

# 國立交通大學

管理學院（資訊管理學程）碩士班

## 碩士論文

結合貝氏網路與激勵理論之推薦機制  
- 電影推薦系統設計

**A Recommendation Mechanism Combined with  
Bayesian Networks and Incentive Theory  
- A Movie Recommend System Design**

研究生：謝金育  
指導教授：李永銘 博士

中華民國 102 年 6 月

結合貝氏網路與激勵理論之推薦機制  
- 電影推薦系統設計

A Recommendation Mechanism Combined with  
Bayesian Networks and Incentive Theory  
- A Movie Recommend System Design

研究生：謝金育

Student:Chin-Yu Hsieh

指導教授：李永銘

Advisor: Dr. Yung-Ming Li



中華民國一百零二年六月

# 結合貝氏網路與激勵理論之推薦機制

## - 電影推薦系統設計

研究生：謝金育

指導教授：李永銘 博士

國立交通大學管理學院（資訊管理學程）碩士班

### 摘要

推薦系統被廣泛應用於網路上，用以幫助使用者快速找到適合或是感興趣的產品，目前已發展出許多的推薦技術，例如內容導向式(Content-based Approach)、協同過濾式(Collaborative Filtering Approach)、及混合式(Hybrid Approach)等推薦方法。雖然推薦技術已日趨成熟，但其仍存在著一些問題，例如在使用者資訊不足的前提下，推薦系統無法正確的提供推薦資訊。

本論文提出結合貝氏網路與激勵理論(Incentive theory)的推薦機制，在使用者資訊不足的情況下仍然可以產生高度正確率的推薦。透過尋找出使用者輪廓檔(user profile)內資訊之關聯規則，將其視為一個訊息源用以擴大使用者輪廓檔，從而避免使用輪廓檔資訊不足的問題。而在新項目的評比方面，透過激勵理論(Incentive theory)，在符合個人理性(individual rationality)條件與誘因相容(incentive compatible)條件下鼓勵使用者對於新項目分享。實驗證明，透過拓展使用者輪廓檔與鼓勵使用者評比是可以幫助系統在使用者資訊不足的情況下提升系統的推薦準確率。

關鍵字：關聯規則，貝氏網路，推薦系統

# **A Recommendation Mechanism Combined with Bayesian Networks and Incentive Theory - A Recommend System Design**

Student : Chin-Yu Hsieh

Advisor : Dr. Yung-Ming Li

Master Program of Institute of Information Management  
College of Management  
National Chiao Tung University

## **Abstract**

The recommendation systems are widely used on the network to help users quickly find suitable or interested products. In this area, many recommendation techniques, such as Content-based Approach, Collaborative Filtering Approach, and Hybrid Approach, have been developed. Although the recommendation technology has become mature, there are still some problems. Recommendation systems cannot provide the correct information if we don't have enough user information. The precision of the recommendation system will be increased dramatically, if a system gets more user potential information.

In this study, we attempt to propose a recommendation mechanism design combined with Bayesian networks and incentive theory. In lack of user profile information, our recommendation mechanism still has high precision. This mechanism approach uses Bayesian network to generate association rules from user profiles, a source of information used to expand the user profile, and avoid the problem of user profile shortage. In the new items problem, based on the theory of incentives, we propose a mechanism for encouraging data sharing. This encouraging mechanism satisfies individual rationality and incentive compatibility. Our experiments show that our proposed mechanism can significantly improve the performance of a recommender system under the short user profile situation.

Keywords : Association Rule, Bayesian network, Recommender systems

## 誌 謝

首先，我要感謝我的指導教授李永銘教授，感謝教授二年來對我的悉心指導，讓我不僅在學術上獲益良多，而且在待人處世上亦得到更多的啟發。

感謝學位論文口試委員：尹邦嚴博士、林妙聰博士與陳炳文博士，他們的指導與建議，讓本論文更臻完善。

感謝博士班學長無尾熊(賴政揚)與發哥(林聯發)，感謝二位在我學習的過程中給予我最大的協助與實質中肯的的建議。

感謝資訊經濟與科技應用實驗室(IEBI)一起學習的伙伴，文平、螢駿、彥勳、良彥、曉珍、千琇與慧珊及交大資管所前後期的同學，能夠一起學習是難忘的回憶，感謝他們在我研究所二年中對我的鼓勵與陪伴使我的求學生涯更多采多姿。

最後我要感謝我的父母、姊姊、婉儀、麗華；好友欽文、易霖、玲億、明輝與永淙的全力支持與鼓勵，使我在沒有後顧之憂的情形下，全力投入學業與工作，並得以順利完成學業。

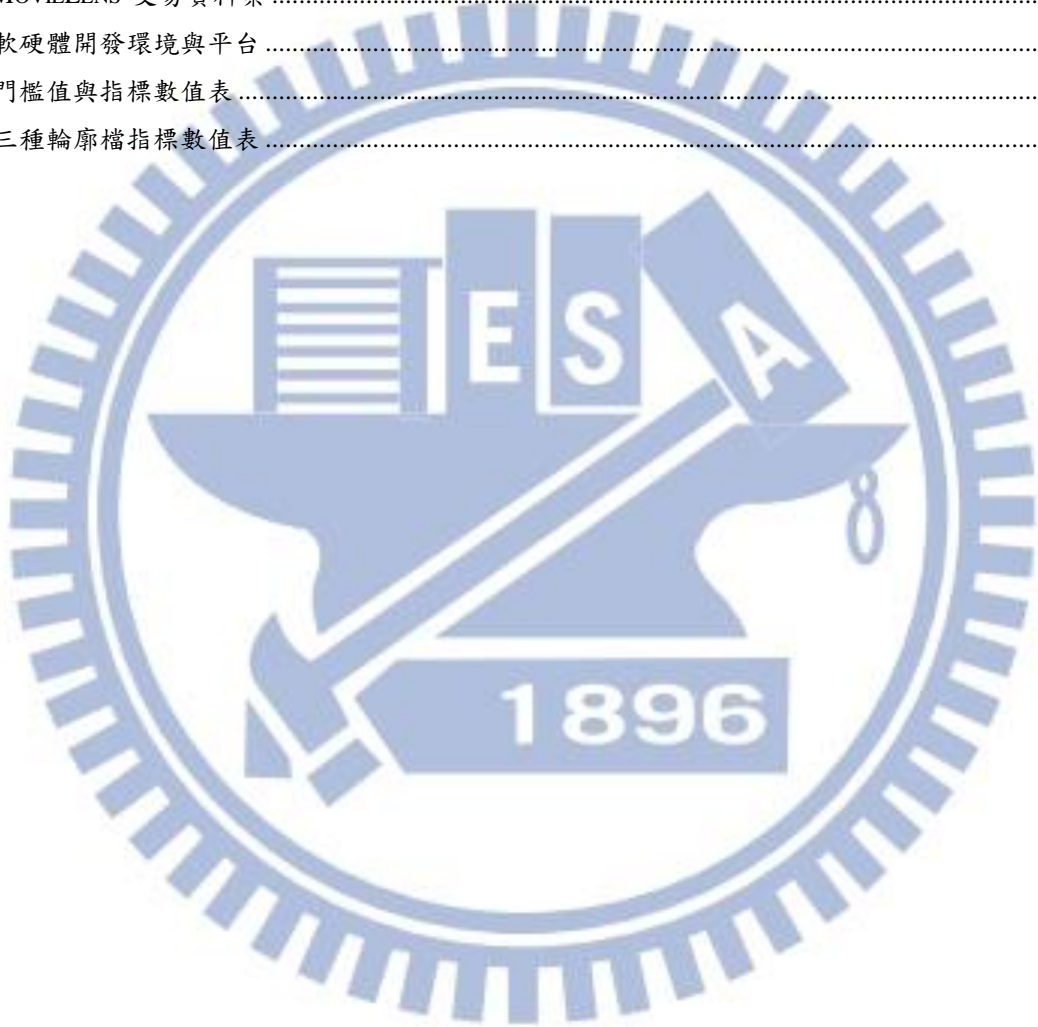
## 目錄

中文摘要 .....	I
英文摘要.....	II
誌謝.....	III
目錄 .....	IV
表目錄.....	VI
圖目錄 .....	VII
<b>第一章、 緒論.....</b>	<b>1</b>
1.1 研究動機.....	1
1.2 研究目的.....	2
1.3 論文架構.....	3
<b>第二章、 文獻探討.....</b>	<b>4</b>
2.1 推薦系統 (RECOMMENDER SYSTEM) .....	4
2.1.1 內容為基礎的推薦系統(CONTENT-BASED) : .....	7
2.1.2 協同過濾式 (COLLABORATIVE FILTERING APPROACH, CF).....	8
2.1.3 混合式 (HYBRID-BASED APPROACH) .....	12
2.2 關聯規則分析 (ASSOCIATION RULE ANALYSIS).....	14
2.2.1 關聯規則 (ASSOCIATION RULES).....	14
2.2.2 APRIORI 演算法 .....	17
2.3 貝氏網路模式 (BAYESIAN NETWORK MODEL) .....	19
2.3.1 貝氏定理 (BAYESIAN THEOREM) .....	19
2.3.2 貝氏網路 (BAYESIAN NETWORK) .....	20
2.4 激勵理論(INCENTIVE THEORY) .....	23
2.4.1 代理問題架構.....	24
2.4.2 委託代理理論基本模型.....	27
2.4.3 激勵效果.....	28
<b>第三章、 研究架構與方法.....</b>	<b>31</b>
3.1 研究架構.....	31

3.1.1	使用者輪廓(USER PROFILE)資訊拓展 .....	33
3.1.1.1	輪廓檔項目分類 .....	33
3.1.1.2	建構貝氏網路結構 .....	34
3.1.1.3	產生條件機率表 .....	37
3.1.2	鼓勵使用者對於新項目(影片)進行評比 .....	42
3.1.2.1	IR 條件 (INDIVIDUAL RATIONALITY 個人理性).....	46
3.1.2.2	IC 條件 (INCENTIVE COMPATIBLE 誘因相容) .....	46
<b>第四章、</b>	<b>實驗設計與結果.....</b>	<b>47</b>
4.1	測試資料來源與實驗環境.....	47
4.2	關聯規則相關工具.....	49
4.3	貝氏網路相關工具.....	50
4.4	推薦系統相關工具.....	51
4.5	使用者輪廓(USER PROFILE)資訊拓展實驗 .....	53
4.5.1	使用者輪廓拓展.....	53
4.5.2	推薦系統效益驗證.....	55
4.6	鼓勵使用者評比實驗.....	59
4.6.1	激勵程度計算.....	59
4.6.2	鼓勵使用者評比驗證.....	63
<b>第五章、</b>	<b>結論與未來方向.....</b>	<b>65</b>
5.1	研究結論.....	65
5.2	管理意涵.....	66
5.3	未來研究方向.....	67
<b>參考文獻-中文部份.....</b>		<b>68</b>
<b>參考文獻-英文部份.....</b>		<b>68</b>

## 表目錄

表 1 SUPPORT 值和 CONFIDENCE 值的意義 .....	16
表 2 高頻項目組 .....	16
表 3 K-ITEMSET .....	18
表 4 項目分類交易紀錄檔 .....	34
表 5 MOVIELENS 交易資料集 .....	48
表 6 軟硬體開發環境與平台 .....	49
表 7 門檻值與指標數值表 .....	56
表 8 三種輪廓檔指標數值表 .....	57

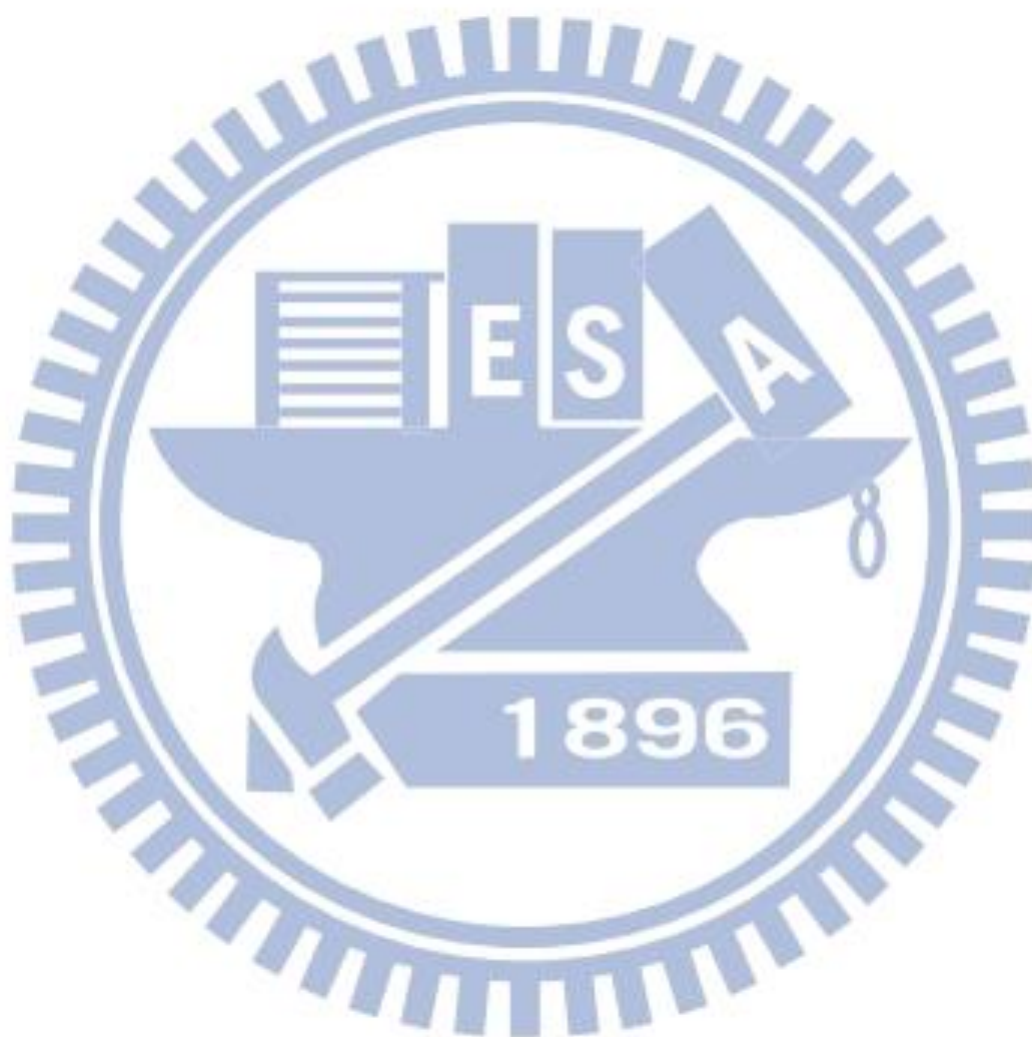




## 圖目錄

圖 1 研究步驟示意圖 .....	3
圖 2 推薦系統架構圖[5] .....	6
圖 3 協同式推薦系統架構[17] .....	10
圖 4 循序式組合 .....	12
圖 5 線性式組合 .....	13
圖 6 APRIORI 演算法之虛擬碼 .....	18
圖 7 APRIORI 演算法之副函式虛擬碼 .....	18
圖 8 DAG .....	20
圖 9 下雨與感冒的貝式網路模型 .....	21
圖 10 含有機率表之貝氏網路 .....	23
圖 11 研究架構 .....	32
圖 12 使用者輪廓檔 .....	33
圖 13 貝氏網路架構法則長度 1 .....	35
圖 14 鏈狀結構 .....	35
圖 15 收斂結構 .....	36
圖 16 網路結構(完整圖形) .....	36
圖 17 氏網路架構圖及機率表 .....	38
圖 18 貝式網路機率狀況示意圖 .....	39
圖 19 使用者輪廓拓展 .....	40
圖 20 使用者輪廓拓展使用事後推論 .....	41
圖 21 使用者 I 的預期支付 .....	44
圖 22 鼓勵分享示意圖 .....	45
圖 23 WEKA 開發工具 .....	50
圖 24 MSBNx 操作畫面 .....	51
圖 25 TASTE 系統結構 .....	52
圖 26 資訊拓展實驗架構 .....	53
圖 27 WEKA 關聯規則參數設定 .....	54
圖 28 使用者輪廓拓展 .....	54
圖 29 貝氏網路關聯 .....	55
圖 30 門檻值與指標數值圖 .....	56
圖 31 三種輪廓檔指標數值圖 .....	58
圖 32 鼓勵評比實驗架構 .....	59
圖 33 使用者 1-10 預期效用 .....	60
圖 34 使用者 11-20 預期效用 .....	62

圖 35 使用者 21-30 預期效用 .....	62
圖 36 使用者回評 .....	64
圖 37 使用者回評比例 .....	64



# 第一章、緒論

## 1.1 研究動機

由於市場白熱化的競爭，越來越多的企業利用網路來從事服務或商品行銷，除了商品本身品質的提升外，服務的品質更是成為關注的焦點，如何留住既有的客戶以及尋找潛在的客戶，更是成為當前主要的議題。網際網路逐漸成為現代人取得資訊的重要管道之一，且無時無刻都有新的資訊產生。網路上充斥著大量的資訊與資料，資訊超載(Information Overload)的問題日趨嚴重，而推薦系統可從大量的資訊裡篩選出使用者有興趣的項目，進而推薦給使用者。常見的推薦系統主要有內容導向式推薦(Content-Based Filtering)、協同式推薦(Collaborative-Filtering)以及混合式推薦 (Hybrid Recommendation Method)，而協同式推薦技術相關研究近來更是成為推薦系統發展的主流。

協同式推薦的方法通常是利用使用者的輪廓(User Profile)，收集使用者的所有資訊(例如使用者曾經瀏覽過的網站，網站屬性)，記錄著使用者個人的背景與相關興趣等等，再經由不同的推薦機制，推薦合適的產品給使用者。不同的推薦機制皆有各自的優缺點，大多是利用使用者的喜好來做為推薦的依據。當使用者的輪廓(User Profile)和/或適用的資料集具有高訊息密度時，推薦系統所推薦給使用者的產品通常是正確而且是有效率的。但是當使用者輪廓或適用的資料集資訊不完整時這些推薦機制卻都有著相同的冷啟始問題(Cold-Start problem)，亦即系統中存在著「新使用者」與「新推薦項目」。

冷啟始問題於電影的推薦中更是一個需要被關注的議題，因「新使用者」與「新推薦項目」情況的產生在電影推薦系統上是屢見不鮮的。當新的影片被存入系統時，由於新的影片沒有被使用者評分過，所以系統無法進行有效的推薦。而新進使用者在一開始進入系統時，由於系統並不知道新進使用者的喜好，所以無法正確推薦影片給使用者。

一般而言，改善新進使用者於推薦系統正確率不高的方法是採取隨機推薦的方式，即先隨機挑選幾部影片讓使用者試看，使用者在看過系統所挑選的影片過後可以給予評分，得以讓系統瞭解使用者的影片種類的喜好，再根據評分

結果推薦影片給使用者。或是利用分類法則將產品分類定找出使用者有興趣相類似的產品進行推薦[1] [2]。或是使用資料探勘中關聯規則(Association rules)的技術更進一步來擴大使用者輪廓資訊[3]。雖然關聯規則可以找出於使用者輪廓資訊內產品間潛在的關係，但其仍會因為信賴度門檻定義選擇的問題而使得某些規則遺失使得推薦正確率降低。

而改善新推薦項目的問題，大多是採用問卷方式來調查使用者的興趣，再根據問卷收集結果依照最新加入或最熱門.....等方式做推薦，來吸引使用者。不過此種方式是牽涉到問卷設計與問卷發放與回收的機制，由推薦的時效性方面來看似乎是緩不濟急。

## 1.2 研究目的

根據上述研究動機，本論文預期達成的目的可歸納為下列二項：

- (1) 提供更正確的使用者輪廓(User Profile)資訊擴大方式，避免因為關聯規則信賴度門檻定義選擇的問題而使得某些規則遺失使得推薦正確率降低。
- (2) 鼓勵使用者對於新項目(影片)進行評比，避免因為一個新的物件或產品，因大多數的人都不願意第一個去嘗試，進而使得新產品被推薦的機會不高。使用者與推薦系統產生良性互動，透過正確率移轉價值方式結合使用者利益與推薦系統利益，此方式可增加使用者轉換成本進而產生鎖定效果(Lock-in Effect)。

### 1.3 論文架構

本文共分為五個章節，第二章為文獻探討，說明目前推薦系統的類型與相關文獻的討論；第三章為研究架構與方法，說明本論文所提出的方法；第四章分析實驗數據之結果；第五章結論，總結本論文所提出的論點與未來研究方向。



圖 1 研究步驟示意圖

## 第二章、文獻探討

本章將介紹一些與本論文有關聯的前人研究。在第一節裡，先探討各種推薦系統常用的方法，包含以內容為基礎的推薦系統、協同式推薦系統與混合式推薦系統等等；第二節中，將探討對於資料探勘領域裡關聯規則(Association Rules)的研究，包含其演算法步驟之說明；第三節為說明貝氏網路理論，包含貝氏機率與貝氏圖型模式；第四節則是介紹激勵理論中代理人機制，包含個人理性條件 (Individual Rationality, IR) 與誘因相容條件 (Incentive Compatible, IC) 等機制之說明。

### 2.1 推薦系統 (Recommender System)

推薦系統 (Recommender System) 是一種資訊過濾 (Information Filtering, IF) 機制，主要是為了減少使用者在搜尋資訊過程中所附加的額外成本，其可依據使用者的喜好、興趣、行為或需求，推薦出使用者潛在需求的資訊、服務或產品[4]。在資訊過濾系統中，會將資料庫中每筆資料或所新增資料與使用者輪廓資訊加以比對，藉此過濾出與使用者相關的內容，並加以整合、分類、註解或索引，方便去判斷彼此的相關性如圖 2 [5]所示。

隨著網路科技的進步，有太多資訊充斥在我們生活中，要如何將這些有用或無用的資訊過濾出來，是網路智慧領域上的重要問題。依據推薦資訊的方式推薦系統區分為二類[6]：

(1) 非個人化推薦系統 (Non-Personalized Recommender System)：

此類推薦只將熱門產品或資訊全部呈現給使用者，因此每位使用者將會看到相同的產品或資訊。

(2) 個人化推薦系統(Personalized Recommender System):

根據使用者的興趣和購買行為，向使用者推薦感興趣的訊息或商品。隨著電子商務規模的擴大，商品類型快速增長，使用者需要花費大量的時間才能找到自己想買的商品。為了解決這些問題，個性化推薦系統應運而生。

- 屬性導向推薦系統 (Attribute-Based Recommender System) :  
依據使用者對產品特定屬性的喜好，提供符合該項屬性的資訊或產品。

- 項目相關導向推薦系統 (Item-to-Item Correlation Recommender System) :

利用歷史紀錄比對出同時購買的項目，即找出項目的關聯規則，再利用此規則進行項目之間相互推薦。

- 使用者相關導向推薦系統 (People-to-People Correlation Recommender System) :

針對使用者找出具有相類似喜好的其他使用者，再依據同好對產品或資訊的評價找出適合的項目推薦給使用者。

推薦機制的技術大致上可以分成以下三種：以內容為基礎的推薦系統 (Content-Based) [7][8]、協同式推薦系統(Collaborative Filtering)[9]、混合式推薦系統(Hybrid)[7][10]，以下分別說明。

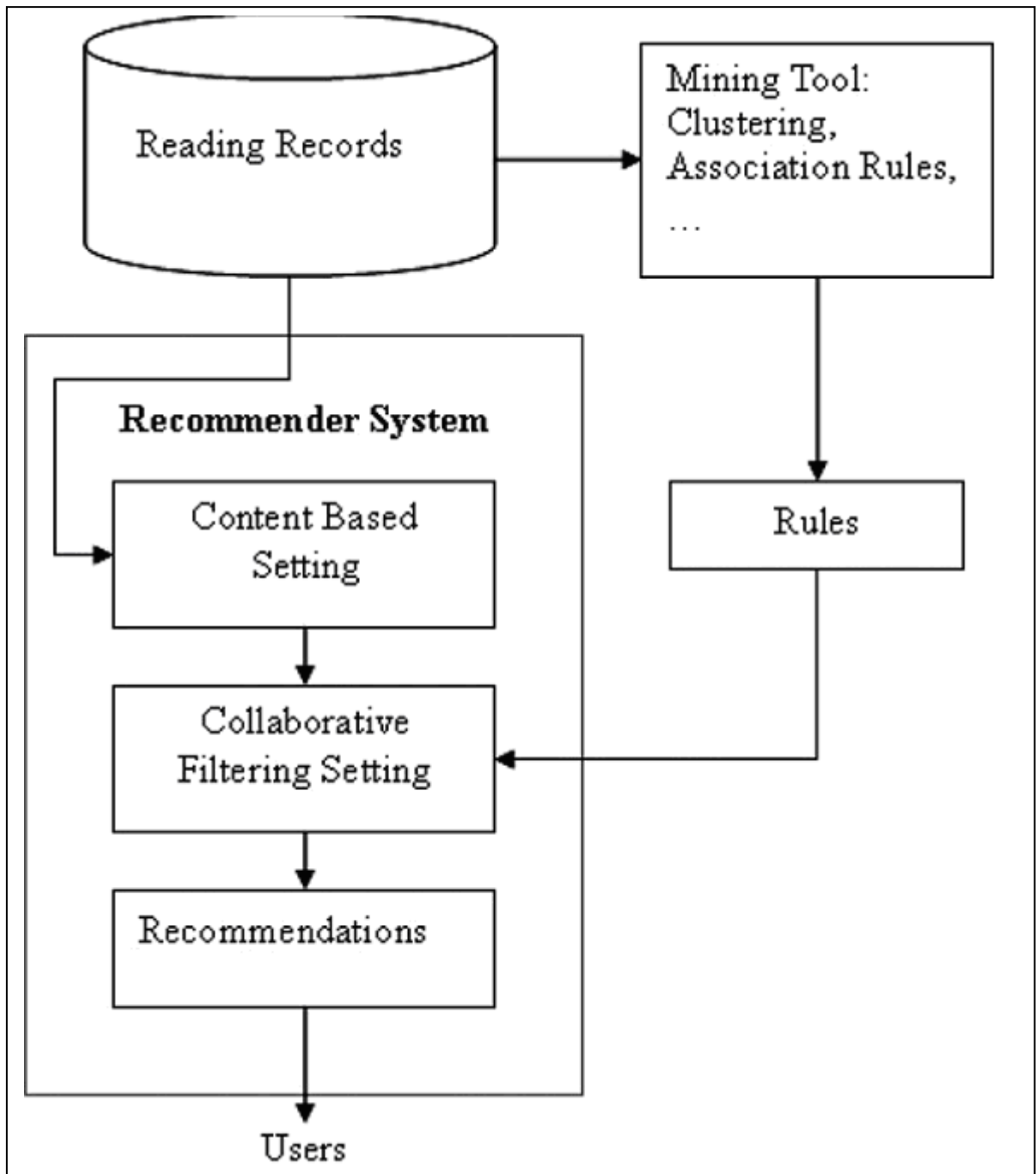


圖 2 推薦系統架構圖[5]



### 2.1.1 內容為基礎的推薦系統(Content-Based)：

使用者面對選擇之時，會依據印象中接近或是類似程度較高的產品。所選擇的產品都會包含個人喜好的特徵。內容式推薦系統，又稱為以特徵為基礎過濾(Feature-Based Filtering)。其依據的準則是對物品內容分析，而不是其他使用者的評價。

其概念主要源自於「資訊擷取(Information Retrieval, IR)」中的方法，主要著重於項目 (Item) 的分析，藉由分析項目的屬性特徵，再與使用者的輪廓資料做比對，藉此推薦出使用者可能會有興趣的項目。這種技術一般都稱之為內容式資訊過濾技術推薦(Content-Based Recommender)。對於此種推薦方式而言，了解使用者的喜好性趣是相當重要的問題。收集使用者的喜好可分為以下兩種方式：

- 隱性(Implicit)資料收集：

系統針對使用者行為進行記錄，此種紀錄可能是在使用者不知情的情況下，或是知情但是在不與使用者有互動，或鮮少互動的過程中進行。其目的為方便推測使用者背後的行為。不過也因與使用者的互動極少，可能造成系統推測的錯誤率提高。

- 顯性(Explicit)資料收集：

在使用者使用系統的過程中(可能為「使用前」或「使用中」或「使用後」)，系統會透過詢問使用者的方式收集使用者對系統的評價或感受。此種方法是得到使用者偏好最直接的方法，但是也因此有可能會引發使用者的負面情緒反應。

使用者與系統互動程度的多寡是顯性與隱性最大的差別，交流程度越高，所得到的喜好資訊將可能為使用者真實的偏好，但此種方式對使用者造比較大的負擔。一個可以完全反應出使用者偏好，又可降低使用者負擔的資訊取得方式，將是設計推薦系統需考量因素之一。而內容為基礎的推薦系統，有以下缺點：

- 特殊化問題(Over-specialization Problem)：

舉例來說，於線上書店推薦系統中，當系統知道使用者特別喜歡看勵

志小說的時候，系統只會推薦勵志小說給使用者閱讀，這種方法，比較會導致使用者只想專注的看某一些類型的小說，一個好的推薦系統事要避免這種情形的發生。

- 有限分析問題(Limited Analysis Problem)：

物件會以許多形式上的出現，如音樂、雜誌、電腦、視訊、廣告... 等等，目前現階段僅對於處理文字技術方面比較成熟。

- 新使用者問題(New User Problem)：

新加入的使用者，在系統上沒有任何使用記錄時或是使用程度不足時，會導致無法很正確且即時的做出有效的推薦。

### 2.1.2 協同過濾式 (Collaborative Filtering Approach, CF)

協同過濾主要是利用全體的觀點來產生推薦項目給特定的個人使用，主要是重點在於建立大量使用者的輪廓(User Profile)，除了紀錄每一個使用者所設定的主題之外，還記錄使用者個人的背景與相關興趣等等，之後針對使用者所查詢的主題，透過用戶模型(User Modeling)技術形成社群 (Community)，亦即在某些特定行為或偏好上有相類似特性成員的集合，並透過這些鄰近者所組成的鄰近群組之意見或建議，來產生使用者有興趣的推薦資訊給特定的使用者做為參考[11]如圖 3 [17]所示。而依據所採用的演算法又將協同過濾概分為二種主要的類別[12]：

- (1) 以記憶為基之協同過濾 (Memory-Based CF)：

依據使用者的歷史記錄，與資料庫中每一筆紀錄比較其相似度以找出與使用者喜好相類似的鄰近族群，再依族群的喜好進行推薦。其中最常使用的方法為最接近鄰居法 (Nearest Neighbors)。以記憶為基之協同過濾又可以細分為以下二項

- 以使用者為基礎 (User-based) 的協同過濾：

用相似統計的方式取得具有相同愛好或特徵的使用者，所以稱之為以使用者為基礎 (User-based) 的協同過濾或基於鄰居的協同過濾(Neighbor-based Collaborative Filtering)。其方法步驟如下：

### I. 使用者資訊收集：

收集可以代表使用者喜好的資訊並給與評分。其評分方式有二種。一般的網站的使用評分的方式稱為「主動評分」。另外一種稱為「被動評分」，根據使用者的行為模式由系統代替使用者進行評價，不需要使用者直接打分或輸入評價資料，例如電子商務網站記錄使用者購買的商品記錄。

### II. 最近鄰搜索(Nearest neighbor search, NNS)

其出發點為與使用者興趣喜好相似的另一使用者，換句話說即為計算兩個使用者的相似度。例如：找出 10 個和 A 有相類似喜好的使用者，將其對 B 的評分作為 A 對 B 的評分預測。一般即會根據資料之不同採用不同的演算法，目前較多使用的相似度演算法有 Person Correlation Coefficient、Cosine-based Similarity 與 Adjusted Cosine Similarity。

### III. 產生推薦結果

有了最近鄰居集合，即可對使用者的喜好進行預測，進而產生推薦結果。依據推薦目的之不同採取不同形式推薦，一般而言，常見推薦方式有 Top-N 推薦和關聯規則推薦。Top-N 推薦為針對單獨使用者產生，對每個使用者有不同的結果，關聯推薦是對最近鄰居使用者的記錄進行關聯規則 (association rules) 探勘。

#### ● 以項目為基礎(Item-based)的協同過濾

隨著使用者數量增加，以使用者為基礎的協同推薦的演算法其計算時間會跟著變長，進而造成效能降低。Sarwar 於 2001 年提出基於項目的協同過濾推薦演算法(Item-based Collaborative Filtering Algorithms)。「可以引起使用者興趣的項目，必定和之前高評分的項目相類似」，此為其演算法的基本假設。使用者之間的相似度改採用項目間的相似度來代替。其方法步驟如下：

### I. 收集使用者資訊：

同樣以使用者為基礎 (User-based) 的協同過濾。

## II. 針對項目的最近鄰搜索：

首先計算已經評價過的項目和即將要預測的目標項目的相似度，透過加權(即相似度)各已評價項目的分數，得到待目標項目的預測值。例如：對 A 和 B 進行相似度計算，首先要找出同時對 A 和 B 評過分數的組合，再對找出的組合進行相似度計算，一般常用的演算法與以使用者為基礎 (User-based) 的協同過濾採用的方式相同。

## III. 產生推薦結果

此方式不需考慮使用者間的差異，所以精確度較差。但其優點為不需要使用者的歷史資訊，或對使用者進行識別。因此可以採用離線方式完成相似性計算(大量)，進而降低線上計算。尤其是在使用者數量多於項目數量的情形下可以顯著提高推薦效率。

### (2) 以模式為基之協同過濾 (Model-Based CF)：

將使用者歷史記錄，經由統計或機器學習方式來建構出使用者偏好模型，進而利用此偏好模型來產生推薦。目前常見的方法包括，潛在語意索引(Latent Semantic Indexing, LSI)、關聯式規則(Association Rules)、貝式網路 (Bayesian Network) 或迴歸分析 (Regression Analysis) 等等。

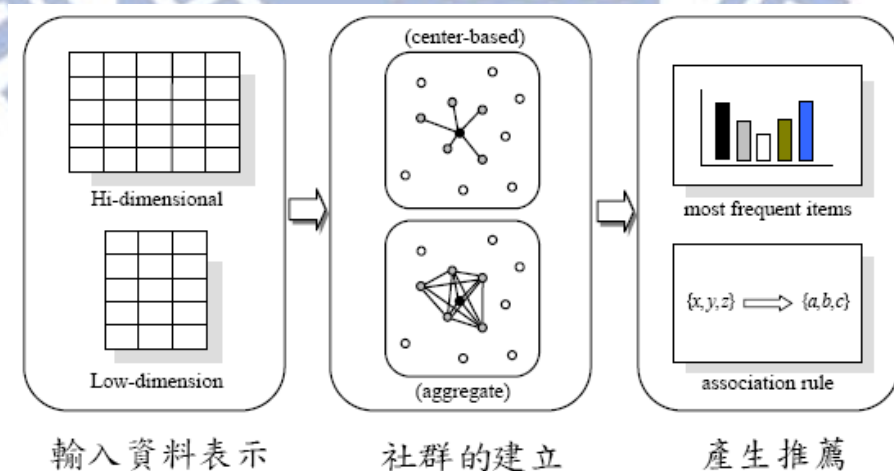


圖 3 協同式推薦系統架構[17]

協同式推薦系統，雖然可以避免以內容為基礎式的推薦可能會發生的問題，如特殊化、分析物件內容有限等，但還是有其他缺點：

- 新使用者問題(New User Problem)：

新加入的使用者，於系統內無任何使用記錄時或是使用程度不高時，將無法正確且即時的做出有效推薦。此問題亦發生於內容為基礎的推薦系統中。

- 稀疏化問題(Sparsity Problem)：

推薦系統通常都會對於數量相當龐大的產品或物件來作為評價與推薦，但是使用者所接觸到的物件大部分都只佔非常小的比例。當產品的數量大於使用者，或兩者的比例十分懸殊時，系統在尋找到適合的推薦者來進行推薦的命中率或推薦成功率就會降低。

- 新物件問題(New Item Problem)：

因為一個新的物件或產品，因大多數的人都不願意第一個去嘗試，進而新產品被推薦的機會不高。

協同過濾是目前使用最成功的推薦系統技術，而其最著名的應用是一個稱為「Tapestry」的郵件篩選分類系統，由 Goldberg 於 1992 年提出[19]，此系統被視為最早應用協同推薦系統的設計。Palo Alto 研究中心發展這項實驗性郵件系統的原因，主要是由於研究中心工作人員每天需接收大量的電子郵件但卻無從篩選分類進而造成工作效率降低。

另一著名的系統為 1994 年所發展的 GroupLens[23]，此系統主要是應用於新聞內容的篩選上，透過此系統來幫助讀者過濾其感興趣的新聞內容。系統建立的前提為讀者以前感興趣的東西在未來也會有興趣閱讀。讀者看過內容後給予一個評比分數，系統將記錄其分數作為系統未來推薦參考之用。

在 GroupLens 後還有相近的 MovieLens(電影推薦)、Ringo(音樂推薦)與 Jster(笑話推薦)等等，經技術不斷的進步，網路越加發達，使用者參與的程度越來越高，使得推薦系統也發展得越來越嚴密。

### 2.1.3 混合式 (Hybrid-Based Approach)

混合式推薦系統(Hybrid Approach)即是結合了內容式推薦與協同式推薦的兩種方法所形成的推薦系統。以內容為基礎的推薦與協同過濾正好可以互補其缺點，取其各自的優點，所以有學者提出結合以內容為基礎與協同過濾這兩種方法[13]。現今技術中主要的兩種混合形式：循序組合(Sequential Combination)如圖 4，線性組合 (Linear Combination) 如圖 5。

- 循序組合：這一個類型的推薦系統主要分成兩步驟  
步驟 1. 內容式過濾法先找出有相同喜好相似的使用者。  
步驟 2. 再透過協同式過濾法進行預測。

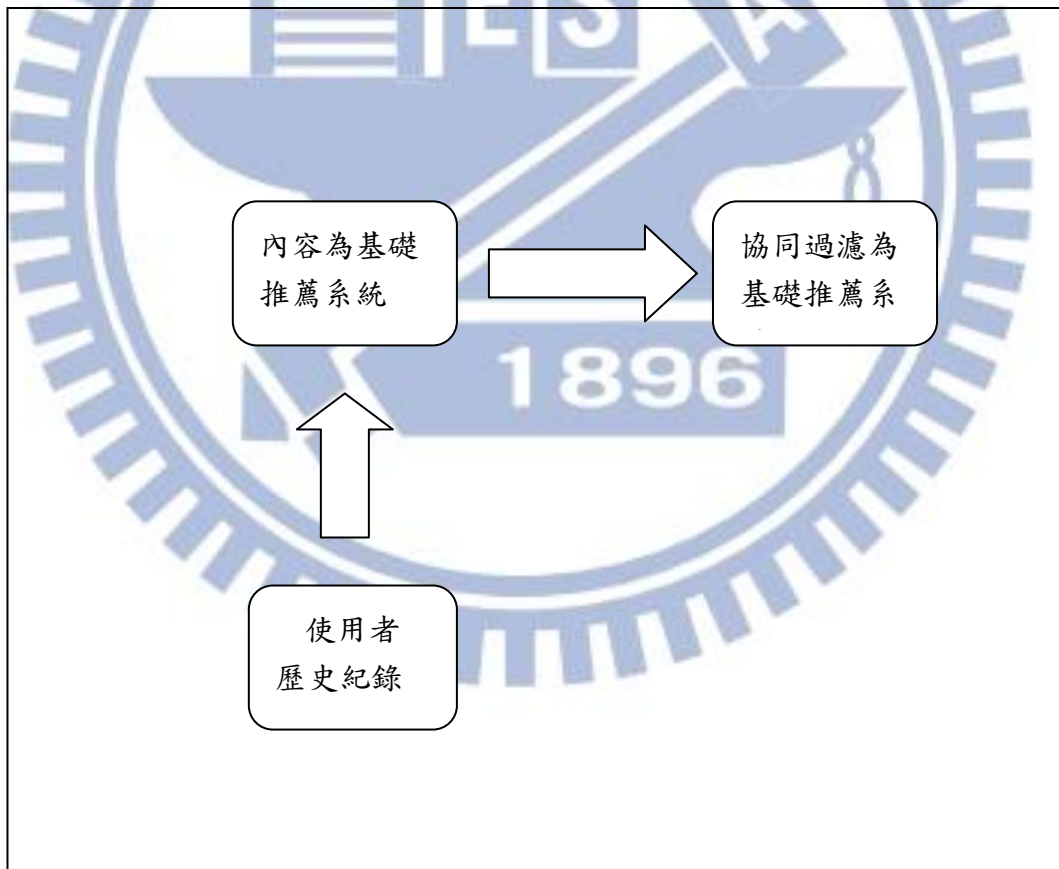


圖 4循序式組合

- 線性組合：

其概念非常直觀與單純，就是將協同式和內容式推薦系統計算出來所得到的數值，或是將內容為基礎和協同式推薦系統所計算的數值分別加權相加。

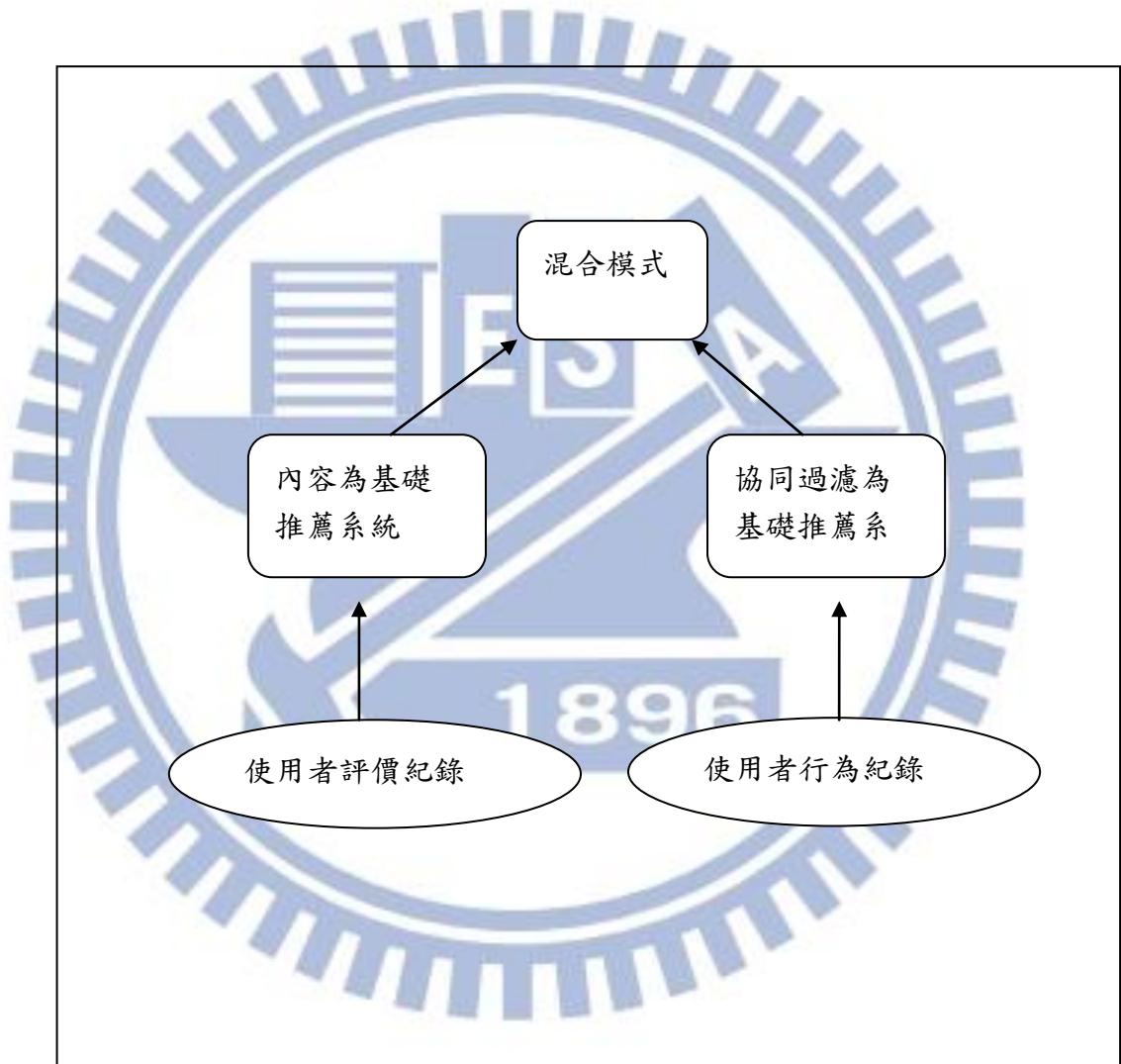


圖 5 線性式組合

## 2.2 關聯規則分析 (Association Rule Analysis)

關聯式規則(Association Rules, AR)，又稱關聯規則，是資料探勘領域中相當重要的研究課題，主要是從大量數據中探勘出有價值的資料項目之間相關性。關聯規則被用來解決的問題如：「若一個消費者購買了 A 產品，那麼他有多大機會將會購買 B 產品？」以及「若一個消費者購買了產品 A 和 B，那麼其還將購買其他哪些產品？」，其他相關研究如，DAN 資料的分析、紋理特徵(Texture Features)的分析等等，因為此項研究的結果與企業決策的制定有相當程度的互動。

關聯規則主要在於處理分析大量且雜亂無章的資料，並將資料分析減少成為易於觀察、理解與解釋的靜態資料。關聯規則最早由 Agrawal et al. (1993) 所提出，並隨後於 1994 年提出 Apriori 演算法運用在關聯規則的推導。後續許多關聯規則相關的研究亦基於此演算法，如 F-P 演算法、以布林運算為基礎的演算法、Pincer-Search 演算法等等，皆是著名的關聯規則探勘的演算法。以下即針對關聯規則的基本定義及 Apriori 演算法分別詳細說明之。

### 2.2.1 關聯規則 (Association Rules)

關聯規則的定義為：在交易資料庫中有多筆的銷售交易記錄，每一筆交易記錄可能含有一項或多項商品以形成該次交易紀錄。

令  $I = \{I_1, I_2, \dots, I_m\}$  是所有相異項目 (Items) 所成的集合。假設資料庫  $D$  有  $n$  個交易紀錄， $T = \{T_1, T_2, \dots, T_n\}$ ，而每筆交易記錄  $T_p$ ，都是由一部份的項目所組成之集合，即  $T_p \subseteq I$ ，其中  $p = 1, 2, \dots, n$ 。

某一集合  $X$ ，若  $X \subseteq I$ ，則稱  $X$  為項目集 (itemset)，若此  $X$  包含  $k$  個項目，則稱  $X$  為  $k$ -項目集 ( $k$ -itemset)。

簡言之，商品項目組所成的集合所代表的就是某些特定商品組合，如： $\{\text{麵包}, \text{牛奶}\}$ 、 $\{\text{啤酒}, \text{尿布}, \text{汽水}\}$ 。若在項目集中  $X$  與  $Y$  之間有一關聯規則存在，其表示為  $X \Rightarrow Y$ ， $X, Y \subseteq I$  且  $X \cap Y = \phi$ ，其中  $I$  是交易資料庫中所有項的集合，如： $X \Rightarrow Y$ ，說明成「購買  $X$  產品的前提下，同時也會購買  $Y$ 」。

然而，所找出來的關係不一定是強烈相關的，所以需要一些指標來判斷這



些關係是否具有可信度。而在關聯式演算法中最常使用的兩個衡量指標，信賴度 (Confidence) 與支持度 (Support)。

支持度 (Support) 代表某項目集在所有交易中出現的頻率，其計算公式為如下：

$$P(X \cap Y) = \frac{\text{交易中同時包含X和Y產品}}{\text{總交易}}$$

假設：交易資料庫中有 100 筆交易，若其有 50 筆交易被發現同時含有 {啤酒, 尿布, 汽水}，則 {啤酒, 尿布, 汽水} 項目集的支持度即為 50/100=50%。在規則探勘的過程中會發掘出各種組合的項目集，支持度太低的項目集不具代表性，因此需設定一個最低門檻值(Threshold)做為過濾的依據，稱之為最小支持度 (Minimum Support)。所找出的項目集的支持度若超過最小支持度則稱為一個頻繁項目集 (Frequent Itemset)。若支持度越高，就表示此項目集同時出現的機率就越高。

信賴度 (Confidence) 其利用了條件機率的概念來定義關聯規則可信賴的程度。也就是 X 出現的條件下，Y 也會跟著出現的條件機率。數值越高，規則的可靠度也就越高。信賴度的計算方式如下：

$$P(Y | X) = \frac{\text{交易中同時包含X和Y}}{\text{包含X產品的交易}}$$

假設：交易資料庫中有 100 筆交易，若其有 50 筆交易被發現同時含有 {啤酒, 尿布, 汽水}，而有 60 筆交易含有 {啤酒}，則關聯規則「啤酒=>汽水, 尿布」的信賴度即為 50/60=83.3%。若信賴度太低則表示該關聯規則沒有代表性，因此需設定一個最低門檻值(Threshold)做為過濾的依據，稱之為最小信賴度 (Minimum Confidence)。若信賴度越高，就表示這條法則成立的機率就越高。經由關聯規則計算結果所代表之意義，如表 1所示：

表 1 Support 值和 Confidence 值的意義

	低 Confidence	高 Confidence
高 Support	發生頻率高，但關聯規則不正確。	發生頻率高，且關聯規則正確。
低 Support	發生頻率低，且關聯規則不正確。	發生頻率不高，但關聯規則正確率很高。

由支持度與與信賴度的公式不難發現， $X \rightarrow Y$  之支持度等同於  $Y \rightarrow X$  之支持度，即說明支持度是具有反身性的 (Reflexive)。但信賴度並不具有反身性，即  $X \rightarrow Y$  的信賴度和  $Y \rightarrow X$  的信賴度是不同的。由於信賴度沒有反身性，如果信賴度門檻值定義不佳，可能就會失去其中一組規則。以表 2 為例，假設表中是四組超過最小支持度的頻繁項目集。假設最小信賴度為 100%， $A \rightarrow B$  因符合門檻值而被採用，而  $B \rightarrow A$  則因不符門檻值而被刪除。但若 A 產品和 C 產品互為替代品，則對顧客而言，於購買 B 產品的前提下，A 產品和 C 產品都是會考慮選擇的產品。但因為 A 產品與 C 產品交替選擇的關係進而造成  $B \rightarrow A$  因不符最小信賴度而被刪除。

表 2 高頻項目組

L <sub>3</sub>	頻繁項目集
	AEB
	CDB
	ABD
	CBF

## 2.2.2 Apriori 演算法

在關聯規則探勘技術的研究中，由 Agrawal et al. (1994) 所發表的 Apriori 演算法最廣為人知，此一演算法已成為關聯規則中最具代表性的演算法之一。此演算法的核心概念就是藉由不斷地掃瞄資料庫的交易記錄，並計算出各種可能的項目組合（候選項目集）的支持度，最後只留下支持度達到門檻值的頻繁項目集。Apriori 演算法運作步驟如下[14]：

- (1) 計算每一個 items 的 support，找出大於或等於 minsupport 的 itemsets，稱為 seed set (large itemsets)
- (2) 利用(1)所找出的 large itemsets 產生出下一階段可能成為 large itemsets 的 itemsets，稱為 candidate itemsets。
- (3) 計算 candidate itemsets 的 support
- (4) 找出(3)中大於或等於 minsupport 的 itemsets，成為新的 seed set
- (5) 重複(1)~(4)，直到找不到新的 seed set 為止

圖 6與圖 7為 Apriori 演算法之虛擬碼，演算法中的表示符號說明如下：資料庫欄位為  $\langle TID, item \rangle$ ，欄位 item 的中順序皆按字典順序(lexicographic)排列。itemset 中 item 個數為 size，k-itemset 表示此 itemset 中有 k 個 item。 $L_k$  為 large itemsets 的集合， $C_k$  為 Candidate itemset 的集合，其中 k 為 item 的個數(如表 3)。

Apriori 演算法是目前資料探勘領域中最常使用之演算法其優點與缺點如下所述：

- 優點：
  - 簡單易懂、實現複雜度低
- 缺點
  - 在找高頻項目集合時，會產生大量的候選項目集合(Candidate Itemset)
  - 必須經常掃瞄整個資料庫，造成執行效率不佳

表 3 k-itemset

K- itemset	項目集含有 K 個 Item
$L_k$	Large k-itemsets 集合(皆大於 minsupport) 1) itemset , 2)support count
$C_k$	Candidate k-itemsets 集合 1) itemset , 2)support count

```

1)  $L_1 = \{\text{large 1-itemsets}\};$ 
2) for ( $k = 2; L_{k-1} \neq 0; k++$ ) do begin
3)    $C_k = \text{apriori-gen}(L_{k-1});$  //New candidates
4)   forall transactions  $t \in D$  do begin
5)      $C_t = \text{subset}(C_k, t)$  //Candidates contained t
6)     forall candidates  $c \in C_t$  do
7)        $c.\text{count} ++;$ 
8)     end
9)    $L_k = \{c \in C_k \mid c.\text{count} \geq \text{minsup}\}$ 
10) end
11) Answer =  $\bigcup_k L_k$ 

```

圖 6 Apriori 演算法之虛擬碼

```

1) insert into  $C_k$ 
2) select p.item1, p.item2, ..., p.itemk-1, q.itemk-1
3) from  $L_{k-1}p, L_{k-1}q$ 
4) where p.item-1 = q.item-1, p.item-k-2 = q.item-k-2, p.item-k-1 < q.item-k-1;
5) forall itemsets  $c \in C_k$  do
6)   forall (k-1)-subsets  $s$  of  $c$  do
7)     if ( $s \in L_{k-1}$ ) then
8)       delete  $c$  from  $C_k$ 

```

圖 7 Apriori 演算法之副函式虛擬碼

## 2.3 貝氏網路模式 (Bayesian Network Model)

貝氏信任網路 (Bayesian Belief Network, BBN) 是結合貝氏機率理論與圖形模式所發展出來的知識表示方式，藉由節點、連結與條件機率表三部分[15]。貝氏網路是一個有向的非循環圖，每個節點代表一個隨機變數，每條連結線指出兩個變數之間的交互關係[16]。簡言之，這個有向圖是這些變數之聯合機率分佈的分解表示法。以下將分別針對貝氏定理及貝氏網路做一較為詳細之說明。

### 2.3.1 貝氏定理 (Bayesian Theorem)

假設  $C_1, C_2, \dots, C_n$  為樣本空間 (Sample Space)  $S$  的分割，且有一事件 (Event)  $A$ ，在此前提下有兩定理存在：

**定理 1：總機率法則 (Law of Total Probability)**

$$P(A) = \sum_i P(C_i)P(A|C_i)$$

**定理 2：貝氏定理 (Bayes' Rule)**

$$P(C_j|A) = \frac{P(C_j)P(A|C_j)}{P(A) = \sum_i P(C_i)P(A|C_i)}$$

其中  $P(C_i)$ : 事前機率 (Prior Probability)

$P(A|C_i)$ : 樣本機率 (Sample Probability)

$P(C_j|A)$ : 事後機率 (Posterior Probability)

從公式可知，貝氏定理可結合事前機率 (Prior Probability) 與樣本機率 (Sample Probability) 來推算事後機率 (Posterior Probability)。較一般利用統計方法來說，貝氏定理更有效的運用有限的樣本資訊與經驗值，使得分析資料時不需太多樣本資訊即可得到理想的統計數值，進而做更有完整與有效率的推論。

### 2.3.2 貝氏網路 (Bayesian Network)

在圖形模式上，貝氏網路是一個有向的非循環圖 (Directed Acyclic Graphs, DAG)，每個節點代表一個隨機變數，每條連結線指出兩個變數之間的交互關係。換句話說，這個有向圖是這些變數之聯合機率分佈的分解表示法。圖 8 就是 DAG 的一個簡單例子。

圖中的節點 (Node) 表示一個隨機變數，變數之間的連線 (Link) 則表示事件的交互關係，其影響程度藉由條件機率來表達。因此圖 8 又可以轉換為以下數學式子：

$$P(A,B,C,D,E) = P(E | C)P(D | B, C)P(C | A)P(B | A)P(A)$$

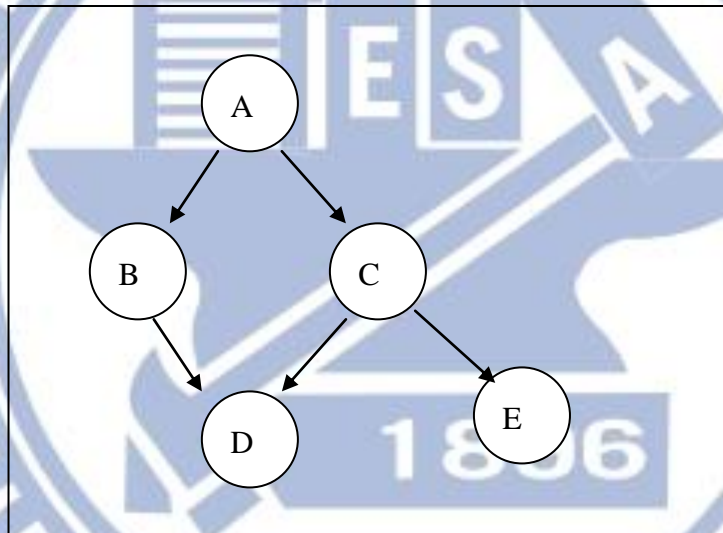


圖 8 DAG

以下將用一個簡單的例子（下雨與感冒）來說明貝氏網路原理與過程，此例將說明貝氏網路中的因果關係並非是決定性的，而是推論性，而節點之間的推測關係是透過條件機率表來呈現。

假設有 A、B 二位民眾出門沒帶雨具，而下雨將可能會導致感冒。根據天氣預報下雨的機率為 60%。令淋雨導致感冒的機率為 70%，而未下雨但得到感冒的機率為 10%，其表示方式如下

R 表示下雨事件， $A'$ ,  $B'$  表示 A, B 得到感冒的變數。

$$P(R = Y) = 0.6, P(R = N) = 0.4。$$

$P(A' | R=Y)=0.7, P(B' | R=Y)=0.7$  淋雨導致感冒的機率。

$P(A' | R=N)=0.1, P(B' | R=N)=0.1$  未下雨但得到感冒。

圖 9為下雨與感冒的貝式網路模型。透過此模型可以推導出在 A 已經得到感冒的情形下 B 得到感冒的機會。

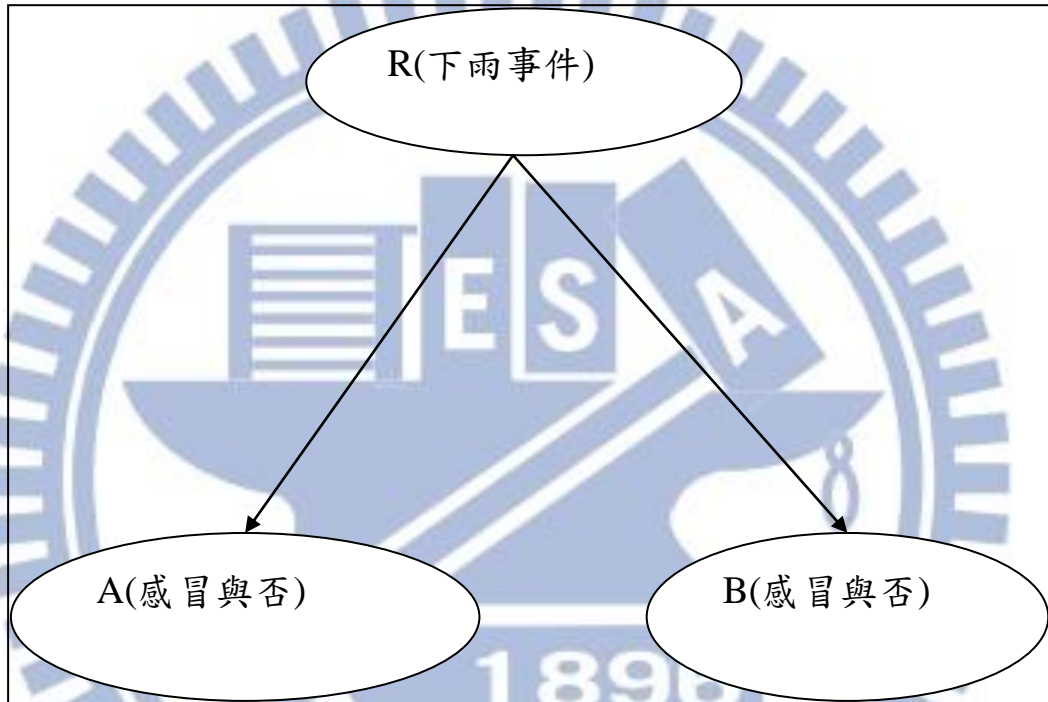


圖 9下雨與感冒的貝式網路模型

- 計算在沒有任何資訊的提供下 A 和 B 發生感冒的機率

$$P(A \text{ 感冒, 下雨}) = P(A' = Y | R = Y) \times P(R = Y) = 0.7 \times 0.6 = 0.42$$

$$P(A \text{ 感冒, 沒下雨}) = P(A' = Y | R = N) \times P(R = N) = 0.1 \times 0.4 = 0.04$$

$$P(A \text{ 沒感冒, 下雨}) = P(A' = N | R = Y) \times P(R = Y) = 0.1 \times 0.6 = 0.09$$

$$P(A \text{ 沒感冒, 沒下雨}) = P(A' = N | R = N) \times P(R = N) = 0.9 \times 0.4 = 0.36$$

$$P(A \text{ 感冒}) = 0.42 + 0.04 = 0.46 \quad P(A \text{ 沒感冒}) = 0.09 + 0.36 = 0.45$$

B 的行況同 A 所以機率為

$$P(B\text{感冒}) = 0.42 + 0.04 = 0.46 \quad P(B\text{沒感冒}) = 0.09 + 0.36 = 0.45$$

- A 已發生感冒前提下，推導其他節點(R 和 B)的機率值

$P(R=Y)^*$  與  $P(R=N)^*$  表示觀察到 A 已發生感冒時，發生下雨機率的變化。

$$P(R=Y)^* = P(R=Y | A'=Y) = \frac{P(A'=Y | R=Y)}{P(A'=Y)} = \frac{0.7 \times 0.6}{0.46} = 0.91$$

$$P(R=N)^* = P(R=N | A'=Y) = \frac{P(A'=Y | R=N)}{P(A'=Y)} = \frac{0.1 \times 0.4}{0.46} = 0.09$$

若觀察到 A 已發生感冒，下雨的機率則有可能提升。

- 當下雨機率改變時，推導 B 發生感冒機率的變化

$$P(B\text{感冒, 下雨}) = P(B'=Y | R=Y) \times P(R=Y)^* = 0.7 \times 0.91 = 0.637$$

$$P(B\text{感冒, 沒下雨}) = P(B'=Y | R=N) \times P(R=N)^* = 0.1 \times 0.09 = 0.009$$

$$P(B\text{感冒}) = 0.637 + 0.009 = 0.646$$

由上述例子可知貝氏網路以整體性觀點來調整整個網路，換句話說當機率值調整時，貝氏網路上所有相關節點可根據其條件機率而同時調整，此特性使得在推論過程，若中有新資訊進來便能立即反應出整體事件可能發生的機率。如上述例子中，當以知 A 已發生感冒，則下雨的機率由 0.6 提升為 0.91，相同情況下 B 發生感冒的機率也因下雨機率的增加而提升至 0.646。

圖 10 為含有機率表之貝氏網路架構圖，假定圖中節點是我們想了解問題變數，以本論文來說就是過去被評價過的項目，而圖中的機率表，則是表示父節點於各種狀態下，子節點被評價的機率。



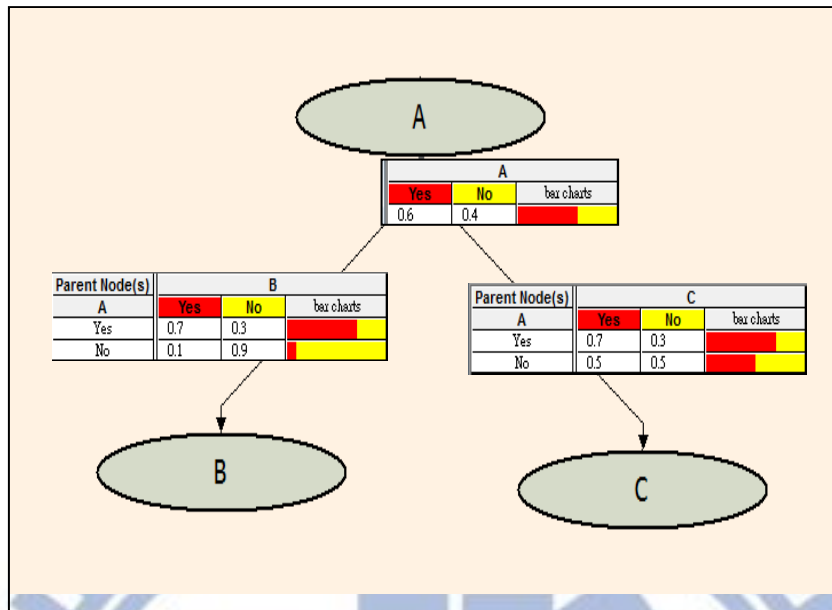


圖 10 含有機率表之貝氏網路

## 2.4 激勵理論(Incentive theory)

在經濟發展的過程中，分工與交易的出現帶來了激勵問題。激勵理論是用於處理需要、動機、目標和行為四者間關係的理論。30 年代經濟學家開始關心被傳統經濟理論所忽視的企業內部管理效率問題，認識到激勵的重要性，並開始對激勵問題進行研究。與管理學通過對人的多種需求研究激勵不同，經濟學對激勵的研究是以經濟人為出發點，以利潤最大化或效用最大化為目的。進入 70 年代以後，激勵問題成為其中非常重要的研究課題。而委託代理理論 (Principal-agent Theory) 亦是激勵理論的主要內容之一，其主要研究為委託代理關係。指一個或多個行為主體依據明示或隱含的契約，指定另一些行為主體為其服務，並授予一定的決策權利，並根據後者提供服務數量和質量對其支付相應的報酬。授權者就是委託人，被授權者就是代理人。如何設計一套激勵機制來激勵代理人努力工作即為委託代理理論的核心工作。本論文即透過此激勵模式來激勵使用者進行新產品的推薦與評比。以下分別簡述代理問題架構和幾項委託代理理論基本模型。

## 2.4.1 代理問題架構

代理問題即為一種機制設計(mechanism design)[17]與選擇的問題。此類問題當中將有一個「委託人」(principal)和一個或多個「代理人」(agents)；委託人的支付函數式(Payoff)為共同的知識，代理人的支付函數只有代理人自己知道，委託人和其他代理人皆不知道。舉例來說，在一級密封價格拍賣(The first-price sealed auction)中，賣方不知道買方對拍賣品的評價；而在雙方叫價拍賣中，拍賣人(auctioneer)亦不知道買方的評價，也不知道賣方的商品成本；以企業訂價的角度來看，在企業(壟斷型)對商品的訂價中，企業亦不知道消費者願意付出的最高價格。

委託人或許可以直接要求代理人說明自己的類型，但代理人可能不會說實話，除非委託人能提供給代理人足夠的「激勵」(Incentive)，而激勵可以為貨幣或非貨幣的型態。因提供激勵是有成本的，因此委託人將面臨成本與收益的問題。

委託人設計機制目的是最大化自己的期望效用函數。其將會面臨二個條件(約束)。<sup>[17][C3]</sup>

- 個人理性條件(individual rationality, IR)：

如果一個理性的代理人有任何興趣接受委託人設計的機制，則其代理人在此機制下所獲得的期望效用必須「不小於」其在「不接受」這個機制時所得到的最大期望效用。此條件又稱為「參與約束」(participation constraint)。另外在「不接受」這個機制時所得到的最大期望效用稱為保留效用(reservation utility)。

- 誘因相容條件(incentive compatible, IC)：

此條件又稱為激勵相容條件，在委託人不知道代理人類型的前提下，於此機制內代理人必須有其積極性來選擇委託人希望他選擇的行動。顯然，只有當代理人選擇委託人所希望的行動時得到的期望效用「不小於」他選擇其他行動時得到的效用時，代理人才有積極性選擇委託人所希望的行動。

滿足「個人理性條件」的機制又稱為「可行機制」(feasible mechanism)；

滿足「誘因相容條件」的機制又稱為「可實施機制」(implementable)，如果一個機制滿足「個人理性條件」與「誘因相容條件」時，可稱此機制是「可行的可實施機制」。委託人的問題是選擇一個「可行的可實施機制」以最大化其預期效用(「可行的可實施機制」可能有很多個)。

以下將透過上述原則以工資設計機制為例說明代理問題。假設有兩名參賽者，委託人與代理人，委託人付工資  $E$  雇用代理人進行工作。代理人可能會以不同努力的程度來進行工作，以  $m \in \{m_H, m_L\}$  表示代理人努力程度有兩種可能， $m_H$  與  $m_L$  分別是高、低努力。 $\pi$  表示代理人工作的成果，是代理人幫委託人賺到的利潤。

委託人的報酬函數是利潤減去工資  $\pi - E$ ，代理人的報酬是由工資與努力工作程度決定，寫成  $u(E, m)$ 。試想利潤有  $N$  種可能金額，寫成  $\pi = (\pi_1, \pi_2, \dots, \pi_N)$ ，其中  $\pi_1 < \pi_2 < \dots < \pi_N$ 。工作努力程度與  $m$  與利潤大小之間有隨機關係，努力程度  $m_i$  的代理人賺取不同金額利潤的機率是  $p_i = (p_{i1}, p_{i2}, \dots, p_{iN})$ ，其中  $i = H, L$  表示努力程度高、低，而  $p_{in}$  則表示  $m_i$  努力代理人賺到  $\pi_n$  利潤的機率， $n = 1, 2, \dots, N$ 。

用一個例子來說明以上架構， $\pi = (0, 200, 600)$  表示代理人創造的利潤可能是 0 或 200 或 600，而  $p_H = (0.1, 0.2, 0.6)$  與  $p_L = (0.5, 0.4, 0.3)$  分別說明  $m_H$  和  $m_L$  代理人賺到不同利潤的機率。由上述資料可算出  $m_H$  與  $m_L$  代理人幫委託人賺到的預期利潤分別是 400 與 260。平均來說， $m_H$  會賺到較高的預期利潤，但是  $m_L$  也有 0.3 的機率會賺到 600 的高利潤。

「資訊不對稱」是代理問題的重點，也就是代理人努力程度是  $m_H$  或是  $m_L$  委託人無法觀察到。所以委託人無法根據代理人實際付出的努力來付工資，而只好由利潤  $\pi$  來決定工資的大小。 $m_H$  (努力) 代理人也可能得到低利潤， $m_L$  (不努力) 代理人也可能賺到高利潤。

我們藉由分析代理人的最適決策。面對  $E(\pi)$  工資函數，代理人必定會以追求  $u(E, m)$  極大化為目標來決定以下二項選擇：

- 是否接受工作。

- (若接受) 選擇何種程度的努力來工作。

所以委託人設計  $E(\pi)$  時必需將代理人對  $E(\pi)$  的最適反應考慮在內。我們將委託人的契約設計(contract design)問題，分為下二步驟解決：

- 先假設委託人的目標是希望代理人選擇  $\hat{m}$  努力程度，則最佳工資契約

$E^*(\pi)$  應該滿足下二條件：

- 個人理性條件 (individual rationality, IR) :

拒絕工作的代理人將獲得保留效用水準  $\underline{u}$ ，而在  $E^*(\pi)$  工資函數下，代理人接受契約並用  $\hat{m}$  的努力程度工作得到的效用水準。最佳工資契約必需使得  $u(E^*(\pi), \hat{m}) \geq \underline{u}$ 。如果此條件不滿足，理性的代理人將會選擇拒絕工作。

- 誘因相容條件 (incentive compatible, IC) :

以  $\tilde{m}$  表示其他任意努力程度， $\tilde{m} \neq \hat{m}$ 。在  $E^*(\pi)$  之下，代理人用  $\tilde{m}$  的努力程度工作會得到  $u(E^*(\pi), \tilde{m})$  效用，工人用  $\hat{m}$  的努力程度工作會得到  $u(E^*(\pi), \hat{m})$ 。最佳工資契約必需使得  $u(E^*(\pi), \hat{m}) \geq u(E^*(\pi), \tilde{m})$ ，才能確定代理人會選擇  $\hat{m}$  的努力程度。滿足此條件後，對代理人最好的是選擇  $\hat{m}$  的努力程度工作，這正好和委託人的目標一致，因此稱為誘因相容條件。

一套有效用的激勵機制必須符合誘因相容與個人理性條件。誘因相容就是機制設計者同時思考個人追求與社會追求，使得追求個人利益的行為，正好與社會實現集體價值最大化的目標相吻合。而個人理性條件是指代理人接受合同下的期望收益(效用)要大於其他市場機會下能獲得的最大期望收益。

## 2.4.2 委託代理理論基本模型

委託代理理論是建立於非對稱訊息賽局的基礎上，而非對稱訊息 (asymmetric information) 指的是部份參與人擁有但另一部分參與人不擁有的訊息。從非對稱發生的時間看，可分別稱為事前非對稱和事後非對稱。事前非對稱訊息賽局的模型稱為逆向選擇模型 (adverse selection)，事後非對稱訊息的模型稱為道德風險模型 (moral hazard)。前者是委託人與代理人的資訊不對稱所造成；後者原因為簽訂契約雙方一些行動的不可觀察性或無法驗證性而完成。委託代理理論的任務是研究在利益衝突和訊息不對稱的環境下，委託人如何設計最好契約來激勵代理人來達成委託人想要達成的目標。以下簡述幾個較為著名的委託代理理論模型。

- 代理模型

最早研究委託代理模型的是 Radner, 1981 和 Rubbinstein, 1979。其透過重複賽局模型證明，若委託人和代理人有長期的合作關係時，代理人無法用偷懶的方式提高自身的福利且由於在長期契約的保護下向代理人提供了“個人保險” (self-insurance)，委託人可以免除代理人的風險。即使合同不具法律上的可執行性，出於聲譽的考慮，合同雙方都會各盡義務。

- 效應模型

棘輪效應 (ratchet effects) 指的是委託人以代理人過去的成績作為標準，因為過去的成績包含著有用的訊息。過去的成績與代理人的主觀努力有關，代理人越努力，好成績可能性越大，代理人給自己的“標準”也越高。當代理人意識到努力帶來的結果是“標準”的提高，其努力的積極性將會降低。這種標準成績上升的傾向稱為“棘輪效應”。在長期的過程中，棘輪效應會抵銷激勵機制。

- 風險模型

於委託代理人模型中，主要討論的是代理人道德風險的問題。是實

上，委託人也同樣存在道德風險。其衡量的方式存在很大的主觀意識。代理人可能無法觀察到委託人觀察到的東西。在這種情況下，就存在委託人的道德風險問題。根據契約，當觀察到的產出高時，委託人應該支付給代理人高的報酬，但委託人可以謊報產出不高，進而逃避責任，而把本應支付給代理人的收入佔為己有。相同的如果代理人預設委託人可能耍賴，其將不會積極努力工作。

- 監督模型

存在委託代理關係就無法避免監督問題。事實上，在非對稱訊息的情況下，委託人對代理人訊息的了解程度可以由委託人自己選擇。比如說，通過僱傭監工或花更多的時間和精力，委託人可以在一定程度上更多的了解代理人的訊息。從而加強對代理人的激勵和監督。但訊息的獲取又是具有成本的，於是，委託人面臨著選擇最優監督力度的問題。

由於委託代理關係於實際社會中是存在的，因此委託代理理論被用於解決各種問題。例如老闆與僱員、股東與經營者，選民與官員，醫生與病人，都是委託代理關係。因此，經由激勵機制的影響，設計最佳的激勵機制，將會越來越廣泛的被應用於生活各方面。

### 2.4.3 激勵效果

從效果的角度而言，一般將激勵的效果分為二項，顯性激勵和隱性激勵兩種。以下將分別說明何謂顯性激勵與隱性激勵及優缺點。

- 顯性激勵（explicit incentive mechanism）

委託人為了激勵代理人選擇委託人所冀望的行動，會依據可以觀察的結果進行獎勵或懲罰代理人。此類型的激勵機制稱為「顯性激勵機制」。

換句話說，顯性激勵是指代理人預期可獲得收入的總和，其包含物質或精神的收入，物質收入例如工資、獎金或其他實物收入，精神方面的補償例如職位的提升與獲得榮譽的肯定等等。在實施顯性激勵的過程中，委

託人將處於被動狀態，其只能以自身的行為方式或結果改變自己所處的狀態。隨代理人目的不同，不同的顯性激勵方式效果也會不同，如榮譽的肯定，對某些人極具補償效果而對另外一些人則不然。

#### ■ 優點

其特點是代理人可以感覺到或預期可得到此項激勵。此種激勵機制是經由代理人所盡力去爭取但結果卻不由代理人決定，透過此方誘導代理人選擇委託人所冀望行動的一種激勵方式。

#### ■ 缺點

顯性激勵機制的缺陷在於它並不能使代理人自願發揮自己的全部內在潛力，從而不能使顯性激勵完全發揮其作用。

### ● 隱性激勵(Implicit incentive mechanism)

隱性激勵，主要是針對公開的顯性收入(物質或精神)之外，採用非公開的隱蔽收入進行激勵的一種方式。例如公關費用(非公開沒有既定標準的各種津補貼)。無“薪”隱性的激勵，則更能體現出委託人領導能力。

#### ■ 優點

隱性激勵具有降低激勵成本(由於隱性激勵收入沒有統一的標準，因此激勵的標準可以靈活調整)、增加代理人的收入、激勵效用能夠持久且兼具標準與獎勵名單不公開，能夠降低代理人們心理不平衡因素。

#### ■ 缺點

##### ◆ 價值導向不明確:

一般而言，激勵的重要作用之一，是通過樹立標竿、獎勵績優，而隱性激勵在某些程度上會使得組織能夠傳遞的訊息幅度有限，甚至連被激勵者本人都不清楚具體的激勵標準。

##### ◆ 激勵效果的不確定性：

由於“隱性”的包袱，成績優異代理人能夠得到激勵，卻不能

公開其榮譽，換句話說代理人能夠清楚自身的獎勵情況，卻不清楚獎勵的標準及與其他代理人的對比，進而使得努力的方向變得模糊。此類型隱性激勵效果模糊，無法有效激勵成績優異者，亦不能有效激發成績落後者。

◆ 激勵公平性：

一般而言，「公平、公開與公正」為激勵機制的基本要求，而隱性激勵在某些程度上需避免「公平、公開與公正」原則。一旦公平原則受到質疑，其對組織和諧的破壞性更勝於顯性激勵。





## 第三章、研究架構與方法

### 3.1 研究架構

本論文之研究架構如圖 11所示，其中包含了兩個主要部分，分別為使用者輪廓(User Profile)資訊拓展與鼓勵新項目評比：

- 使用者輪廓(User Profile)資訊拓展：

利用貝氏網路事後推論特性更正確找出使用者輪廓中各個有興趣項目種類的潛在關聯性，避免因為關聯規則信賴度門檻定義選擇的問題而使得某些規則遺失使得推薦正確率降低。

例如 以下二條規則

冒險片 (Adventure) -> 奇幻片 (Fantasy)

奇幻片 (Fantasy) -> 冒險片 (Adventure)

假設

冒險片 (Adventure) -> 奇幻片 (Fantasy) 規則成立

但

奇幻片 (Fantasy) -> 冒險片 (Adventure) 因不符合門檻值而被過濾掉，其被過濾掉的原因為，科幻片 (Sci-Fi) 與 冒險片 (Adventure) 互為替代品，導致出現的次數降低進而使得以下二項規則都會被過濾掉。

奇幻片 (Fantasy) -> 冒險片 (Adventure)

奇幻片 (Fantasy) -> 科幻片 (Sci-Fi)

- 鼓勵使用者對於新項目(影片)進行評比：

透過激勵理論的代理人機制設計滿足個人理性條件與誘因相容條件的激勵模型來最大化其自身(使用者與推薦系統)的期望效用函數，讓使用者願意去嘗試新的物件或產品。

以下將針對研究架構中的兩個核心之細部功能詳細說明之。

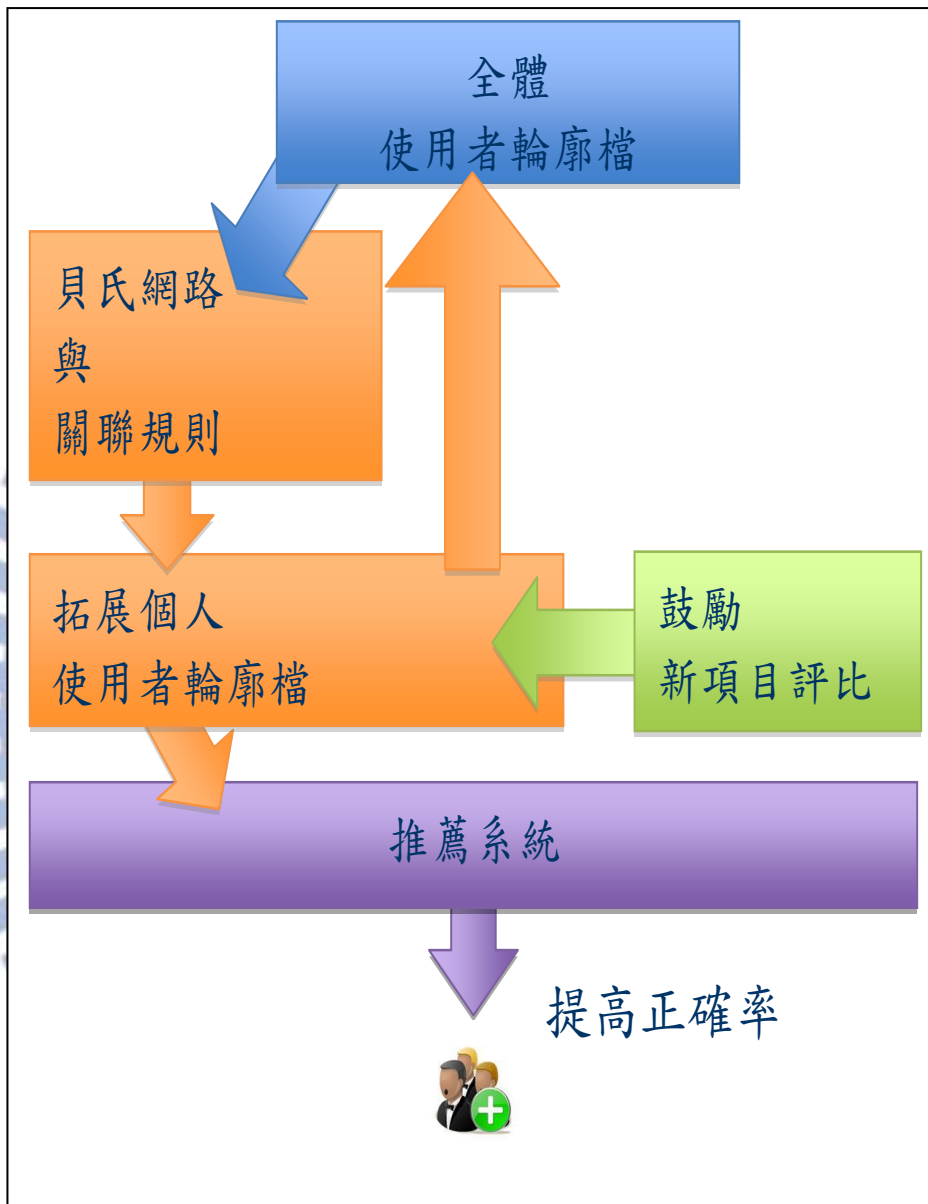


圖 11 研究架構

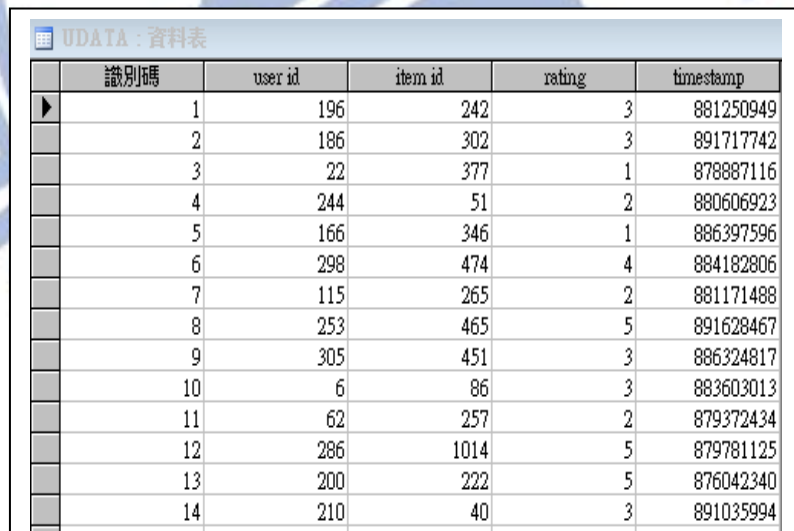
### 3.1.1 使用者輪廓(User Profile)資訊拓展

本論文期望利用全體使用者輪廓檔內項目類別間潛在的關係來拓展使用者輪廓資訊不足的個別用戶。此資訊拓展可分為三部分，第一部分為將輪廓檔內的項目進行分類以利第二部分關聯規則的探勘並建置貝氏網路結構。第三部分則是貝氏網路機率表的產生與拓展使用者輪廓檔。以下對三部份進行說明。

#### 3.1.1.1 輪廓檔項目分類

本論文將利用網路電影資料庫(Internet Movie Database, 簡稱 IMDB)的分類方式進行分類，並將其轉換為交易紀錄檔型態(表 4)，其項目分類有評比過的為 1，反之則紀錄為 0。完成轉換後則利用「Apriori」演算法從表 4 的矩陣中分析出各項目類別間的關聯規則。原始使用者輪廓檔內如圖 12 所示，

- user id：為使用者 ID。
- item id：為電影名稱 ID。
- rating：為使用者對於電影的評價程度。



識別碼	user id	item id	rating	timestamp
1	196	242	3	881250949
2	186	302	3	891717742
3	22	377	1	878887116
4	244	51	2	880606923
5	166	346	1	886397596
6	298	474	4	884182806
7	115	265	2	881171488
8	253	465	5	891628467
9	305	451	3	886324817
10	6	86	3	883603013
11	62	257	2	879372434
12	286	1014	5	879781125
13	200	222	5	876042340
14	210	40	3	891035994

圖 12 使用者輪廓檔

表 4 項目分類交易紀錄檔

使用者 ID (Transaction ID)	Action	Adventure	Animation	Comedy	Crime	...
1	0	1	1	0	0	...
2	0	0	1	0	1	...
...	...	...	...	...	...	...
943	1	1	1	0	1	...

### 3.1.1.2 建構貝氏網路結構

由現存的資料中建構出"最佳化"的貝氏網路結構，此問題已被證明為 NP-Hard 的問題。因此，本論文除了希望利用貝氏網路事後機率推論的優點來避免信賴度門檻定義選擇的問題外，更近一步的利用關聯規則的結果來協助貝氏網路的建立。在建立關聯規則之後，則利用以法則條件(Rule Body)與法則結論(Rule Head)轉換為父代與子代結點的方式[CI]逐條檢視並建構出完整的貝氏網路架構。

透過法則轉換為父代與子代結點的方式須特別注意「法則的長度」，因法則長度將影響貝氏網路建構時的複雜度。所謂的法則長度，是指法則條件(Rule Body)的個數，如法則 $\{A, B\} \rightarrow C$ ，法則長度為 2。若法則長度過長，則項目之間的關係將會變得複雜，在轉換建構貝氏網路過程中時可能性亦會變多。

因此，為使網路架構簡單，本論文將法則長度上限訂為 2。以下舉例說明法則長度 1 跟 2 與構成網路結構(完整圖形)的情況：

- 法則長度 1

法則長度為 1 時，法則條件屬性視為關聯中父代節點，法則決策屬性則是視為關聯中的子代節點如圖 13 所示。

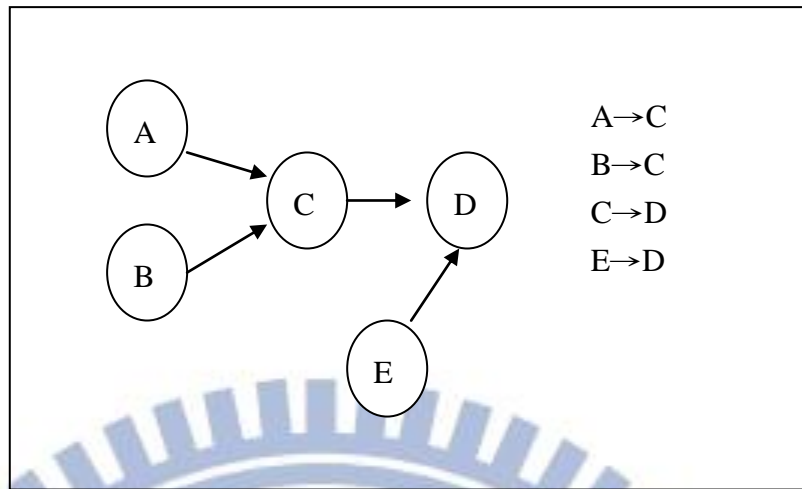


圖 13 貝氏網路架構法則長度 1

● 法則長度 2

法則長度為 2 時，仍是以法則條件屬性視為關聯中父代節點，法則決策屬性則是視為關聯中的子代節點方式建構。但其會有以下二種不同的結構：

- 鏈狀結構：若條件屬性中的兩屬性，也有法則存在時形成此結構，如圖 14

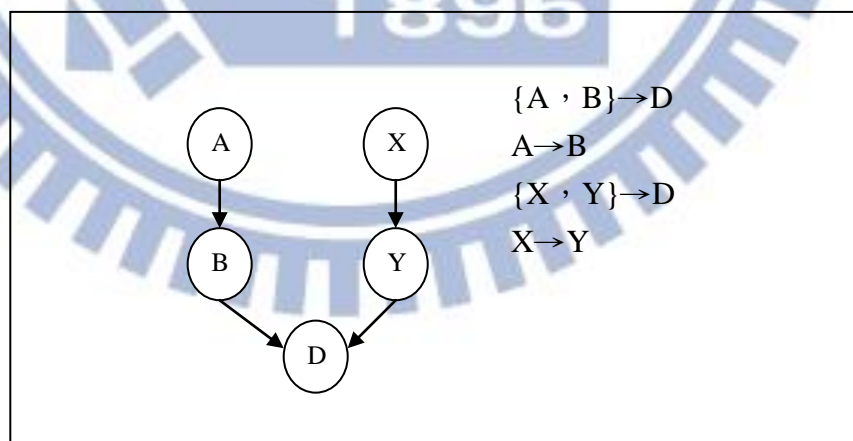


圖 14 鏈狀結構

- 收斂結構：若條件屬性中的兩屬性，沒有法則存在時形成此結構，如圖 15

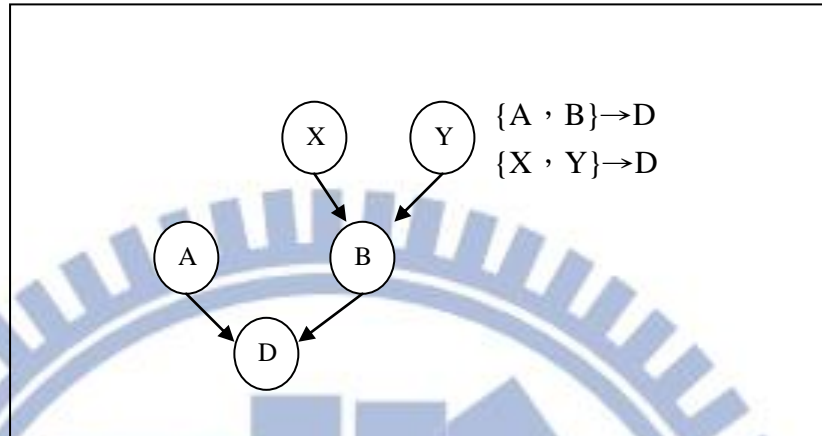


圖 15 收斂結構

- 構成網路結構：

若任三個節點出現完整圖形時，如圖 16。會存在條件獨立的情形，此時計算子節點與父節點之間的「條件熵(Conditional entropy)」值（或稱為 unexplained information 值）的大小，並透過去除「條件熵」值較低的關聯，以簡化貝氏網路。

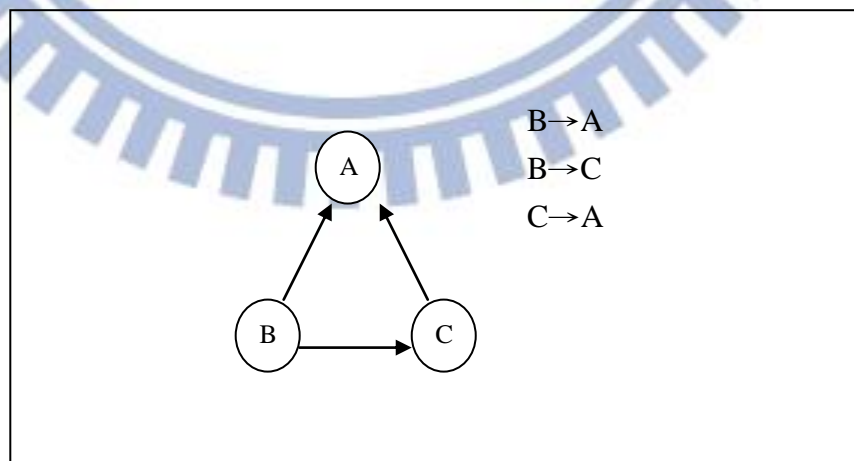


圖 16 網路結構(完整圖形)

### 3.1.1.3 產生條件機率表

透過法則條件(Rule Body)與法則結論(Rule Head)轉換為父代與子代結點的方式逐一檢視關聯規則並建構完成貝氏網路的架構後，接下來即可透過貝氏網路聯合機率分配定理計算圖中每個節點的機率表，如圖 17所示。

$$p(x) = \prod_{i=1}^n p(x_i | \text{parents}(x))$$

表示為在父節點發生的前提下， $x$ 發生的機率

$\text{Parents}(x)$ :  $x$  節點父節點之機率。

圖 18為貝氏網路架構與機率示意圖，以下說明圖中各節點所表示的意義。

- 圖中橢圓表示推薦系統中各項目(於本論文中為電影類別)間之關係。
- 灰色橢圓節點(Fantasy、Comedy、War)則表示為使用者輪廓檔中已評比之項目。
- 節點數字部分表示在灰色橢圓節點發生的前提下，各節點的機率值。
- 虛線節點(Adventure、Action、Thriller)則表示其機率大於所設定的門檻值(此處假設為 0.7)，超過門檻值之節點將會添加至使用者輪廓資訊不足的個別用戶中。

建構完成貝氏網路架構與機率計算後即可透過貝氏網路所推論之項目類關係添加至使用者輪廓資訊不足的個別用戶中，其方式如圖 19、圖 20所示

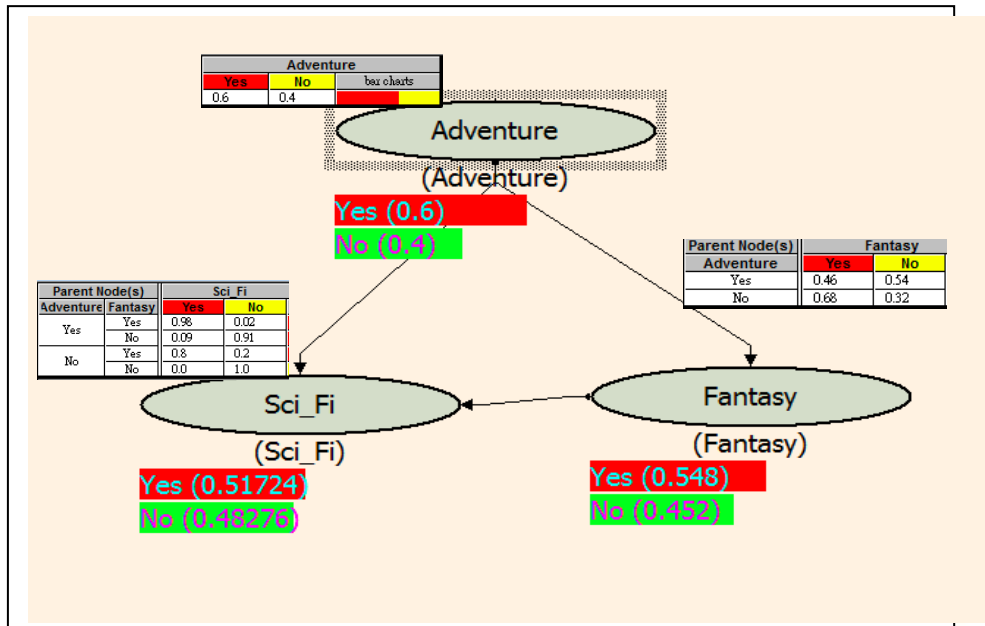


圖 17 氏網路架構圖及機率表





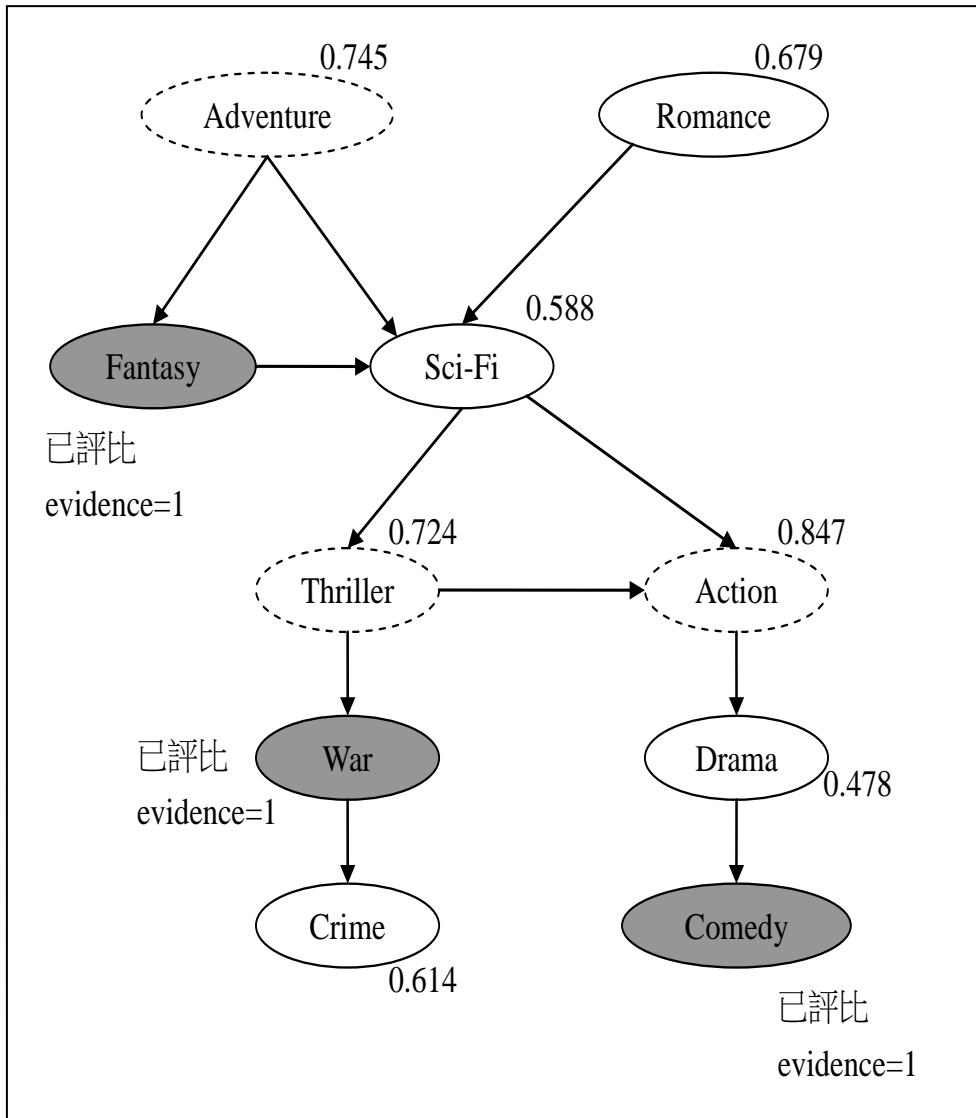
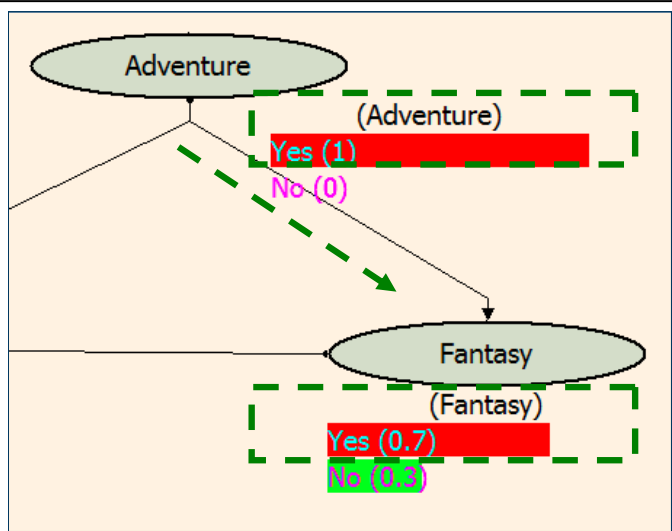


圖 18 貝式網路機率狀況示意圖



在 Adventure 已發生的前提下，Fantasy 發生的機率超過門檻值(0.7)

規則：Adventure（冒險片），==> Fantasy（奇幻片）

資訊不足之輪廓檔(User 20)

0, Adventure（冒險片）, 0,0,0, 0, 0,0, 0, 0, 0, 0, 0, 0, 0, 0, War, 0



拓展輪廓檔(User 20)

0, Adventure（冒險片）, **Fantasy（奇幻片）**, 0,0, 0, 0,0, 0, 0, 0, 0, 0, 0, War, 0

圖 19 使用者輪廓拓展

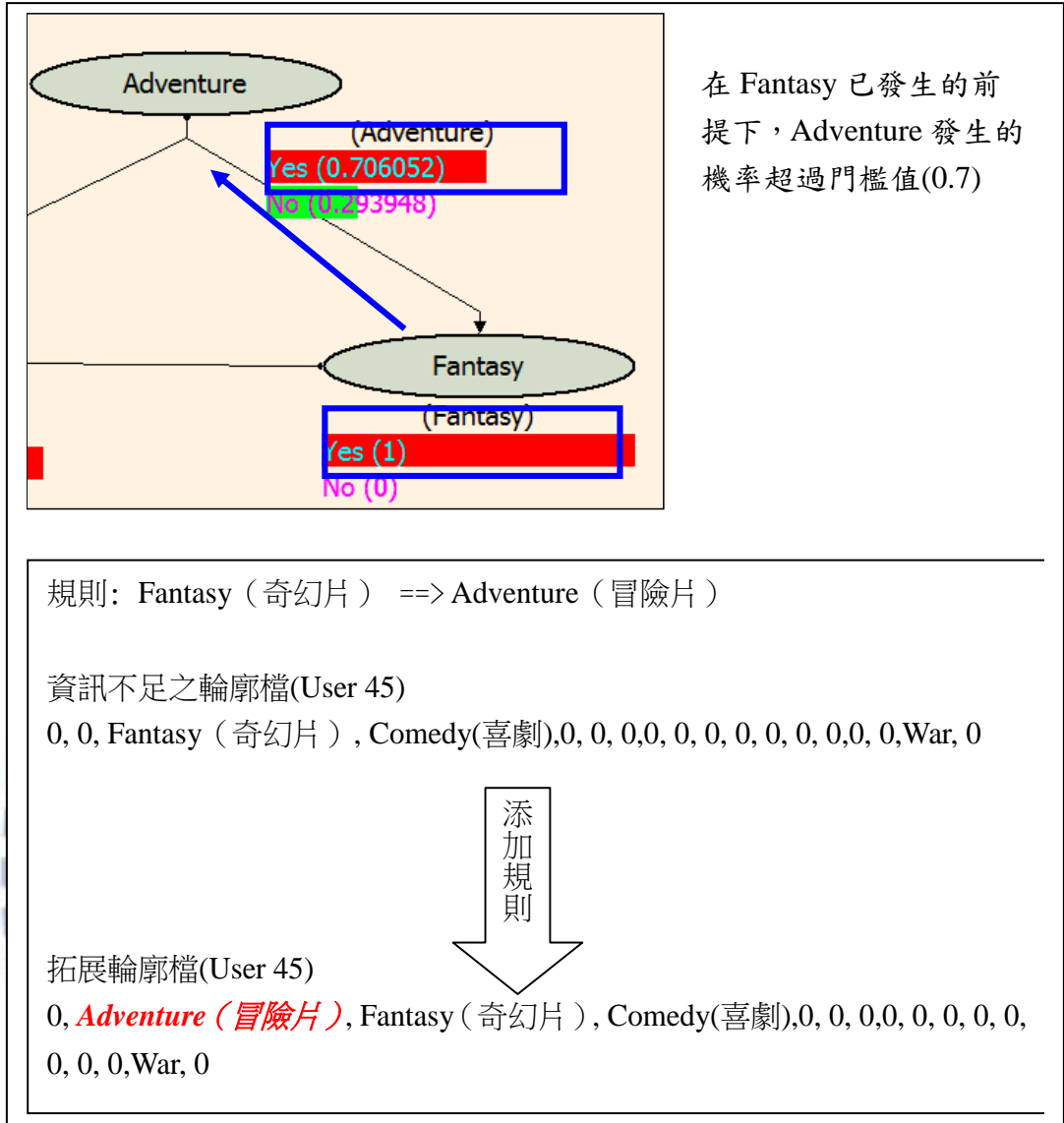


圖 20使用者輪廓拓展使用事後推論

### 3.1.2 鼓勵使用者對於新項目(影片)進行評比

激勵模型(incentive model)又稱經濟模型，主要目的在於刺激與鼓勵參與者積極主動參與活動。本節將描述本論文激勵機制的設計並分析激勵模型在影片推薦系統方面任務。

為了方便說明起見，將先定義本論文中激勵機制所使用的共用符號(notation)。

1. 參與者(Player):  $P_1, P_2, \dots, P_n$ 。
2.  $n$  個使用者其所得之結果(Outcome)為集合  $A$ 。
3. 每一位參與者  $P_i \in \{P_1, \dots, P_n\}$ ，皆有  $r_i$  的資訊，其表示對於推薦項目評比的資料。定義  $R$  為  $r_i$  向量的集合。
4. 每一位參與者  $P_i \in \{P_1, \dots, P_n\}$ ，選擇  $r'_i$ ，其中  $r'_i$  可能與  $r_i$  不相等，定義  $R'$  為  $r'_i$  向量的集合。
5. 向量  $R = (r_1, r_2, \dots, r_n)$ 。
6. 向量  $R_{-i} = (r_1, r_2, \dots, r_{i-1}, r_{i+1}, \dots, r_n)$ 。即向量  $R$  中去除  $r_i$ 。

舉例來說，假設評比項目共有 10 項， $r_i$  表示使用者  $i$  將會評比「全部」10 個項目，而  $r'_i$  則表示使用者  $i$  只願意評比「最多」10 個項目。

7. 參與者將提供  $R'$  給與推薦系統，產生推薦建議，將此推薦系統模式(model)稱為  $S$ 。而  $acc(S(R'))$  稱為此模式的正確率。
8. 假設選擇向量  $R$  參與推薦的正確率將大於或等於選擇  $R'$  參與推薦的正確率， $acc(S(R)) \geq acc(S(R'))$ 。亦即願意評比「全部」項目的推薦正確率將大於或等於只願意評比「部分」項目的推薦正確率。
9. 假設選擇向量  $R$  參與推薦的正確率將大於或等於只有  $R_{-i}$  參與推薦的正確率， $acc(S(R)) \geq acc(S(R_{-i}))$ 。亦即「全部」使用者皆參與評比，其推薦之正確率將大於或等於只有「部分」使用者參與評比之推薦正確率。

10. 每一位參與者(Player):  $P_1, P_2, \dots, P_n$  , 其效用函數為

$$u_i(r_i, S(R')) = \text{Max}\{v_i(S(R')) - v_i(S(R'_{-r'_i})), 0\}$$

$u_i(r_i, S(R'))$  表示使用者  $i$  在所有參與者提供  $R'$  給推薦系統時所獲得之

效用函數, 圖 21 表示使用者  $i$  的預期支付, 圖 22 為鼓勵分享示意圖。

$v_i(S(R'))$ : 在提供  $R'$  於推薦系統時的結果(Outcome)

$v_i(S(R'_{-r'_i}))$ : 無  $i$  時推薦系統時的結果(Outcome)

上述效用函數其需符合以下二個條件

- 個人理性條件(individual rationality)

$$u_i(r_i, S(R)) \geq 0$$

代表使用者在「參與」此機制下得到的期望效用必須不小於「不參與」這個機制所得到的最大期望效用。

- 誘因相容條件(incentive compatible)

$$u_i(r_i, S(R)) \geq u_i(r'_i, S(R')), \text{ 其中 } R = r_i \cap R_{-r_i}, R' = r'_i \cap R'_{-r'_i}$$

代表使用者選擇參與「全部」項目評比所得到的期望效用將不小於選擇參與「部分」項目評比所得到的期望效用。

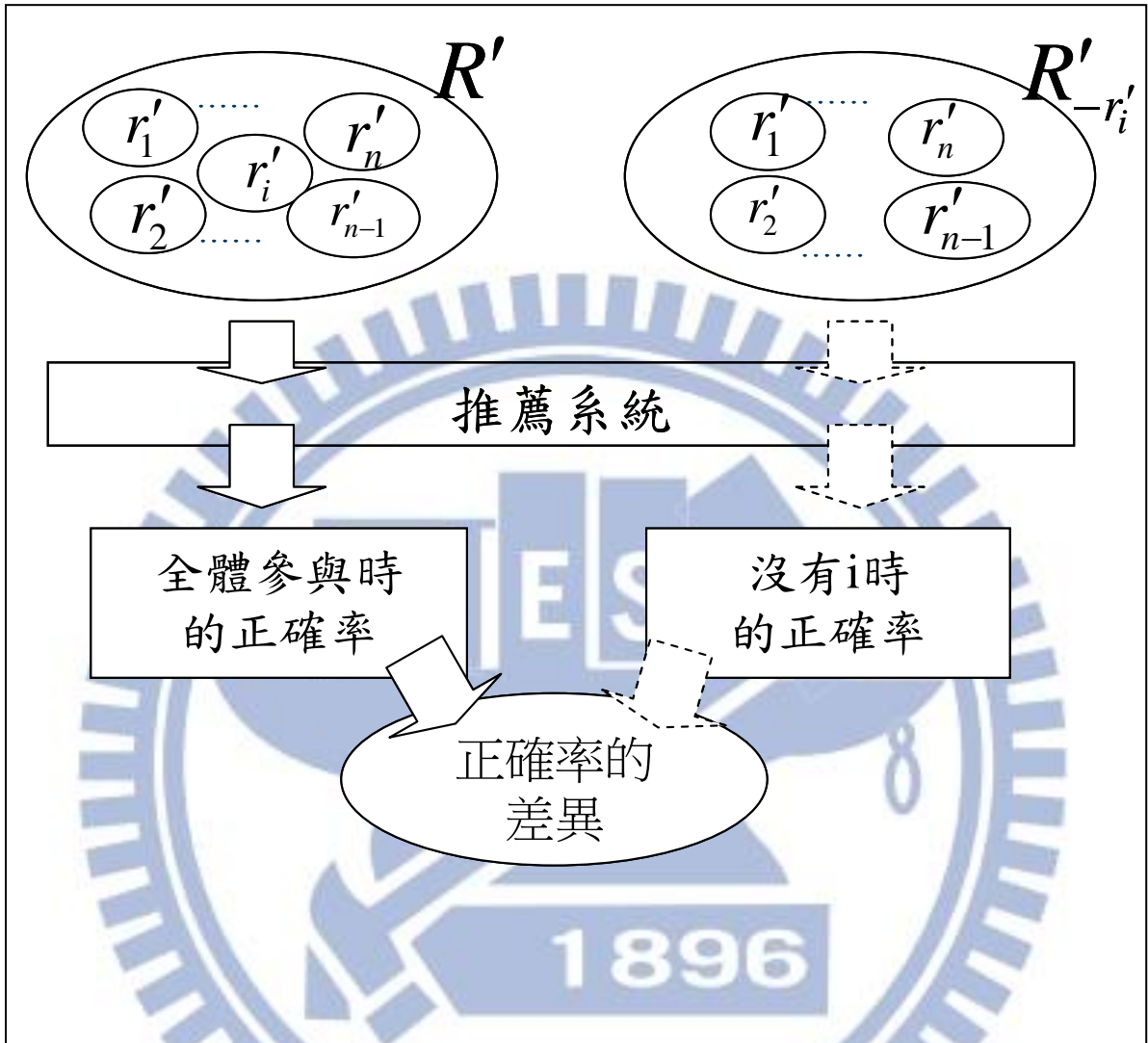


圖 21 使用者  $i$  的預期支付

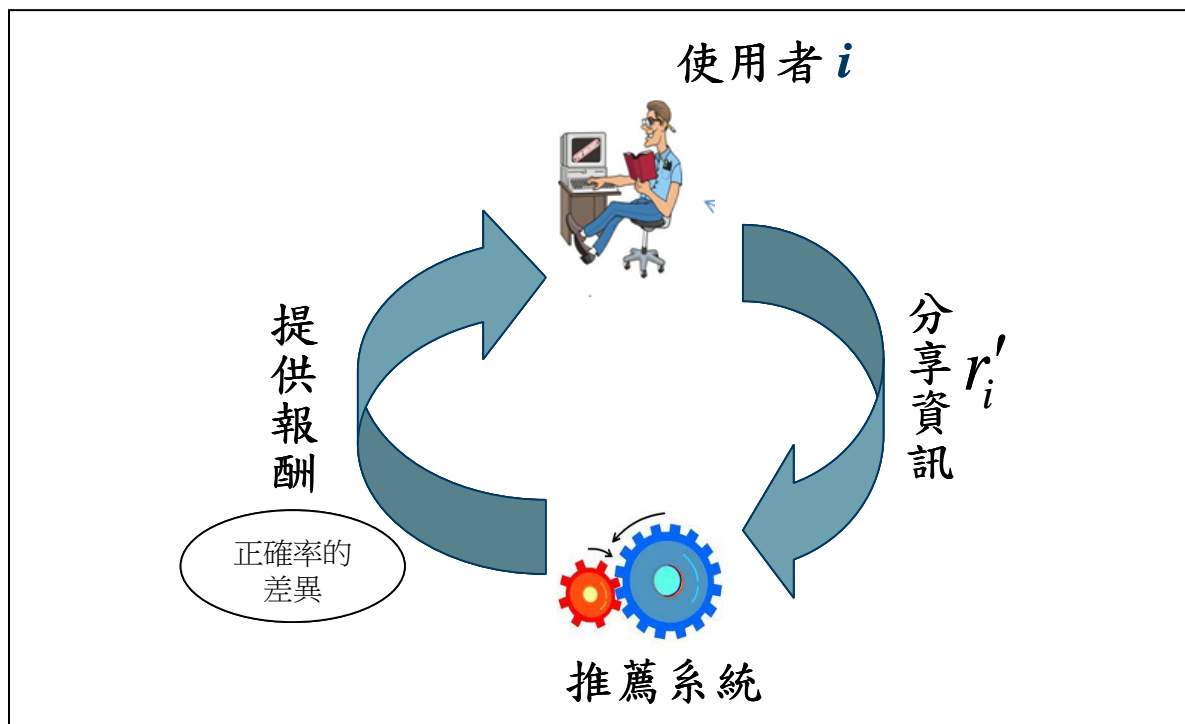


圖 22 鼓勵分享示意圖



### 3.1.2.1 IR 條件 (individual rationality 個人理性)

本節將說明參與者的效用函數相符合個人理性(individual rationality)條件。即參與者看到效用函數後，決定是否接受或拒絕參影片評比。如果拒絕，參與者得到最低保留效用水準  $\underline{u}$ ，並規範(normalize)每個使用者的保留效用，若參與者選擇不參加評比時其效用函數為 0。

證明  $u_i(r_i, S(R)) \geq 0$

$$u_i(r_i, S(R)) = \text{Max}\{v_i(S(R)) - v_i(S(R_{-i})), 0\}$$

由於  $\text{Max}\{v_i(S(R)) - v_i(S(R_{-i})), 0\}$  至少為 0，

且  $\text{acc}(S(R)) \geq \text{acc}(S(R_{-i}))$ ，前提假設選擇向量  $R$  參與推薦的正確率將大於或等於只有  $R_{-i}$  參與推薦的正確率。所以此機制符合個人理性條件(Individual Rationality)

### 3.1.2.2 IC 條件 (incentive compatible 誘因相容)

本節將說明參與者的效用函數相符合誘因相容(incentive compatible)條件。最佳效用函數必需使得  $u_i(r_i, S(R)) \geq u_i(r'_i, S(R'))$ ，才能確定參與者會選擇  $R$  的影片評比程度，亦即滿足此條件後，在此機制之下對參與者最好的情況是選擇  $R$  的影片評比程度。

證明  $u_i(r_i, S(R)) \geq u_i(r'_i, S(R'))$ ，其中  $R = r_i \cap R_{-r_i}$  且  $R' = r'_i \cap R'_{-r'_i}$ 。

$$u_i(r_i, S(R)) = \text{Max}\{v_i(S(R)) - v_i(S(R_{-r_i})), 0\} \geq$$

$$u_i(r'_i, S(R')) = \text{Max}\{v_i(S(R')) - v_i(S(R'_{-r'_i})), 0\}$$

$\text{Max}\{v_i(S(R)) - v_i(S(R_{-r_i})), 0\}$  與  $\text{Max}\{v_i(S(R')) - v_i(S(R'_{-r'_i})), 0\}$  皆大於等於 0，且因  $\text{acc}(S(R)) \geq \text{acc}(S(R'))$ ，前提假設選擇向量  $R$  參與推薦的正確率將大於或等於選擇  $R'$  參與推薦的正確率。  $u_i(r_i, S(R)) \geq u_i(r'_i, S(R'))$  所以此機制符合誘因相容(incentive compatible)條件。



## 第四章、實驗設計與結果

本章將分為三個部分敘述：首先說明測試資料來源、實驗環境與實驗過程中所採用的相關應用軟體，第二部分說明使用者輪廓(User Profile)資訊拓展實驗的架構方式，最後說明鼓勵使用者評比實驗的架構方式。

### 4.1 測試資料來源與實驗環境

測試資料方面本論文採用 MovieLens 資料集，其中包含的使用者，電影和由使用者對電影的評分等三個部分。共有 943 位使用者，1,700 部電影與 100,000 筆評比資料。(MovieLens, <http://www.grouplens.org/node/73>)

在電影分類方面則採用網路電影資料庫(Internet Movie Database, 簡稱 IMDB, <http://www.imdb.com/genre/>)分類方式，其分類方式如下所述：

- 動作片(Action)
- 冒險片(Adventure)
- 動畫片(Animation)
- 喜劇片(Comedy)
- 犯罪片(Crime)
- 紀錄片(Documentary)
- 劇情片(Drama)
- 家庭片(Family)
- 奇幻片(Fantasy)
- 黑色片(Film-Noir)
- 恐怖片(Horror)
- 音樂片(Music)
- 歌舞片(Musical)
- 懸疑片(Mystery)
- 愛情片(Romance)
- 科幻片(Sci-Fi)
- 短片(Short)

- 運動片(Sport)
- 驚悚片(Thriller)
- 戰爭片(War)
- 西部片(Western)

為了拓展使用者輪廓(User Profile)資訊，本次實驗透過 MovieLens 資料集與網路電影資料庫分類方式，建了一個 (transactional dataset)其中包含了 943 位使用者(transactios)與 19 種電影分類，資料集格式如表 5所示。

在評估推薦正確率方面本論文採用精確率 (Precision)、涵蓋率 (Coverage) 與 F1 指標(F1-Measure)來確定整體推薦系統的性能，並將使用者輪廓(User Profile)資訊隨機分為訓練集(80%)和測試集(20%)二部分。本論文實驗中所使用的軟硬體開發環境與平台如表 6所述。

表 5 MovieLens 交易資料集

使用者 ID (Transaction ID)	Items
1	unknown,Action,Adventure,Animation,Childrens,Comedy,Crime,Documentary,Drama,Fantasy,Film_Noir,Horror, Musical,Mystery,Romance,Sci_Fi,Thriller,War,Western
2	0,Action,Adventure,Animation,Childrens,Comedy,Crime,0,Drama,Fantasy,Film_Noir,Horror,Musical,Mystery,Romance,Sci_Fi,Thriller,War,0
...	...
943	0,Action,Adventure,Animation,Childrens,Comedy,Crime,0,Drama,Fantasy,0,Horror,Musical,Mystery,Romance,Sci_Fi,Thriller,War,Western

表 6 軟硬體開發環境與平台

硬體設備	Intel Pentium 4 CPU 3.40GHz / 1.0GB RAM
作業系統	Microsoft Windows XP Professional; Linux
程式語言	ASP,Java
資料庫	Microsoft Access ; Text File
關聯規則軟體	Weka 3.9.6
貝氏網路軟體	MSBNX
推薦系統基礎架構	Apache Mahout ; Taste

## 4.2 關聯規則相關工具

WEKA (Waikato Environment for Knowledge Analysis) 是由 University of Waikato 使用 Java 開發的資料探勘常用軟體，WEKA 為一個公開的資料探勘工作平臺，集合了大量資料探勘任務的機器學習算法，包括對數據進行預處理，分類、聚類、關聯規則等，並且完成了圖形化的使用者介面。

WEKA 系統已得到了學術領域廣泛的認可，被譽為資料探勘和機器學習歷史上的里程碑，是現今最完備的資料探勘工具之一，本論文中關聯規則的建立即是利用 WEKA 開發工具來協助完成。WEKA 可以提供之功能簡述如下(圖 23 為 WEKA 操作介面)：

- 資料預處理。
- 分群(Cluster)：包含 K-means、EM (Estimation Maximization)、CLOPE、FarthestFirst..... 等等。
- 分類(Classify)：包含 Tree(ID3、Radom Tree)、bayes..... 等等。
- 關聯規則(Association Rule)：Apriori、FPGrowth、Predictive Apriori、Filtered Associator..... 等等。

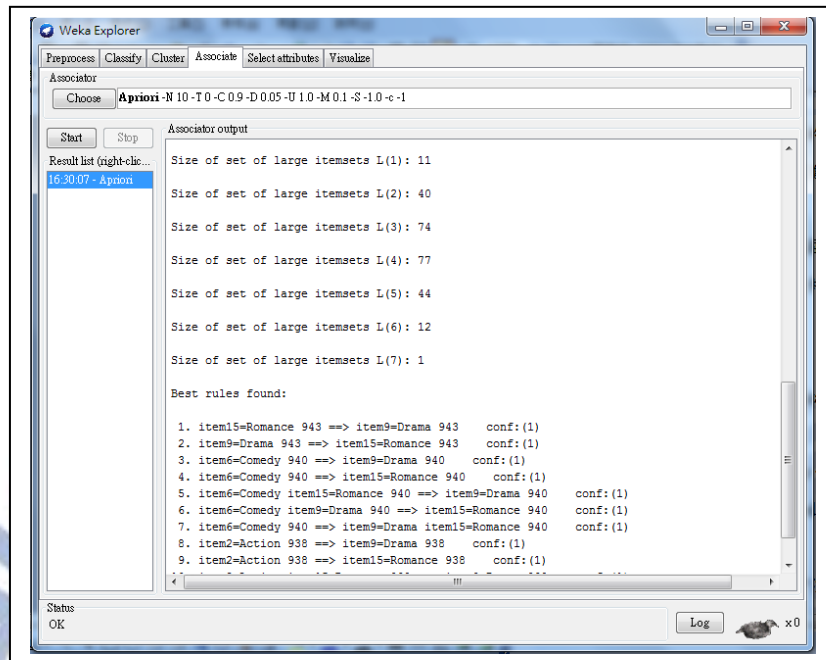


圖 23 WEKA 開發工具

### 4.3 貝氏網路相關工具

目前已有許多研究單位投入研發 Bayesian Belief Network 的開發工具，如由微軟所開發的 MSBNx(Microsoft Bayesian Network Editor)、由英國 Open University 所發展的 BKD(Bayesian Knowledge Discoverer)、Norsys 軟體組織發展的 Netica 及 Pittsburgh University 決策系統實驗室開發的 GeNIE 等等…。這些工具的主要功能有提供建立貝氏網路的介面與貝氏推論引擎。藉由這些工具，使用者可以很快速的建立貝氏網路並進行推論的工作。

微軟所開發的 MSBNx 則有提供一套完整的程式庫，讓使用者可以很方便的利用 Visual Basic 及 C/C++ 等程式語言來控制及操作貝氏網路，此外 MSBNx 也支援將所建構之貝氏網路轉化成 XML 格式，以方便後續程式的開發與操控，圖 24 為 MSBNx 操作畫面。本論文採用微軟所開發的 MSBNx 開發工具來協助貝氏網路的建立。

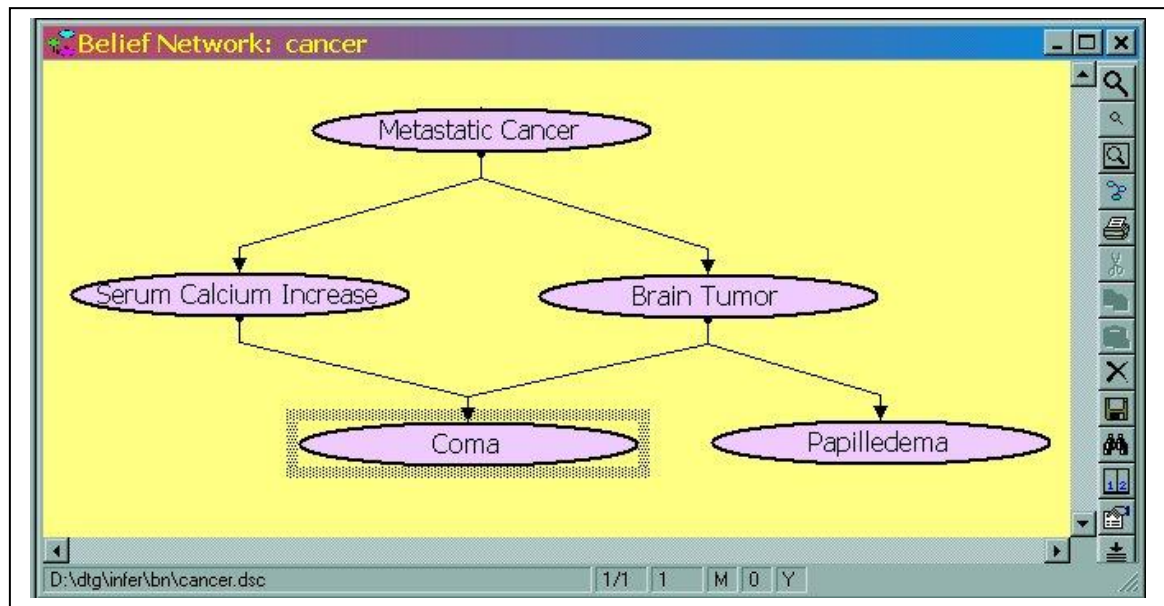


圖 24 MSBNx 操作畫面

#### 4.4 推薦系統相關工具

Apache Mahout 是 Apache Software Foundation(ASF)的一個開放原始碼專案，提供可擴展機器學習領域的經典演算法，幫助開發人員更加方便快捷地開發應用程序。其演算法包括聚類、分類、協同過濾等等

Taste 是 Apache Mahout 所提供的一個協同過濾演算法，它是一個利用 Java 開發的高效率推薦引擎。Tast 實現了最基本「基於用戶」和「基於內容」的推薦演算法，同時也提供了擴展介面，使用戶可以方便的定義和實現自己的推薦演算法。同時，Taste 不僅僅適用於 Java 應用程序，其設計亦使它能滿足企業對推薦引擎在性能、靈活性和可擴展性等方面的要求。本論文採用 Apache Mahout 之基礎架構與 Taste 所提供之 GenericUserBasedRecommender 演算法來協助推薦系統的建立。

Taste 由以下五個主要的組件組成，其系統結構如圖 25所示[18]：

- DataModel：

DataModel 是使用者資訊輸入的抽象介面，其支持從任意類型的資料來源抽取使用者資訊輸入。 Taste 分別支持從資料和文件中讀取用戶的

資訊 ( JDBCDataModel 和 FileDataModel )。

- UserSimilarity 和 ItemSimilarity :

UserSimilarity 用於定義兩個用戶之間的相似度，ItemSimilarity 則是用於定義兩個項目之間的相似度的界面。

- UserNeighborhood :

用於基於用戶相似度的推薦方法，推薦的內容是基於找到與當前用戶喜好相似的“鄰居用戶”。

- Recommender :

Recommender 是推薦引擎的抽象介面，Taste 中的核心組件。

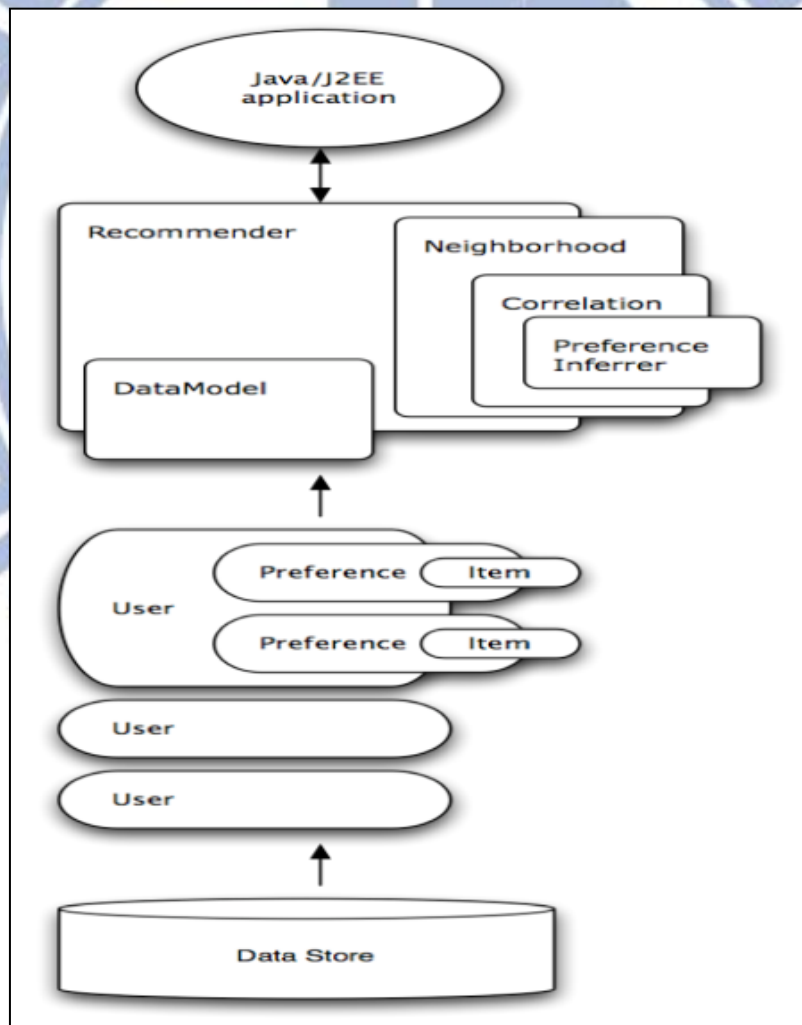


圖 25 Taste 系統結構

## 4.5 使用者輪廓(User Profile)資訊拓展實驗

實驗分二部分，首先分別利用關聯規則與貝氏網路拓展原始使用者輪廓檔，第二部分再分別將原始使用者輪廓檔與已拓展使用者輪廓檔(關聯規則與貝氏網路)輸入推薦系統進行精確率 (Precision)、涵蓋率 (Coverage)、F1 指標的比較。其實驗架構如圖 26所示。

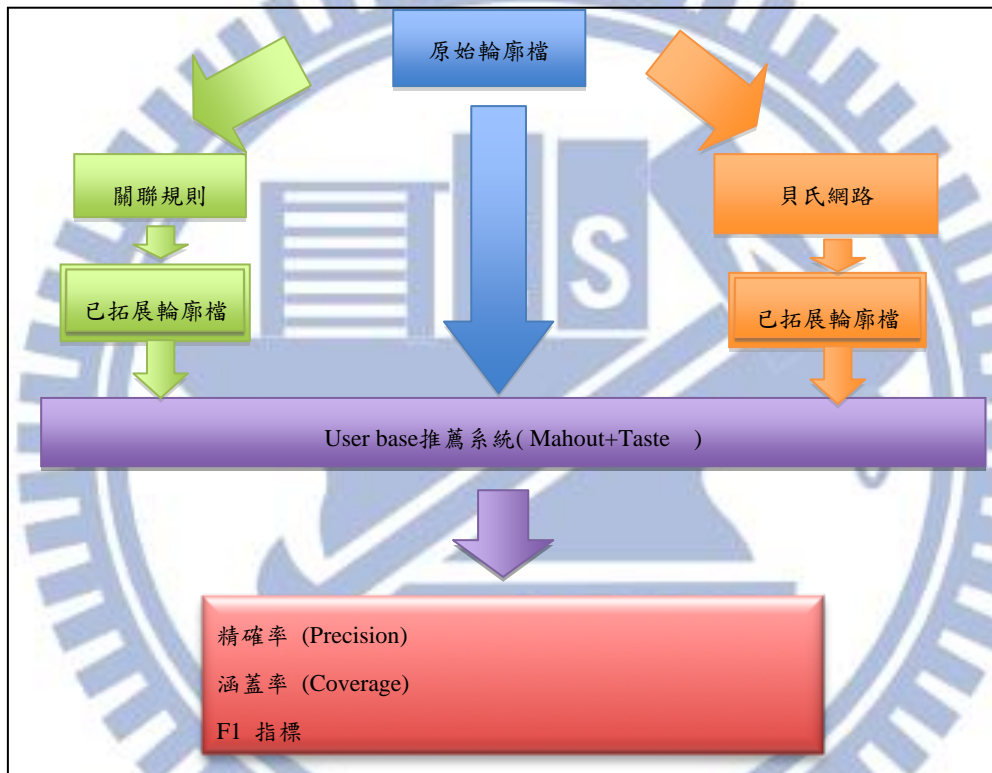


圖 26 資訊拓展實驗架構

### 4.5.1 使用者輪廓拓展

此部分實驗以全體 (943 位) 使用者輪廓為基礎，分別利用關聯規則與貝氏網路找出項目間關聯性，再藉由其關聯性拓展使用者輪廓，方式如下所述：

- 以關聯規則找出使用者輪廓中項目的關聯性：

利用 WEKA (<http://www.cs.waikato.ac.nz/ml/weka/>) 於支持度大於 60%，信賴度大於 70%，增益值大於 3 的條件所得出的 50 條關聯規則，

並進一步篩選出法則長度為三以下之規則，其設定方式如圖 27所示 (weka.associations.Apriori -N 50 -T 1 -C 3.0 -D 0.05 -U 1.0 -M 0.1 -S -1.0 -c -1)，並將所得之結果添加至原始輪廓檔中，其方式如圖 28所示。

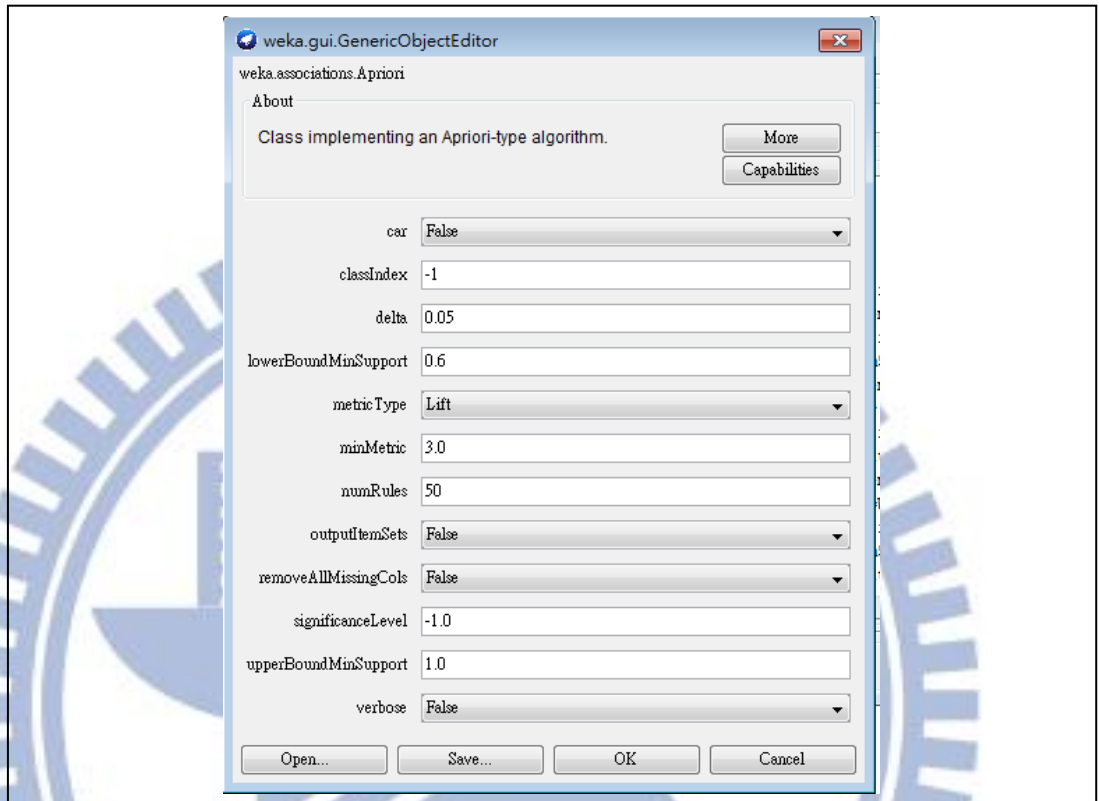


圖 27WEKA 關聯規則參數設定

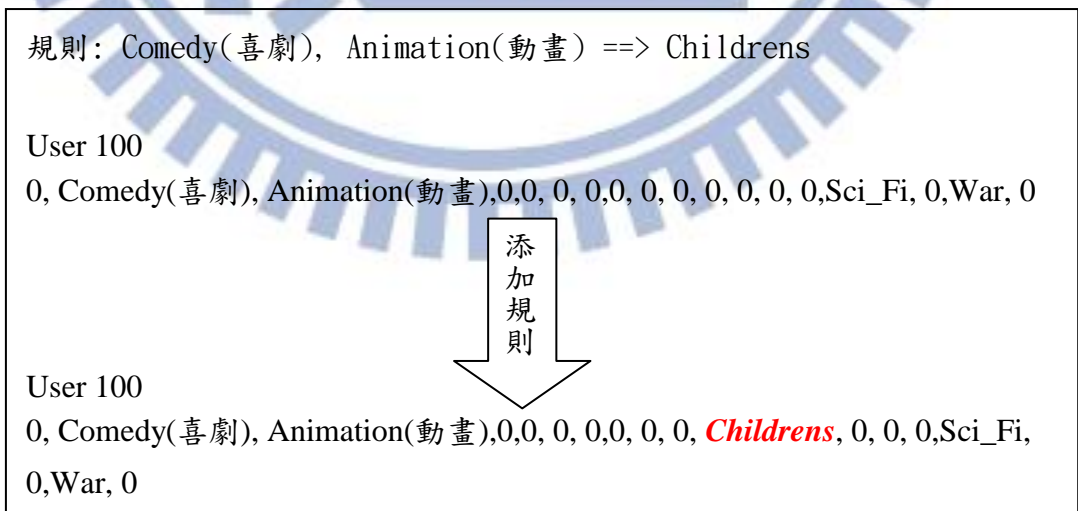


圖 28使用者輪廓拓展



- 以貝氏網路(MSBNX)找出使用者輪廓中項目的關聯性：

以 WEKA 所找出之關聯規則為基礎，並進一步計算各個結點相互發生機率，其結果如圖 29所示，並將門檻值(機率)大於 0.7 的規則添加至原始輪廓檔中(0.7 為本論文之最佳門檻值)，其添加方式同圖 28所示。

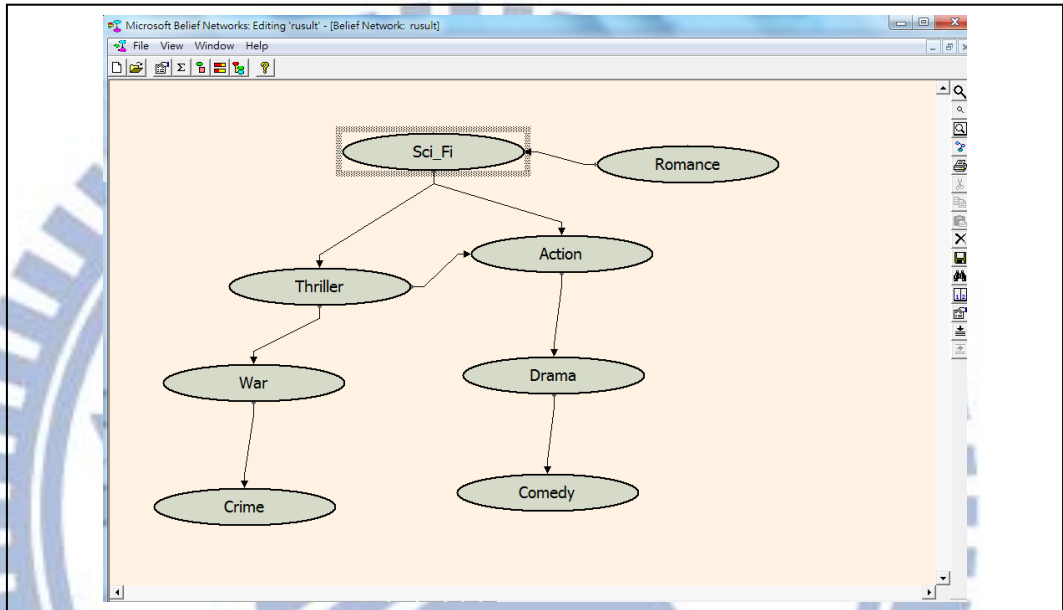


圖 29 貝氏網路關聯

#### 4.5.2 推薦系統效益驗證

將「原始使用者輪廓檔」、「關聯規則使用者輪廓檔」與「貝氏網路使用者輪廓檔」，三種使用者輪廓檔以 mahout 的 taste 推薦系統引擎 (GenericUserBasedRecommender) 分別進行 10 次推薦，並比較其精確率 (Precision)、涵蓋率 (Coverage)、F1 指標 (F1-Measure)。由實驗結果得知，以關聯規則所拓展的使用者輪廓檔於門檻值 0.7 時其精確率與涵蓋率都達到最大如表 7 與圖 30 所示，亦即說明門檻值 0.7 時為本論文之最佳門檻值，以下將基於最佳門檻值條件下進行效益分析。

表 7 門檻值與指標數值表

門檻值	0.3	0.4	0.5	0.6	<b>0.7</b>	0.8	0.9
精確率(Precision)	0.458	0.597	0.624	0.677	<b>0.685</b>	0.774	0.897
涵蓋率(Coverage)	0.854	0.874	0.787	0.621	<b>0.774</b>	0.514	0.347
F1 指標(F1-Measure)	0.596	0.709	0.696	0.648	<b>0.727</b>	0.618	0.500

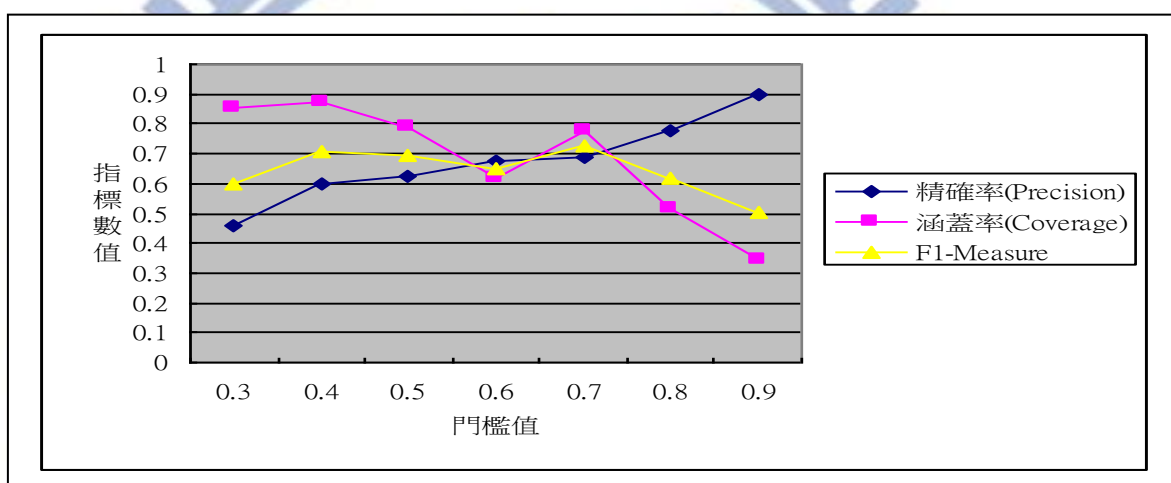


圖 30 門檻值與指標數值圖

使用三種使用者輪廓檔分別進行 10 次推薦之精確率 (Precision)、涵蓋率 (Coverage)、F1 指標(F1-Measure)其結果如表 8 所示，其中以利用貝氏網路所拓展之輪廓檔進行推薦所得之各項指標，精確率(0.754)、涵蓋率(0.807)與 F1 指標(0.780)皆優於其他輪廓檔。

推薦評估指標之提升主要的貢獻來自於貝氏事後推論，由於關聯規則在信賴度指標方面是缺少具反身性的。亦即若項目之間有替代效應或是搭配性，則項目出現次數降低的原因可能是因與其他項目交替出現而被拉低次數，因此若信賴度門檻值定義不佳，可能就會失去某些規則。關於此部分，本論文希望利

用貝氏事後推論的方式來避免這方面的推薦誤差。因此於實驗過程中特別紀錄那一些推論是由父結點(法則條件)而來的，即可知道那一些推薦結論是貝氏網路事後機率推論所貢獻。從實驗結果得知，在符合條件的推薦中，有 26.68% 的推薦皆是來自貝氏網路事後機率推論的結果，亦即使用貝氏事後推論的確可以避免關聯規則在缺少具反身性上導致失去某些規則所造成之推薦誤差。由以上的比較分析可知，本論文所提之推薦方法確有明顯之效益。

表 8三種輪廓檔指標數值表

方法	精確率	%	涵蓋率	%	F1-Measure	%
原始輪廓檔	0.615		0.622		0.618	
關聯規則輪廓檔	0.685	11.38%	0.774	24.44%	0.727	17.51%
貝氏網路輪廓檔	0.754	22.60%	0.807	29.74%	0.780	26.05%

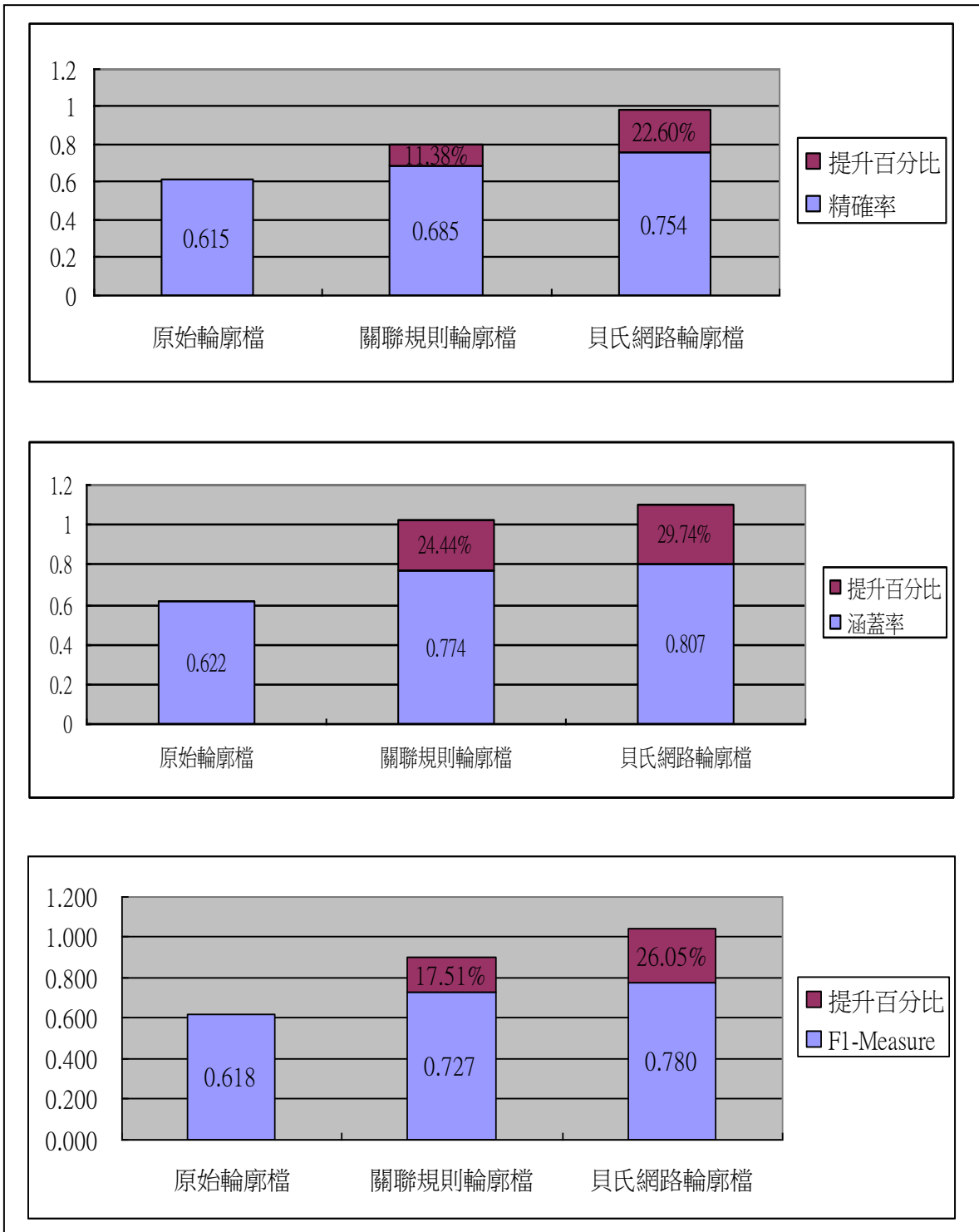


圖 31三種輪廓檔指標數值圖

## 4.6 鼓勵使用者評比實驗

在第三章合理的假設前提之下，已證明該鼓勵使用者評比機制是符合個人理性(individual rationality)條件與誘因相容(incentive compatible)條件。本次實驗分二部分，第一部分為激勵程度計算，利用原始使用者輪廓檔與刪除使用者輪廓檔內已評比過之項目之輪廓檔進行推薦並計算其精確率，第二部分則依據精確率給予使用者不同的激勵程度並觀察使用者重新評比已刪除項目的比例。其實驗架構如圖 32所示。

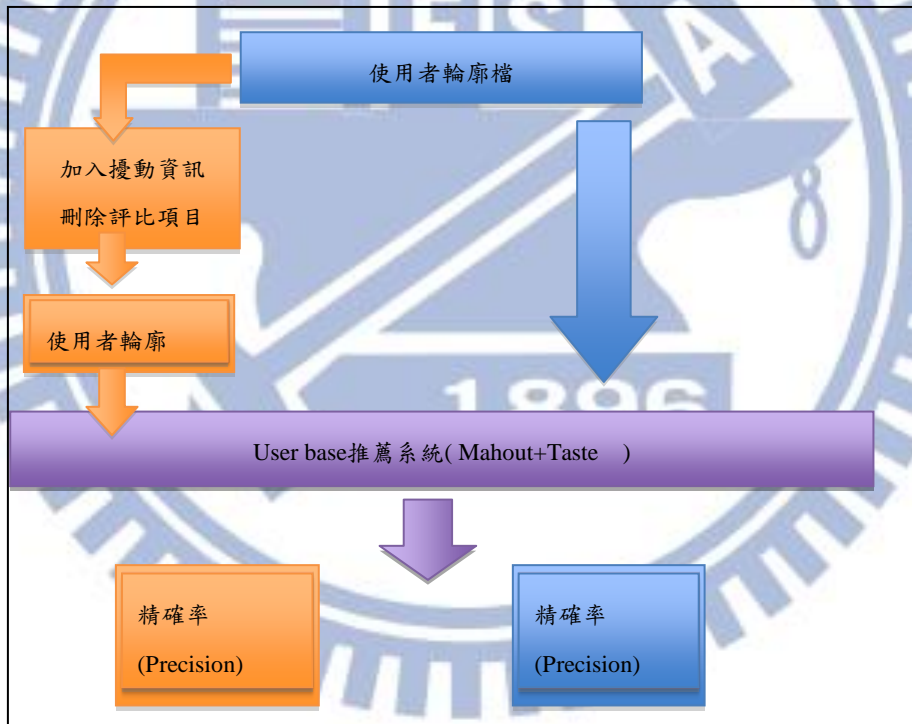


圖 32鼓勵評比實驗架構

### 4.6.1 激勵程度計算

利用 30 位使用者原始輪廓檔進行 50 次推薦並計算正確率。而刪除已評比過之項目部分，則分別計算刪除項目比例為 0.1(例如 user1 原評比 272 筆資料，隨機去除 27「 $272 * 10\% = 27.2$ 」筆資料後，評比數量降為 245 筆資料)、0.2、0.4、

0.6 與 0.8 後之推薦精確率。而使用者的預期收益則為整體推薦系統之精確率與使用者未參與評比情形下精確率之差。30 位使用者整體精確率與預期收益分別顯示圖 33、圖 34 與圖 35。

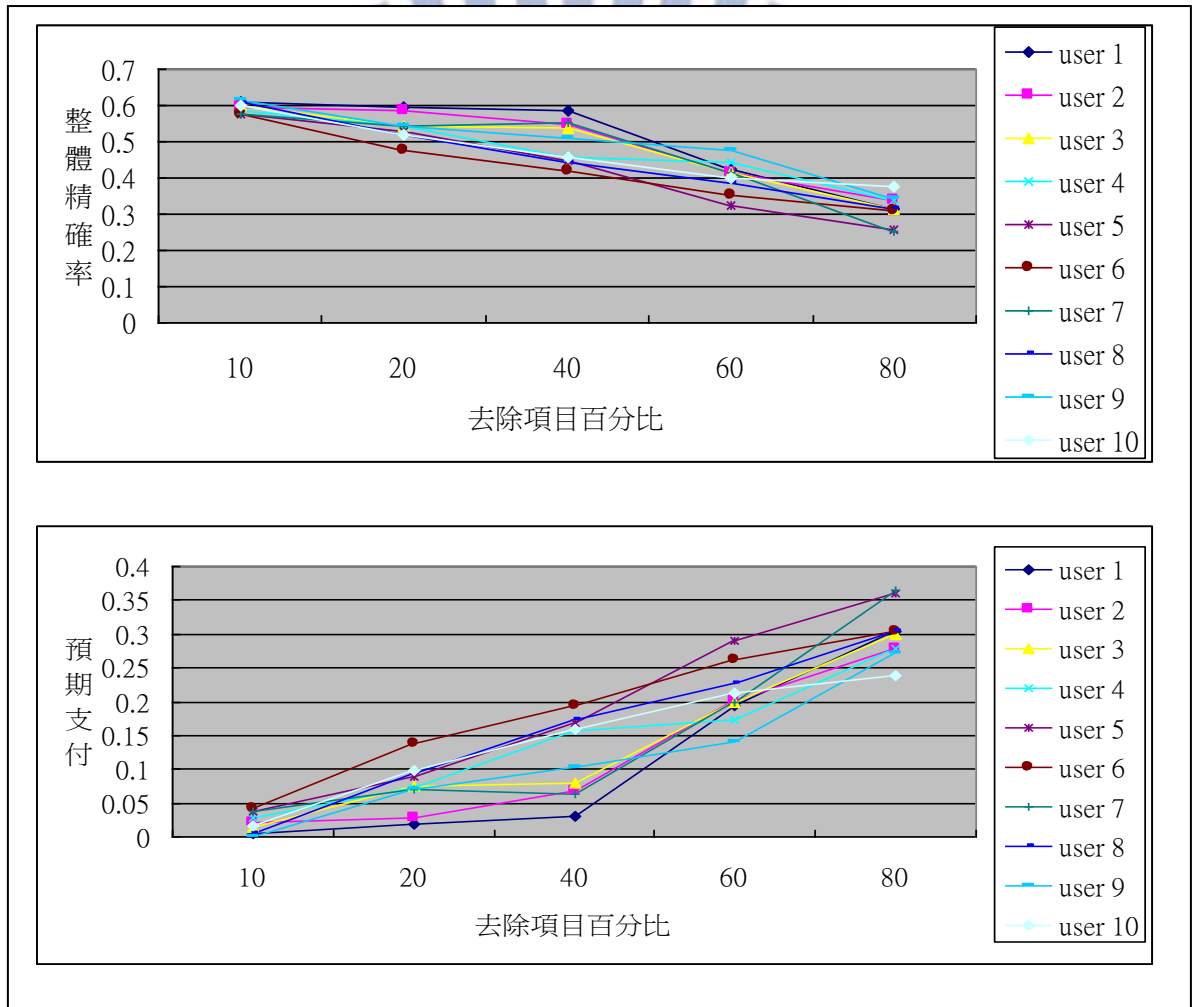


圖 33 使用者 1-10 預期效用

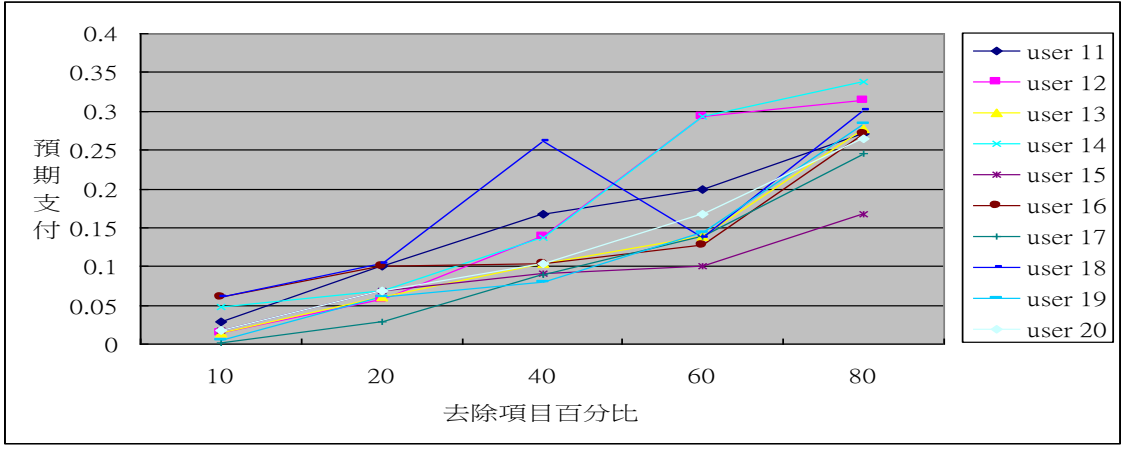
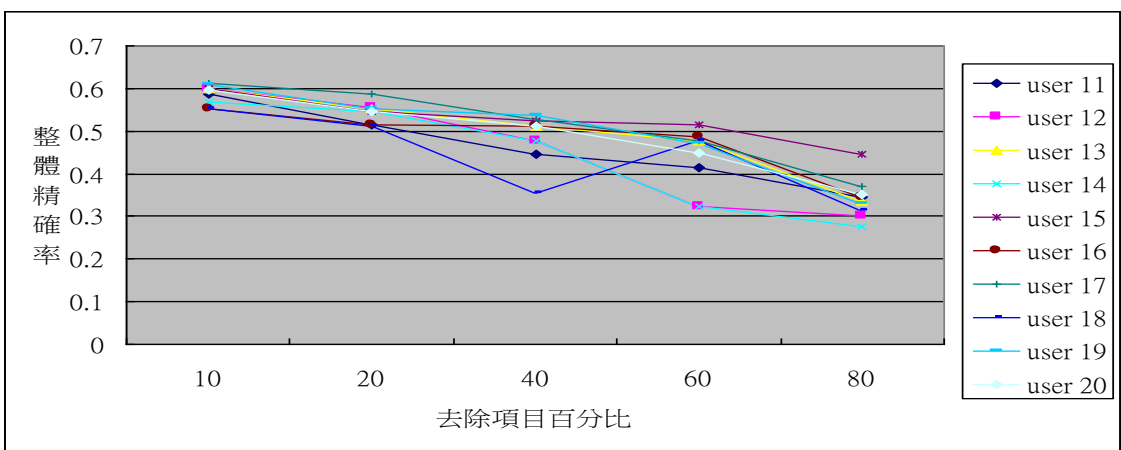


圖 34 使用者 11-20 預期效用

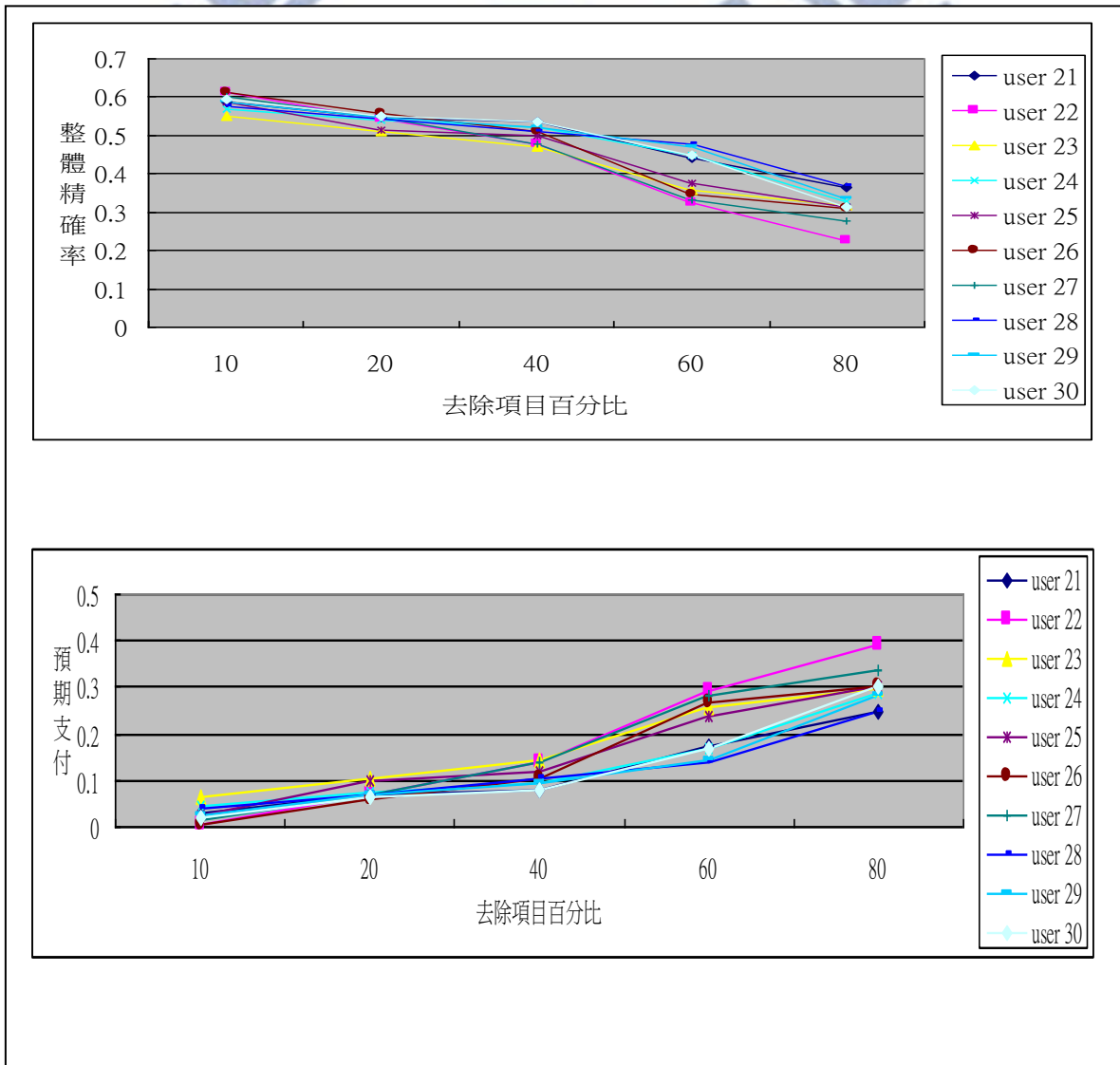


圖 35 使用者 21-30 預期效用



#### 4.6.2 鼓勵使用者評比驗證

在驗證鼓勵使用者評比方面，再給予激勵程度(預期收益)的條件下，觀察 15 位使用者對於 5 種(10%、20%、40%、60%、80%)刪除比例條件下回評的比例(例如，回評比例為 6%，user 1 評比 100 個項目，隨機刪除 10 項「 $100*10%=10$ 」後剩下 90 個評比項目。激勵後 user1 增加 6 項評比，使得評比項目由 90 提升為 96 個評比項目)。圖 36 為在刪除項目的比例為 10% 與 20% 的條件下給予相對應之激勵程度(預期收益)，並觀察 10% 與 20% 的條件下各 15 位使用者對於已刪除項目回評的比例。10% 的條件下其平均回評的比例為 8.74%，超過一半的項目被回評。而在 20% 的條件下其平均回評的比例為 17.64%。

另外觀察發現於 10%、20%、40% 條件下其回評比例皆超過一半，而在 60%、80% 條件下則小於一半，如圖 37，其原因經分析有以下二項：

- 刪除比例過高，使得原有高評比數量的使用者因刪除數量過多導致使用者回評意願降低。
- 刪除比例超過 50%，導致推薦正確率大幅降低，使得使用者回評意願降低。

經實驗與觀察結果發現於刪除項目的比例為 10%、20% 與 40% 的條件下給予相對應之激勵程度(預期收益)，其效益是最大的。由以上的比較分析可知，本論文所提之推薦方法確有明顯之效益。

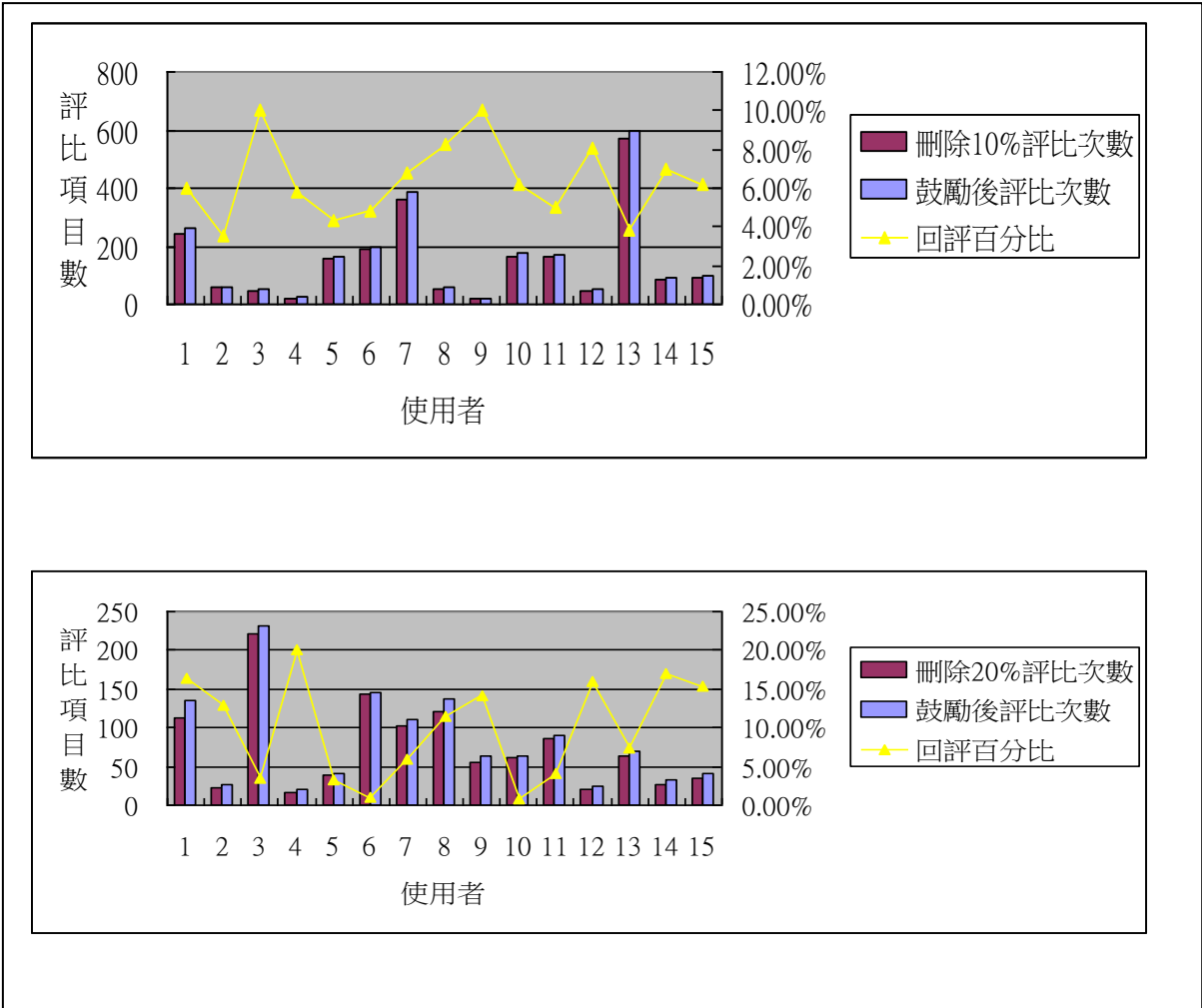


圖 36 使用者回評

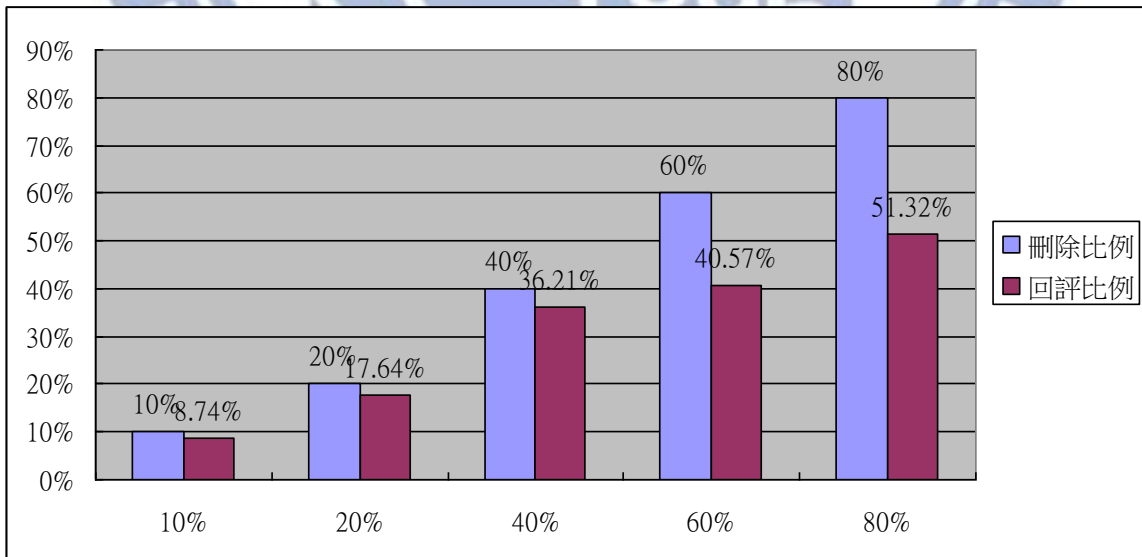


圖 37 使用者回評比例

## 第五章、結論與未來方向

### 5.1 研究結論

協同式推薦技術相關研究是目前推薦系統發展的主流。經由收集使用者的所有資訊(例如背景與相關興趣等等)來充實使用者輪廓(User Profile)的資訊。但是當使用者輪廓或適用的數據集資訊不完整時,推薦機制將面臨著冷啟始的問題,亦即系統中存在著「新使用者」與「新推薦項目」。

電影推薦系統中的冷啟始問題更是一個需要被關注的議題,當新進使用者進入系統時,由於系統一開始並不知道新進使用者的影片種類喜好,一般會先隨機挑選影片讓使用者試看,使用者在觀看系統所挑選的影片過後可以給予評分,讓系統得以知道使用者的影片喜好,再根據評分結果推薦影片給使用者,但是系統需要經過多次的詢答才能找到使用者的真正喜好,往往造成冷啟始的時間過長。當新的影片被存入系統時,由於新的影片沒有被使用者評分過,所以系統無法進行有效的推薦。而新進使用者在一開始進入系統時,由於系統並不知道新進使用者的喜好,所以無法正確推薦影片給使用者。

本論文所提出的拓展使用者輪廓檔與鼓勵評比經由實驗可將其結果歸納為以下四點:

- 在改善新進使用者輪廓檔資訊不足方面,本論文利用分類導向的方式找出使用者輪廓檔內項目間的類別,透過貝氏網路建構其類別關聯性來拓展使用者輪廓,並利用貝氏事後推論來避免關聯規則之推薦的誤差。
- 在改善新物件問題方面,提出一個以激勵模型為基礎的鼓勵評比方式,在滿足個人理性(individual rationality)條件與誘因相容(incentive compatible)條件下以經濟激勵來鼓勵使用者參與新物件評論。
- 在推薦正確率方面,協同式推薦系統透過較為完整的使用者資訊(以擴充之輪廓檔資訊),不僅提高了推薦正確率,亦可經由正確率的提升而達到留住既有的客戶以及尋找潛在的客戶的目標。
- 冷起始問題方面,避免因資訊不完整而造成冷啟始的時間過長,進而導致推薦效能降低。

## 5.2 管理意涵

本論文於考量在資訊不完整的前提下，探討協同式推薦技術與其他推薦方式所面臨的冷起始問題，經由擴大使用者輪廓檔，採用激勵理論鼓勵資訊分享，進而避免資訊不足的問題。本研究之貢獻主要有以下幾點：

- 利用激勵方式來鼓勵使用者參與評比

本論文中提出透過激勵方式來鼓勵使用者參與新物件評比，此方式可擴大其範圍進而討論推薦過程中的「需求與價值」。透握完整的激勵順序(包括：需求、價值、目標、績效、獎酬、及滿意)來分析如何因地制宜提出適切激勵方案，增加推薦效率與效能

- 資訊不足情況下分析資料關聯性

透過貝氏網路於資料缺乏的狀況下仍可分析因果關係之特性，在推薦系統只具有少量與不完整的資料時，依然可以產生推薦。

- 避免關聯法則所產生推薦遺漏的缺憾

透過信賴度 (Confidence)來衡量指標過濾有意義的規則，在產品之間有替代關係存在時容易不符合信賴度而造成推薦上的缺憾。經由貝氏網路來推論法則條件(Rule Body)與法則結論(Rule Head)間機率的狀況進而避免推薦的缺憾。

- 貝氏網路的互動與圖形表現特性，使得推薦的結果或推薦過程的溝通解釋更加容易。

- 使用者與推薦系統產生良性互動

透過正確率移轉價值方式結合使用者利益與推薦系統利益，此方式可增加使用者轉換成本進而產生鎖定效果 (Lock-in Effect)。

- 協助推薦及推薦策略的制定

本研究最終之目的就是協助推薦網站(系統)找出使用者所喜好的關聯產品。推薦網站(系統)可藉由本論文所提之方式來制定未來及推薦策略。例如在成本考量前提下，將使用者所喜好的關聯商品結合在一起推薦。

### 5.3 未來研究方向

在未來的研究方向，建議可朝以下幾個方面進行：

- 在產生關聯規則方面，本論文只單純的比較了 WEKA 所提供的 Apriori 演算法來產生規則，因此未來後續可以嘗試別種演算法，並比較在建構貝氏網路的上是否能獲得更好的成果。
- 於鼓勵使用者參與評比部分，未來可加入考慮使用者風險評估取向，包含代理人是風險中立(risk-neutral)或是風險趨避(risk-averse)。
- 雖然本論文實驗所使用之資料集為真實資料，但由於本論文並未於真實的推薦網站運作，因此在實驗中觀察使用者回評已刪除項目部分，亦受限於模擬長期的交易狀況。因此未來可將本論文所提之方法運用在真實情境下，以了解是否也能有如本實驗中之效益。



## 參考文獻-中文部份

- [C1] 鄭祐昇，「運用資料探勘技術於知識圖之建立」，東海大學，碩士論文，2003。
- [C2] 吳思怡，「運用貝氏網路與關聯規則於產品推薦之研究」，輔仁大學，碩士論文，2006。
- [C3] 張維迎、劉楚俊，賽局理論與訊息經濟學，頁數179-220。

## 參考文獻-英文部份

- [1] Weng, L.T., Xu, Y., Li, Y., Nayak, R., "Exploiting Item Taxonomy for Solving Coldstart Problem in Recommendation Making.", IEEE International Conference on Tools with Artificial Intelligence (ICTAI 2008), Dayton, Ohio, USA, November 2008, pp. 113-120 (2008).
- [2] Ziegler, C.N., Lausen, G., Schmidt-Thieme, L., "Taxonomy-driven Computation of Product Recommendations.", International Conference on Information and Knowledge Management (CIKM 2004), Washington, D.C., USA, November 2004, pp. 406-415 (2004).
- [3] Gavin Shaw, Yue Xu, and Shlomo., "Geva Using association rules to solve the cold-start problem in recommender systems", PAKDD'10 Proceedings of the 14th Pacific-Asia conference on Advances in Knowledge Discovery and Data Mining - Volume Part I Pages 340-347.
- [4] Lam, S. K., McNee, S. M., Konstan, J. A. and Riedl, J., "Getting to Know You: Learning New User Preferences in Recommender System. ", Proceedings of the International Conference on Intelligent User Interfaces, 2002, pp. 127-134.

- [5] Hsu, Mei-Hua., "A personalized English learning recommender system for ESL students. ", *Expert Systems with Applications* Volume. 34, Issue: 1, January, 2008, pp. 683-688.
- [6] Good, N., Schafer J. B, Konstan, J., Borchers A., Sarwar B., Herlocker J. and. Riedl J., "Combining Collaborative Filtering with Personal Agents for Better Recommendations. ", *Proceedings of the 1999 Conference of the American Association of Artificial Intelligence* , 1999, pp. 439-446.
- [7] David W. McDonald, "Ubiquitous Recommendation Systems," *Computer*, Vol. 36, No. 10, pp. 111-112, Oct. 2003.
- [8] Konstan, J. A., Miller, B. N., and Maltz, D., "GroupLens: Applying collaborative filtering to Usenet news," *Communications of the ACM*, vol.40, no.3, 1997, pp.77-87.
- [9] Ansari, A., Essegaier, S., and Kohli, R. (2000), "Internet Recommendation Systems," *Journal of Marketing Research*, Vol. 37, No. 3, pp. 363-375.
- [10] Chee, S. H. S., Han, J., and Wang, K. (2001), "RecTree: An Efficient Collaborative Filtering Method, *Proceedings of International Conference on Data Warehouse and Knowledge Discovery (DaWaK'01)*, pp141-151.
- [11] Zacharis Z. Nick, and Panayiotopoulos Themis, "Web search using a genetic algorithm," *Internet Computing IEEE*, Vol. 5 , No. 2 , pp. 18 – 26, Mar./Apr. 2001.
- [12] Breese, J. S., Heckerman, D. and Kadie, C. , *Empirical Analysis of Predictive Algorithms for Collaborative Filtering, Proceedings of the 14th Annual Conference on Uncertainty in Artificial Intelligence*, 1998, pp. 43-52.
- [13] Balabanovic, M. and Shoham, Y., "Fab: Content-Based, Collaborative Recommendation. ", *Communications of ACM*, Vol. 40, No. 3, 1997, pp. 66-72.
- [14] R. Agrawal, T. Imelinski, and A. Swami, "Mining Association Rules between Set of Itemset in Large Database," *Proc. ACM Special Interest Group on Management of Data (SIGMOD)*, May 1993.
- [15] Heckerman, D. and Wellman M., "Bayesian Networks.," *Communication of the ACM*, Vol.38, No.3, 1995, pp. 27-30.
- [16] Jensen, F.V., "An Introduction to Bayesian Networks.," Springer Verlag, 1996
- [17] B. M. Sarwar, G. Karypis, J. Konstan, and J. Riedl, "Analysis of recommender

- algorithms for e-commerce”, In Proceedings of the 2nd ACM E-Commerce Conference (EC’00), October 2000, Minneapolis.
- [18] N. Nisan. “Introduction to mechanism design (for computer scientists).”, *Algorithmic Game Theory*, pages 209–242, 2007.
- [19] Goldberg, D., Nichols, D., Oki, B. M., and Terry D., “Using Collaborative Filtering to Weave an Information Tapestry.” *Communication of the ACM*, December 1992, pp.61-70.
- [20] E. Rasmusen. “Games and information: An introduction to game theory.”, Blackwell Pub, 2007.
- [21] C. C. Aggrawal and P. S. Yu, “A New Approach to Online Generation of Association Rules,” *IEEE Transactions on Knowledge and Data Engineering*, vol. 13, no. 4, pp. 527-540, Jul./Aug., 2001.
- [22] M. S. Chen, J. Han. and P. S. Yu, “Data Mining: An Overview from a Database Perspective,” *IEEE Transactions on Knowledge and Data Engineering*, Vol. 8, no. 6, pp. 866-882, Dec., 1996.
- [23] R. Layfield, M. Kantarcioglu, and B. Thuraisingham. “Incentive and Trust Issues in Assured Information Sharing”. In *Collaborative Computing: Networking, Applications and Worksharing: 4th International Conference, CollaborateCom 2008, Orlando, FL, USA, November 13-16, 2008, Revised Selected Papers*, page 113. Springer, 2009.
- [24] R. Agrawal and E. Terzi. “On honesty in sovereign information sharing”. *Lecture Notes in Computer Science*, 3896:240, 2006.
- [25] Hofmann, T., “Latent Semantic Models for Collaborative Filtering”, *ACM Transactions on Information System*, Vol. 2, No. 1, 2004, pp89-115
- [26] Johathan, L., Herlocker, J., Konstan, J. and Riedl, J., “Explaining Collaborative Filtering Recommendations”, *ACM 2000 conference on Computer Supported Cooperative Work*, 2000, pp.241-252
- [27] Heckerman, David, “A Tutorial on Learning With Bayesian Network”, *Technical Report MSR-TR-95-06*, Microsoft Research, Microsoft Corporation, 1995, pp. 301-354.
- [28] Margaritis, D., “Learning Bayesian Network Model Structure from Data,” submitted in partial fulfillment requirements for the degree of Doctor of Philosophy, Carnegie Mellon University, Pittsburgh, 2003.



- [29] Singh, M. and Macro, V., “Construction of Bayesian Network Structures from Data: A Brief Survey and an Efficient Algorithm,” International Journal of Approximate Reasoning, Vol. 12, 111-131, 1995.
- [30] Apache Mahout and taste, <http://mahout.apache.org/>
- [31] Weka, <http://www.cs.waikato.ac.nz/ml/weka/>
- [32] Msbnx, <http://research.microsoft.com/en-us/um/redmond/groups/adapt/msbnx/>
- [33] GroupLens, <http://www.grouplens.org/>

