的相關係數為主要權重的計算。

2. 問：最早提到的 trust，是如何定義它？

學生：已在 P16 補充，請參。

3. 問：此定義與前述文獻的 trust(transform)的關係是什麼？

學生：本研究以(O'Donovan, J. and Smyth, B., 2005)數據上的信任來作本研究信任值門檻的概念，與前述文獻(Golbeck, J., 2005)稍有不同，朋友間真實給予信任值，且透過 Tidal Trust(transform)演算可推導出不同階層朋友各別與目標使用者的之間的信任值。

4. 問：因果關係圖

學生：已在 P22 and P26 補充，請參。

➤ 盧希鵬老師

1. 問：Golbeck, J. (2005)此篇是融合相似度與信任值計算相關係數，兩者為相關並非相等，或許在我的論文中使用”trust”較不恰當。

學生：詳細補充說明已在 P16 補充，請參。

2. 問：準確率要多高的情況下，可以被信以為真。

學生：本研究以 MAE 的方式檢測與目標使用者真實評分值與本研究推估出來的推薦值之間的絕對值平均差距為主要準確度衡量方式，對於目標使用者是會不會採納推薦而給予購買行為，為本研究不探討範圍，感謝老師

提供學生未來研究的建議。學生已補充在 P51，請參。

3. 問：研究問題”朋友在推薦系統中，所佔的比例為多少時，推薦能力最適當”，是否可以回學生，如果無法則要更改寫法。

學生：本研究將 P_{friend} 與 P_{self} 之間以 λ 作為可變動參數，尋找出目標使用者在最適當的 λ 值時有最低誤差，經由盧老師指導，根據 ELM(Elaboration Likelihood Model) 我們需要更多的資訊才可以判斷目標使用者在朋友與自己的意見之間的關係，也就是說，此研究對於人格特質上的資料些許不足，未來如果需要更精準的 λ 值時，建議參考 ELM 的方法一起探討。

➤ 謝宛霖老師

1. 問：level2 的朋友是否為 Bella 的第一層朋友，只是在不同群的第一層朋友中朋友？

學生：第二層的朋友是在情境中 Mary 的朋友（請參 p20），也就是 Bella 不認識 level2 的朋友，須透過第一層朋友連結取得 level2 的資料，由於在問題假設第 2 項時，將朋友設為同相同喜好與品味相似的群體，故各自的 level 中不給予再分群。

2. 問：user-based 的關連是以 item-based 為何還需要 item-based？因為聽起來是用 item 的格式計算使用者相似度。

學生：已在 P23 and P26 補充，請參。

➤ 王偉華老師

1. 問：信任的名詞與相似度的名詞如何區分

學生：本研究信任名詞已有詳細補充說明已在 P16 補充，請參；關於相似度的名詞解釋，為計算物件之間的相關係數或者關聯性；本研究 Pearson correlation 為計算相關係數的方法，分母為兩組物件項目之間的共變異數和分子為項目之間的標準差進行評估。