

以顧客價值分析與權重漸進探勘來進行協力式音樂推薦

Application of Incremental Mining and Customer's Value Analysis to Collaborative Music Recommendations

林朝興 (Chow-Sing Lin)

電腦工程博士 助教授

資訊管理所 南台科技大學

710 台南縣永康市南台街一號

Email: mikelin@mail.stut.edu.tw

06-2533131 ext. 4328

唐瑩荃 (Ying-Quan Tang)

資訊管理碩士

南台科技大學資訊管理所

台南市北園街80巷12號

Email: godinfender@yahoo.com.tw

(06) 275-6535

摘要

現今網際網路快速的發展，大量的數位音樂，已經廣泛在網路上傳播，造成使用者無法隨心所欲的找到想要的音樂歌曲。許多電子商務更是發展音樂推薦系統來提高顧客需求慾望。而一般線上音樂推薦系統，記錄了使用者所有歷史交易資料，並全部進行分析。因此，便增加執行時所耗費的成本、時間及是否符合使用者目前真正喜好的項目。本論文利用 RFM 模組來分析顧客價值，並且將相同顧客價值歸為同一群組，進而達到分群的動作。結合使用者最近習慣，提出以權重漸進探勘（Incremental Mining based on Weight）的構想，以漸進增加交易資料量的方式來探勘最近規則，而不需將全部交易資料都做分析，藉以節省計算成本、時間，並以 Apriori 演算法來探勘關聯式規則。而用相似向量矩陣計算使用者們之間的相似度關係，便利相似聚集。最後利用協力式推薦（Collaborative Filtering）的概念，由推薦模組將音樂推薦給使用者，做為個人化推薦方式。

實驗結果顯示，結合 RFM 模組及相似聚集推薦較單純只使用 RFM 分群方式為佳。此外，利用權重漸進與分群方式，更能夠推薦使用者喜好的音樂。而整體上，本論文的推薦準確率高達 0.78，比其他推薦方法高出 15%~32%，有效的達到個人化推薦的效果。

關鍵字：權重漸進探勘，音樂推薦系統，RFM 模組，協力式推薦，關聯式規則

Abstract

Because of the rapid development of internet network, the large amount of digital music has spread extensively on the Internet. That causes users cannot follow one's bent to find out the music or songs they want. Many e-commerce make further efforts to develop Music Recommendation System to improve customers' demands and desires. The general on-line Music Recommendation System records all user's former transaction and analysis them completely. So, it increases the cost, time, and items which adapt to what users like now or not. This paper combines RFM model to analysis customers' value and classify the same one as the same group. We combine users' Recent Behavior to Incremental Mining based on Weight, which can mine for relations by it, not analyzing all data, to decrease to calculate cost and time. It also prospect for Association Rule Mining by the Apriori algorithms. And then, similar vector matrix is used to calculate the degree of similarity relation between users' to assemble them conveniently. Finally, through the concept of Collaborative Filtering, we take advantage of recommendation model to be the method of individual recommendation that put music up to users.

According to the experiment results, it is better to use the combination of RFM model and similar assembling than RFM classification only. What is more, it also shows that using IMW and classification model can recommend fitfully music what users like. In the whole, the accuracy of this research is up to 0.78, which is higher 15% to 32% than the others, to make up the effect of individual recommendation.

Keywords: Incremental Mining based on Weight, Music recommendation system, RFM model, Collaborative Filtering, Association Rule Mining

一、研究背景與目的

亞馬遜網路書店 (Amazon.com) 的執行長 Jeff Bezos 曾經說過：『假如我在網路上有三百萬個顧客，我就必須有三百萬個網路上的商店』 [19]。這句話突顯了網路商店的重要性。所以在這資訊發達的世界中，如何擁有顧客的高忠誠度，是有其必要的。所以，分析顧客價值性是相當重要的，可讓網站供應商針對高價值顧客做精確的推薦或低價值顧客的行銷策略，以提高顧客的忠誠度。

隨著數位化風潮與高音質服務需求時代的來臨，眾多的數位音樂 (Mp3、Wav、MPEG) 等等，有如洪水傾巢而出，一般消費者要將所有的音樂內容消化幾乎是不可能的。加上琳琅滿目的音樂風格，一般的推薦系統恐怕無法滿足使用者的主觀意識，造成無法符合個人化推薦的主要因素。

然而，全球最暢銷的音樂推薦系統，如：Amazon、CDNOW 等網站，大部分都以產品為導向的銷售推薦而非以顧客為導向的精確策略。因此，若能有一自動化的輔助系統讓使用者得到許多相關的數位音樂歌曲，給予建議或是依據使用者的音樂風格偏好進行推薦，必能提高縮短找尋歌曲的時間及刺激網路音樂的市場，更可以讓使用者接觸自己喜好風格的音樂，增加其新鮮感或知識性。

資料探勘 (Data Mining) 可以用來找出龐大資料庫中隱藏的有用資訊的重要工具。在網站競爭激烈的環境中，許多新興網站紛紛採用不同的資料探勘工具來提供個人化的服務以吸引消費者的瀏覽。所以資料探勘技術的好壞與適切與否，將是影響個人化資訊服務與推薦的重要因素，也是本論文的相關主題。

近年來發展的過濾系統或推薦系統，幫助使用者將大量的資料，轉換為有用的資訊，提供使用者有興趣的產品，對使用者做精確有效的推薦。目前相關的推薦系統發展至今，已經產生出相當多的研究方法了，然而許多方法仍不甚理想。

以關聯規則來探勘使用者之喜好，其技術是將使用者所有的瀏覽記錄做分析並挖掘出結果為依據，因此系統需要過濾全部的交易資料，造成系統效率上的延遲，再則挖掘所有交易項目之結果也不一定是使用者真正目前感到興趣的[14] [22]。另外以階層分析法 (AHP)，採用自問自答的問卷方式來設定使用者最近喜好權重，其缺點是使用者必須自我決策，無任何即時動態的推薦反應，浪費了不少時間[1] [16]。而有關音樂推薦系統的研究，仍然以音樂內容特徵做為音樂推薦之研究最多[8] [13] [17]，然而這些研究需要加強擷取內容特徵的準確率，否則其推薦滿意度通常不佳。其他，如 Ringo 的音樂推薦系統[21]，必須先集合眾人對音樂歌曲的評價分數，再以協力式推薦方式，達到互相推薦的效果，不過這些傳送都必須透過 e-mail 來執行，造成無法線上即時音樂推薦。

後來也有研究學者以最後的交易筆數，當成使用者最近興趣的探勘結果並加入權重計算，可是這些研究方法只探勘有限之交易項目，一旦在大量交易資料下，需要挖掘多少筆交易項目及權重的大小的設定就會變的相當重要與複雜，因此無法更深入探討使用者最近興趣的探勘。然而，權重的主要精神是越接近現在的交易，權重就設的越高。所以加入權重計算可以更得知使用者在最近交易的類別中最感興趣之項目，也因此可以依權重值大小來給予適當的推薦數目。

基於以上的描述，本論文目的有三個構面：

(1). 提供一套個人化音樂推薦的機制

本論文以提供個人化服務的電子商務為目的，使顧客擁有自己量身訂作的個人化服務機制，以提高顧客對網站的高忠誠度。

(2). 取得顧客價值能力

擁有顧客價值能力，可讓企業或供應商，分析顧客的消費、忠誠度等能力，好讓企業針對不同的顧客能力，而有另外的行銷策略。

(3). 分析顧客的最近歷史交易，節省大量計算

本論文最主要就是透過分析最近歷史交易資料來推薦音樂，所以可協助企業瞭解顧客的消費行為與喜好，作為一對一行銷之參考。

本論文共分為五章，第二章為相關文獻探討，回顧之前研究方法與本論文研究相關之論文，第三章介紹如何進行 RFM 模組分群及權重漸進探勘使用者習慣之描述，並介紹相似向量分類及推薦之研究方法，第四章實驗結果與討論，第五章為結論及未來工作。

二、文獻探討

1. 資料探勘的定義

資料探勘的定義就是從大量資料庫中發現知識，將先前並不知道的或潛在有用的

4 以顧客價值分析與權重漸進探勘來進行協力式音樂推薦

資訊從資料庫中挖掘出來的過程[10]。資料探勘的第一步驟，是指定探勘和分析的資料，探勘的資料可以從資料庫或資料倉儲（Data Warehousing）中，取得表格、概略表或記錄文字檔。接著為降低資料量，首先將收集的資料作整理（Data Cleaning）、資料整合（Data Integration）或資料的轉換（Data Transformation）等等，並且確保使用者資料的完整。使用資料探勘的方法有分類、趨勢分析、分群、關聯及循序特徵等，從轉換後的資料中發掘存在的多重特徵及資訊。在探勘之後，我們可以從資料庫存取有趣的知識或規律，並可以從不同角度來觀察或瀏覽之。發現的知識可被用在決策、流程控制、資訊管理、查詢管理等等。因此，資料探勘被資訊業界認為是資料庫系統最重要的前途之一。

2. 電子商務與推薦系統

由於近年來網際網路（Internet）的快速發展，帶來了網路零距離的概念和便利性，網路逐漸成為另一個人們用來交換訊息和溝通的方式。而在網路技術成熟和基礎建設完成之後，網路愈來愈深入人們的生活，使得愈來愈多的人開始學習使用網路，並且逐漸將生活帶到網路上。同時許多商業行為也開始在網路上進行，各種電子網路商店林立，消費者只要輸入網址馬上就被帶到網路商店裡，相當方便而且迅速。

企業可藉由推薦系統來解決網路上的資訊超載，所以利用電子商務網站可以經由顧客歷史資料的提供、供應商的統計資料及探勘後的規則，來推薦個人化的產品。Schafer（1999）提出推薦系統對於電子商務上的效益包括[18]：

- (1). 提高顧客忠誠度：透過推薦機制，可以加強企業與消費者的關係，提高顧客忠誠度，降低流失率。
- (2). 將瀏覽者變成購買者：適時的推薦顧客所喜好的产品，可能會引起購買的慾望，增加產品購買的動作。
- (3). 交叉銷售：推薦系統可以推薦額外的產品給顧客來產生交叉銷售的利益，以提高平均的交易量。

3. 推薦技術的種類

目前推薦系統在電子商務上的應用，比較偏重於個人化的功能，希望能根據每個使用者不同的習慣分別給予使用者符合習慣的推薦，一般較常被應用的推薦方法主要為：

表一：內容式與協力式過濾方法之比較

推薦技術	內容式過濾	協力式過濾
做法	項目間的關聯性	使用者間的關聯性
優點	1. 可以推薦顧客獨特品味或興趣的產品。 2. 可針對提供的推薦解釋推薦的原因。	1. 結合了個人與參考群體的偏好。 2. 所推薦產品可能與顧客之前的興趣大不相同。
缺點	1. 從每一個物件抽取足夠的資訊來表示是不容易的。 2. 無法找到對顧客較特殊的產品。	1. 可能因某物件太少甚至沒有人對其評價，造成系統無法推薦。 2. 系統一開始必須有大量的產品評比，才能找出相似的社群。

- (1). 內容式導向過濾（Content-Based Filtering）：主要是根據資料的內容來篩選資料，為

內容資訊的延伸。並根據使用者輸入的關鍵字，由系統來篩選出符合這些關鍵字或者符合這些關鍵字的語意概念[4] [15] [22]。

(2). 協力式導向過濾 (Collaborative Filtering)：主要是利用眾人的意見來篩選資料，完全跟資料內容無關。此方法結合了個人與參考群體的偏好，能夠減少個人參與程度的同時亦能提供個人化資訊[5] [6] [7] [9]。

目前已有許多的研究針對此兩種方式做比較，見表一。而協力式過濾是目前使用最成功的推薦技術，它可以解決內容式導向的一些問題，幫助顧客得到潛在需要，甚至是有意思想不到的產品。然而，這些推薦方法並無絕對的好壞，只是依照需求不同而各有存在的價值。而本論文在推薦技術上是使用協力式推薦，因為音樂特徵內容過濾相當不易，可能造成準確率的低落，所以我們希望以協力式推薦技術，達到資訊互相分享的結果。

4. 顧客價值分析與 RFM 模型

傳統上，雖然供應商可以很清楚地知道每日的銷售額有多少，但是卻很難能將個別顧客與銷售情況作一連結。而透過資料庫，供應商可以很容易地對顧客進行價值分析，並針對不同價值的顧客進行不同的資源分配，以及採取不同的行銷策略。根據[11]所言：『所謂的顧客終身價值 (Customer Lifetime Value) 是指在未來一段時間之內，企業或供應商可以從個別顧客中獲得之利潤的淨現值』。而供應商可以依據資料庫中顧客的購買記錄，算出每位顧客可能貢獻於供應商的終身價值。透過顧客終身價值的計算，供應商除了可以預測未來的營收情況外，還可以確認出顧客價值的高低，以分配不同的資訊資源於不同價值的顧客身上。

目前評估顧客價值以 RFM 方法廣為運用[12]。供應商必須根據顧客過去的交易記錄，來對其未來的行銷做預測。供應商以三種構面分析顧客價值：最近購買日、購買頻率、購買金額。

首先，最近購買日 (Recency) 為衡量顧客最近一次購買日距離現今時點的日子。一般而言，最近購買日越小，即表示顧客最近一次購買完產品的時間距現今不久，所以被認為是較有可能再度購買的顧客，這群顧客的得分比最近購買日長的顧客要來得高。購買頻率 (Frequency) 指的是一個顧客在一段期間內購買產品的頻率。頻率愈高，則顧客對產品的忠誠度相對較高，此一顧客對公司而言也相對重要。因此，高購買頻率的顧客得分較高。購買金額 (Monetary) 指的是在一段期間內，顧客購買該供應商產品金額的總和。一般而言，金額總和愈高的顧客，其相對於供應商的重要性也愈高；反之，總和價值愈低的顧客得分就相對不高。透過以上三種分析指標，可以加總得到各個顧客對供應商的價值。

根據 Arthur Hughes (1994) 對 RFM 的定義：

(1). 最近購買日 (Recency)

最近購買日是指顧客最近一次購買距離分析時點的天數。一般而言，當顧客最後一次購買的時間距離在很近，即最近購買日較小時，供應商認為該名顧客會再度消費的機會較高，因而其在最近購買日指標分數較高。然而，在最近購買日分數的高低來決定顧客的重要性程度時，應考慮到該供應商產品的特性，不能單憑最近購買日的分數高低便決定顧客的重要性程度。

(2). 購買頻率 (Frequency)

購買頻率是指顧客在一定期間內購買該供應商產品的次數。一般而言，當顧客的購買次數愈多時，則供應商會認為其忠誠度和顧客價值較高。

(3). 購買金額 (Monetary Amount)

購買金額是指在一段期間內，顧客購買該供應商產品的總金額。一般而言，當顧客的購買金額愈高時，則供應商會認為其顧客價值較高。然而，為了避免新顧客因購買次數過少、購買金額過低，而忽略其重要性，因而在處理購買金額時，我們通常以平均購買金額來取代。

5. 關聯規則—Apriori 演算法

關聯規則探勘是資料探勘中的方法之一，經常運用於商品交易記錄資料庫上，針對使用者的交易行為做分析，找出商品間的關聯規則。再根據這些關聯規則來決定搭配促銷商品和商品架位等行銷策略，藉以提高交易量和營業額。例如：80%購買牛奶和果醬的顧客也會同時購買吐司，就是一個典型的關聯式規則，而本論文也基於這個購物規則來挖掘使用者的最近習慣。從 1993 年 Agrawal 提出從交易資料庫中探勘出關聯規則的演算法後[2] [3]，陸續有學者將關聯規則的概念應用到其他領域，提出適用於該領域的演算法。

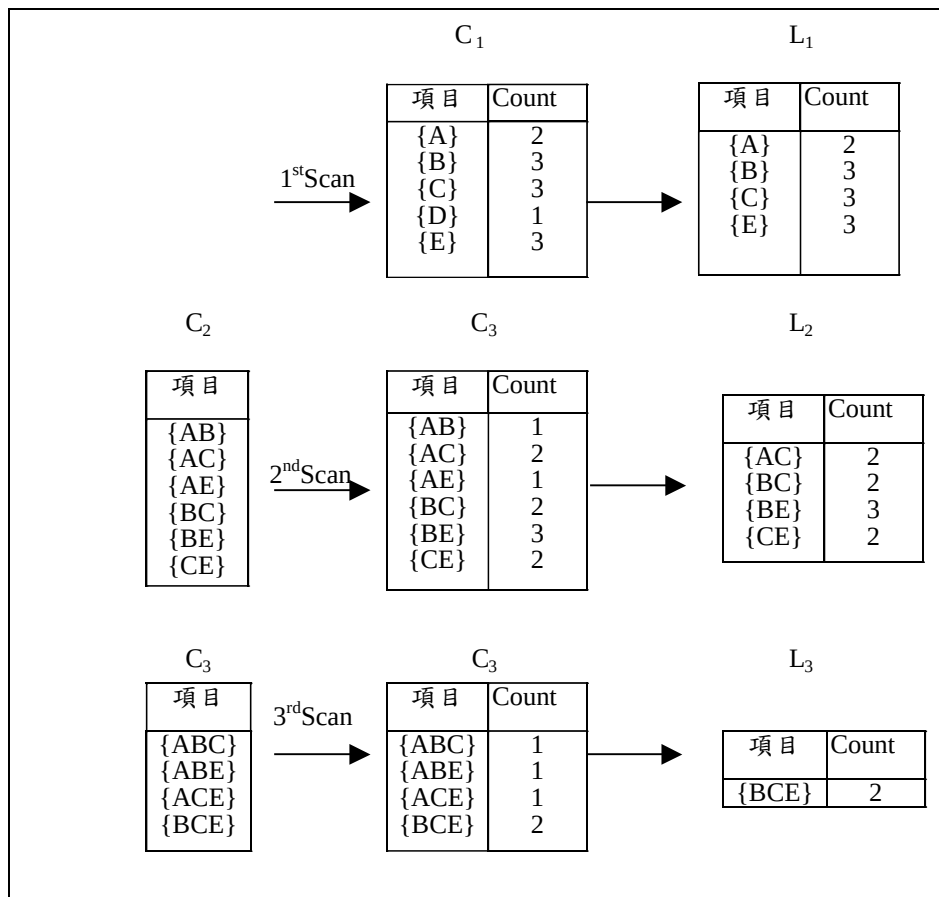
本論文所採用的高頻項目集合產生方式為 Apriori 演算法，其使用一種稱作逐層搜尋的疊帶方法，它會使用 k -項目集合來探索 $(k+1)$ 項目集合。首先，我們找出頻繁 1 -項目集合的集合，該集合記作 L_1 ， L_1 用來找出頻繁 2 -項目集合 L_2 ，而 L_2 用來找 L_3 ，如此下去，直到不能找到頻繁 k -項目集合，找每個 L_k 都需要對資料庫作一次掃描。為了要使用 Apriori 性質，頻繁項目集合的所有非空子集合都必須也是頻繁的。根據定義，如果項目集合 I 不滿足最小支持度 (Min-support) 值，則 I 不是頻繁的。

以下將利用簡單的例子，來看 Apriori 演算法的處理過程：假若資料庫中有四筆交易，每筆交易都具有不同的交易紀錄作代表，而交易中都包含了有數種項目，如表二所示：

表二：資料庫的交易紀錄

交易紀錄	類別項目
T1	ACD
T2	BCE
T3	ABCE
T4	BE

利用圖一，可以清楚的看到 Apriori 演算法的運作過程，先把所有的類別項目當成第一階段的候選項目集，而若是大於使用者自訂的支持度門檻值 (Support Threshold)，在此設定為 2，即可成為第一階段的高頻項目集合，所以刪除 D， L_2 中有 A、B、C、E，四個項目符合限制；進入第二階段循環，候選項目集的產生是由上一階段的项目集合做排列組合，一共產生了 $\{A \rightarrow B\}$ 、 $\{A \rightarrow C\}$ 、 $\{A \rightarrow E\}$ 、 $\{B \rightarrow C\}$ 、 $\{B \rightarrow E\}$ 、 $\{C \rightarrow E\}$ 六個候選項目集 C_2 ，計算其 count 值，並刪除 $\{A \rightarrow B\}$ 、 $\{A \rightarrow E\}$ 兩個項目集合；之後的流程便依此類推，經過不斷的排列組合及資料庫存取，將可在最後產生 $\{BCE\}$ 的高頻項目組合 L_3 。



圖一: Apriori 產生集合範例

6. 最近興趣與權重相關探討

H.-C. Chen and Arbee L. P. Chen (2001) 利用 pitch、duration、loudness，這些音樂標準的特徵抓取，經過分類然後再經由推薦模組找出一個最適合的方式將音樂建議給使用者。此篇論文強調的是三種的推薦模組（CB、COL、SAT），而這三種代表的意義分別為：（1）抓取使用者最近的興趣；（2）利用使用者的群組興趣相互推薦；（3）CB、COL 的混合型。研究論文中的 CB 方法強調最近興趣的抓取，並設定權重。可是當大量交易筆數下，如果只抓取最後幾筆資料來當作最近興趣，實在不足以代表最近興趣挖掘並且權重配置也無任何動態的表示，可能會造成管理者的負擔。而 COL 的應用為冷門的推薦，造成使用者滿意度的低落。另外在擷取音樂特徵中不容易達到成果，造成推薦滿意度的不足。其中 COL 的方法套用了[23]提出的使用者興趣與習慣，興趣的定義為 1 項集合類別，而滿足 2 項集合類別就定義為習慣。並且利用這兩種結果來完成六種分群的動作，由每一群集再進行相似度的比對及協力式推薦。此研究目的是為了找出使用者在最近交易的情形並且得知真正符合最近交易的興趣類別。因此提出演進式探勘公式，其主要精神及步驟都與一般的 Apriori 演算法相同。兩者之間主要的差異在於，掃描資料庫時觀察候選項目集是否存在資料庫中；如果有則將候選項目集的 count 值加一。比對整個資料庫之後，(count 值) / (總次數 - item 第一次出現的值+1) 所儲存的值即為此候選項目集的支持度。接著判斷其支持度值是否有超過支持度門檻值的限制，若是超過即成為真正高頻項目。

表三： 歷史交易資料

交易紀錄	類別項目
T1	AC
T2	AD
T3	ABC
T4	C
T5	ACD
T6	AD

表四： 一項集合演進式探勘過程

類別項目	First	Last	Count	Support
A	T1	T6	5	0.83
B	T3	T3	1	0.25
C	T1	T5	4	0.66
D	T2	T6	3	0.6

表五： 二項集合演進式探勘過程

類別項目	First	Last	Count	Support
AC	T1	T5	3	0.5
AD	T2	T6	3	0.6
CD	T5	T5	1	0.5

這樣的選取方式的用意，是越接近現在的記錄或交易越容易被選取出來，因為越接近現在的紀錄，分母就會越小；越久遠的紀錄，分母會越大。所以演進式探勘的精神即是能隨著時間演進來探勘出客戶的交易或瀏覽習慣。

表六： 時間權重交易

交易紀錄	類別項目	瀏覽時間	權重
T1	AC	03/16~03/22	0.1
T2	AD	03/16~03/22	0.1
T3	ABC	03/23~03/29	0.2
T4	C	03/30~04/05	0.3
T5	ACD	04/06~04/12	0.4
T6	AD	04/13~04/20	0.5

表七： 一項集合時間權重探勘

類別項目	Count	Support
A	5	1.05
B	1	0.19
C	4	0.83
D	3	0.66

表八： 二項集合時間權重探勘

類別項目	Count	Support
AC	3	0.62
AD	3	0.66
CD	1	0.23

我們以表三交易資料為例子，將它的每一類別項目計算他的 count 值及 support 值，接著刪除不足支持度門檻值（假設 45%）的項目 B。然後利用剩下的一項高頻項目組合成二項候選項目，一共產生了 $\{A \rightarrow C\}$ 、 $\{A \rightarrow D\}$ 、 $\{C \rightarrow D\}$ ，三個候選項目，之後依照上述方法計算其 count 值及 support 值。如表四所示：

我們從表五探勘出 $\{A \rightarrow C\}$ 、 $\{A \rightarrow D\}$ 、 $\{C \rightarrow D\}$ ，三個項目規則，我們可以發現，演進式探勘的最主要意義是可以根據交易順序來選取交易資料，越接近目前的交易就越有可能被選取出來。然而，這樣的探勘方式有一個缺點，就是其方法無法分辨 $\{C \rightarrow D\}$ 這一項目集，因為 $\{C \rightarrow D\}$ 只在最近交易中有交易過一次而已，但經由計算，那類別將會成為興趣類別，可是我們可從交易資料表內看出交易的比例， $\{C \rightarrow D\}$ 只佔有一次，所以無法認定它是否是真正使用者的最近興趣。

表九：最近興趣與權重相關比較

作者	研究方法	優點	缺點
Chen, H.-C. (2001)	- 以音樂內容特徵提供推薦。 - 使用三個推薦模組 (CB、COL、SAT)。 - 使用內容導向過濾方法。	- 最近興趣權重計算的推薦。 - SAT 推薦模組提供顧客不感興趣之物品。	- 音樂內容特徵不易擷取。 - 實驗結果，滿意度低。 - 最近興趣交易，代表性不夠。
Wu, Y.-H. (2001)	- 以一項集合代表興趣，二項集合代表習慣。 - 以演進式公式挖掘使用者最近興趣及習慣。 - 使用協力式導向過濾方法。	- 將使用者一開始所選交易分為六種分群。 - 推薦結果呈現滿意度較高。	- 演進式演算法，挖掘結果會產生過多的規則。 - 興趣與習慣的集合，不符合實際。
汪軒楷 (2002)	- 以時間權重及 AHP 方法改善 Apriori 的不足。 - 使用協力式導向過濾方法。	以 AHP 問卷分析，能夠得到較高的滿意度。	- 權重設定易造成管理者負擔。 - AHP 分析，造成時間浪費。
Weng, S.-S. (2004)	- 以產品特徵作個人化推薦。 - 兩階段分群法。 - 使用協力式及內容式導向過濾方法。	- 兩階段分群能夠更精確分類。 - 以產品特徵來分析使用者對產品的偏好，以利個人化推薦。	- 產品特徵的表達易造成管理者負擔。 - 大量資料下，特徵比對使系統效率降低。

汪軒楷、王台平的策略式資料探勘在個人化推薦上之研究 (2002)，提出時間式權重的設定中，利用瀏覽時間來區分權重的大小，越靠近最近交易時間其權重就越大，並利用 Apriori 演算法加上權重的計算挖掘最近習慣。我們同樣以表 3 的交易資料為例子，不過卻增加了使用者的瀏覽時間及權重設定，如表六，我們將一項集合項目的 support 值計算出來，以 A 為例：

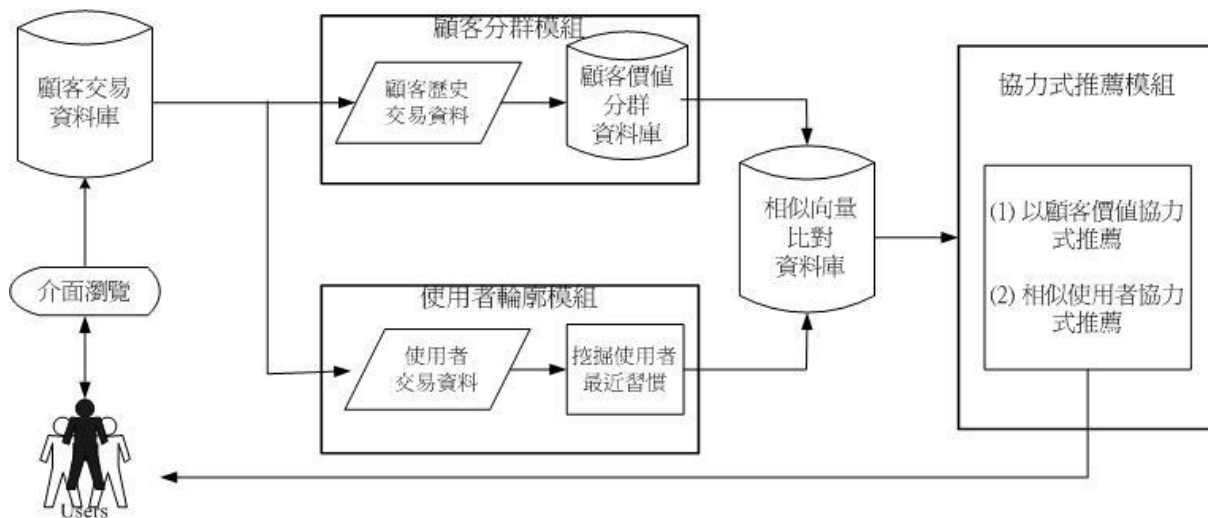
$$(0.1+0.1+0.2+0.4+0.5) / 6 + (5/6) = 1.05$$

見表七，我們計算出每一類別的 support 值，並刪除不足支持度門檻值（假設 45 %）的項目 B，然後組合二項候選項目並計算支持度，如表八並刪除不足支持度門檻值的項目 {C→D}。經由上述說明，其時間式權重的結果是在合理範圍內，可是研究中的缺點為管理者需要自己設定權重值，並且依時間的不同而有所改變，因此可造成管理者設定權重的負擔。除此之外，時間權重的設定很難有一個統一的標準。

綜合以上文獻探討，本論文的目的是希望提供一個人化音樂推薦方式，而且針對使用者最近習慣這方面的研究來提出不同於一般傳統探勘技術，因此利用權重漸進探勘將使用者之前的瀏覽記錄透過本方法來找出最近習慣的類別，見表九。

三、 研究方法

1. 研究流程



圖二： 線上音樂推薦流程圖

本論文將敘述線上音樂推薦系統的基本流程，如圖二所示，主要是為了提供使用者音樂的推薦服務。使用者們透過介面瀏覽的介紹，得知音樂的作者、風格、特徵樂器、歌名、年份、專輯名稱等等內容訊息，因此使用者可試聽音樂檔案並下載使用者喜歡的歌曲，而顧客交易資料庫便是記錄音樂編號、類別風格等等訊息。本論文透過三種模組階段來呈現音樂推薦的過程，(1) 顧客分析模組：透過歷史交易紀錄的行為將資料進行顧客價值分群動作，先達到資料整合及分配不同的資源於不同價值的顧客身上。(2) 使用者輪廓模組：將使用者交易資料進行權重漸進分析並取得使用者高頻交易項目的關聯規則，因此可探勘出使用者在最近習慣中的規則。(3) 協力式推薦模組：基於相似向量比對使用者的相似度後，將可以聚集相似規則使用者，最後進行協力式推薦，將相似比對結果做 top-N 推薦的階段。

2. 顧客價值分群模組之描述

2.1 RFM 模組分群定義

由顧客交易資料庫裡儲存著散亂無整合的歷史資料，所以一開始要先將顧客交易資料庫做一分群的動作並儲存於類別分群資料庫。而所謂資料分群（Data Clustering）的原則，就是『群組內的資料相似度最高，而群組與群組間的資料相似度最低』。因此，我們利用 RFM 模組的構想，將顧客的最近交易日期、交易頻率及交易金額，當成第一階段分群的動作。目的是為了讓往後之推薦能有效的對高價值顧客或低價值顧客做一適當的推薦。

首先 RFM 模組的定義：最近購買日（Recency），是指顧客最近一次購買距離分析時點的天數。購買頻率（Frequency），是指顧客在一定期間內購買該供應商產品的次數。購買金額（Monetary Amount），是指在一段期內，顧客購買該企業產品的總金額，不過本論文要將購買金額的計算方式改為計算類別項目（Itemsets）次數，而類別項目只要點選的越多也代表著顧客會花錢在這個類別項目上，每一個類別項目都代表金額一元。本論文透過價值分析功能可將顧客分群為八個群組，而每一種群組對電子網站供應商，

各擁有不同的推薦策略。

2.2. 顧客價值分群範例

假設使用者 U_1 的歷史交易資料，如表十，分別計算使用者 U_1 的 RFM 三種價值如下：

最近購買日期 (R)：假設目前日期為 2005/06/03，而使用者 U_1 最後一次的交易時間為 2005/05/14，相距天數是 $2005/06/03 - 2005/05/14 = 20$ (日)。

購買頻率 (F)：20 (次)。

購買金額 (M)：42 (元)。

計算使用者 U_1 的 RFM 價值分數，根據每一個使用者的 RFM 價值分數，我們以全部顧客 RFM 的總平均值為一標準，並且以 $\uparrow$ 表示其值大於總平均值，而 $\downarrow$ 小於總平均值。利用這種表示可以將群組分為八個 ($\uparrow\uparrow\uparrow$ 、 $\uparrow\uparrow\downarrow$ 、 $\uparrow\downarrow\uparrow$ 、 $\uparrow\downarrow\downarrow$ 、 $\downarrow\uparrow\uparrow$ 、 $\downarrow\uparrow\downarrow$ 、 $\downarrow\downarrow\uparrow$ 、 $\downarrow\downarrow\downarrow$)，並將每一位使用者分群在符合的群組內。假設表十一，表示著每一位使用者將他的 RFM 價值分數與平均值做一比較的結果，因此可以找出每一位使用者的群組型式。

每一個群組型式表示著行銷上的不同策略，例如使用者 U_3 ，其購買金額高、購買頻率高、最近購買日期近，都足以證明使用者 U_3 對網站供應商而言，是一個高消費、高忠誠的大客戶，因此與使用者 U_3 為同一群組的顧客，都是我們推薦效果最佳的顧客之一。而使用者 U_2 ，可說明為較不常進入網站的人，所以價值性較低，因此在推薦策略上當然是以多量、多類別的混合推薦，期待他下次再度光臨購物。

表十：使用者 U_1 歷史交易紀錄

交易時間	交易紀錄	類別項目	交易時間	交易紀錄	類別項目
2004/01/11	T1	AB	2004/08/30	T11	DE
2004/01/21	T2	CD	2004/09/11	T12	C
2004/02/12	T3	ABE	2004/10/14	T13	ABC
2004/02/15	T4	E	2004/11/15	T14	BD
2004/03/13	T5	CDE	2004/12/16	T15	AC
2004/03/19	T6	AB	2005/01/17	T16	AD
2004/04/21	T7	AE	2005/04/18	T17	ABC
2004/05/24	T8	AB	2005/05/10	T18	C
2004/06/21	T9	AB	2005/05/12	T19	ACD
2004/07/24	T10	CD	2005/05/14	T20	AD

表十一：以 RFM 價值分群表示

使用者	最近購買日期 (R)	購買頻率 (F)	購買金額 (M)	顧客的群組型式
U_1	20	20	42	$R \downarrow F \downarrow M \uparrow$
U_2	28	17	33	$R \uparrow F \downarrow M \downarrow$
U_3	22	30	44	$R \downarrow F \uparrow M \uparrow$
U_4	23	31	45	$R \downarrow F \uparrow M \uparrow$
U_5	19	19	44	$R \downarrow F \downarrow M \uparrow$
U_6	9	20	45	$R \downarrow F \downarrow M \uparrow$
U_7	35	30	47	$R \uparrow F \uparrow M \uparrow$
U_8	58	33	20	$R \uparrow F \uparrow M \downarrow$
總平均	26.75	25	40	

3. 使用者輪廓模組之描述

3.1. 權重漸進規則描述

傳統資料庫由於要計算全部的交易資料，所以要來獲得使用者的高頻繁項目，勢必造成系統執行的時間、成本的增加，影響了電子商務的即時性功能。況且，使用者最近交易的選擇中也不一定會一直圍繞在相同的類別風格，因此我們先說明最近交易的定義：使用者在過去瀏覽中，最接近目前的交易行為，雖然使用者可能會有最近一次與上一次交易相隔太久的情況。不過這些交易也是使用者的喜好內，並不會因時間而來改變。所以我們記錄所有使用者的歷史交易來做探討。

本論文提出 IMW (Incremental Mining based on Weight) 的構想，來找出使用者在最近時間內的交易習慣類別。目的是利用漸進探勘可以省下全部探勘的時間及動態的探勘出最近習慣。然而加入權重更是為了計算何項類別才是最接近現在的交易。如此一來，將可得到一個精確的興趣類別，而不是大約或增加其他類別項目了。然而漸進次數的不同，所得到的結果排列也會不一樣，而影響推薦的數目及精確度，所以支持度的設置就格外重要了。

為了計算每一類別交易項目是否達到支持度的門檻，本論文必須設定一個權重支持度 (Weight Support--WS)，並設定門檻支持度 Support 值，只要通過 Support 的類別，就將其視為高頻繁類別項目，並且刪除不足最小支持度 MinSupport 值的類別項目。

WS_i^j 為第 i 個類別項目在第 j 次漸進交易資料中的權重支持度， $Count_i^j$ 是第 i 個類別項目出現在第 j 次漸進交易的出現次數總和， j 為漸進探勘的次數。 W_j 是權重值大小的計算， β_j 為 β 的 j 次方，為一常數 ($\beta < 1$)。

$$\begin{aligned} WS_i^j &= WS_i^{j-1} + (Count_i^j \times W_j) \\ \text{where,} \\ WS_i^0 &= 0, \quad W_j = \beta^{j-1}, \quad j = 1, 2, \dots, n \end{aligned} \quad (1)$$

3.2. 權重漸進基於 Apriori 演算法探勘規則描述

本論文利用 Apriori 來探勘最近習慣規則，目的是找出有用的最近習慣規則 (Recent behavior rules—Rbr) 來進行精確的推薦。因此本方法是將上述權重漸進的構想加入 Apriori 內，希望透過這種簡單的規則思維，求得研究中理想的規則。以下是描述挖掘規則的步驟。

- (1). 假設先抓取交易資料內的最後 n 筆交易，並計算每一項集合類別 i 的次數值。
- (2). 以公式 (1) 計算 WS 值，比對每一項集合類別 i 是否符合 Support 的要求。
- (3). 如果不符合就再抓取下一個 n 筆交易資料，並且再重新計算 WS 值，直到其中有某個類別項目到達了支持度設定，就停止計算。
- (4). 刪除不足 MinSupport 之一項集合項目。並將通過一項集合類別之高頻繁項目，組合成二項候選集合，重複上述的方式計算每一個二項集合類別 i 。

- (5).二項候選集合依據之前的一項類別集合所做的漸進次數 ($j=1,2,\dots,n$)。接著刪除不足 MinSupport 的集合。
- (6).最後剩下的二項集合類別將視為使用者的最近習慣規則 (Rbr)。公式 (2)，將 WS 大於 MinSupport 及小於 Support 的二項類別項目歸為最近習慣規則。其中 $[i \rightarrow j]$ 為可能類別項目的表示。

$$\text{Rbr} = \{[i \rightarrow j] \mid \text{MinSupport} \leq \text{WS}([i \rightarrow j]) \leq \text{Support}\} \quad (2)$$

3.3. 權重漸進規則探勘範例

假設 $\text{Support}=4$ 、 $\text{MinSupport}=2$ 、漸進交易量 $n=2$ 、 $\beta=0.9$ 、而類別 i 分別為 A、B、C、D、E 為代表。

由表十假設有二十筆使用者 U_1 交易項目之歷史記錄，首先挖掘一項集合類別，假設一開始先抓取最後兩筆交易項目 ($n=2$)，即 T20、T19。計算每一類別項目的次數並經由公式 (1) 計算出權重支持度 WS 值並無任何一項類別項目符合 $\text{Support}=4$ ，如表十二。因此必須再加入另外兩筆交易項目 T18、T17，而且再一次進行權重漸進計算，見表十三依舊未符合 Support 值，所以就一直重複漸進資料的計算，直到完成。表十四為第三次權重漸進的結果，其中的 A 類別之權重支持度符合 Support 值，因此停止漸進計算，並且將 B、E 類別項目刪除，因未達到最小支持度 ($\text{MinSupport}=2$)。

表十二：第一次權重漸進計算

交易項目	類別項目	權重支持度 (WS)
T19 {ACD} T20 {AD}	A	$2*1=2$
	B	$0*1=0$
	C	$1*1=1$
	D	$2*1=2$
	E	$0*1=0$

表十三：第二次權重漸進計算

交易項目	類別項目	權重支持度 (WS)
T17 {ABC} T18 {C} T19 {ACD} T20 {AD}	A	$2+1*0.9=2.9$
	B	$0+1*0.9=1.9$
	C	$1+2*0.9=2.8$
	D	$2+0*0.9=2$
	E	$0+0*0.9=0$

表十四：第三次權重漸進計算

交易項目	類別項目	權重支持度 (WS)
T15 {AC} T16 {AD} T17 {ABC} T18 {C} T19 {ACD} T20 {AD}	A	$2.9+2*0.81=4.52$
	B	$1.9+0*0.81=1.9$
	C	$2.8+1*0.81=3.61$
	D	$2+1*0.81=2.81$
	E	$0+0*0.81=0$

表十五：二項集合第一次權重漸進計算

交易項目	類別項目	權重支持度 (WS)
T19 {ACD} T20 {AD}	AC	$1*1=1$
	AD	$2*1=2$
	CD	$1*1=1$

表十六：二項集合第二次權重漸進計算

交易項目	類別項目	權重支持度 (WS)
T17 {ABC}	AC	1+1*0.9=1.9
T18 {C}		
T19 {ACD}	AD	2+0*0.9=2
T20 {AD}		
	CD	1+0*0.9=1

表十七：二項集合第三次權重漸進計算

交易項目	類別項目	權重支持度 (WS)
T15 {AC}	AC	1.9+1*0.81=2.71
T16 {AD}		
T17 {ABC}	AD	2+1*0.81=2.81
T18 {C}		
T19 {ACD}	CD	1+0*0.8=1
T20 {AD}		

A、C、D 類別項目是一項集合的高頻繁項目，依照 Apriori 演算法組成二項候選集合， $\{A \rightarrow C\}$ 、 $\{A \rightarrow D\}$ 、 $\{C \rightarrow D\}$ 。表十五由一項集合類別的高頻繁項目組成二項候選集合類別，而二項候選集合類別的計算次數是依照一項集合類別總共做了幾次的漸進計算為停止依據 ($j=3$)，所以只要計算至第三次漸進就停止了，理由是要與一項集合的漸進一致，避免二項集合的漸進過程中，因無法到達支持度，而不斷的漸進計算，浪費不必要的成本。由表十六、表十七可看出計算的過程，最後二項集合類別的高頻繁項目規則為 $\{A \rightarrow C\}$ 、 $\{A \rightarrow D\}$ 兩項規則，而 $\{C \rightarrow D\}$ 這一項因支持度過低，因此刪除。最後將二項集合類別的高頻繁項目作為使用者最近習慣的依據，並進行相似使用者的聚集。

最後的結果規則 $\{A \rightarrow C\}$ 、 $\{A \rightarrow D\}$ ，而且 $\{A \rightarrow D\}$ 的權重支持度最大，因此往後的推薦過程， $\{A \rightarrow D\}$ 相關類別產品將推薦較多數目給使用者。我們也可以比較在文獻探討中的另外兩種方法，雖然最後結果相差不遠，可是本論文方法解決了演進式探勘的缺點及時間權重探勘之權重設定負擔，所以本論文可得出較精確的探勘結果。

4. 使用者相似類別聚集描述

4.1. 相似矩陣建置之探討

這一章節將說明如何再一次聚集相似使用者，並且依據使用者最近習慣得出的規則來做聚集分類。目的是為了聚集相似類別項目的使用者，找出使用者之間更相似的群組，達到真正協力式資訊分享的作用。定義如下：

$SM(X)$ 為使用者 X 最近習慣規則 (Recent behavior rules) 展開結合成相似矩陣 (Similar Matrix)，矩陣內第 i 列第 j 行的元素為 0，Rbr 為最近習慣規則 $[i \rightarrow j]$ 可能類別項目，公式 (3)。

$$SM(X) = [sm_{ij}]_{m \times n} ; (m, n = 1, 2, 3, \dots, p)$$

$$\text{where} \quad (3)$$

$$sm_{ij} = \begin{cases} 1, & \text{if } [i \rightarrow j] \in Rbr \\ 0, & \text{otherwise} \end{cases}$$

表十八假設有 A、B、C、D、E 五個類別項目，而使用者 X 最近習慣類別為 $\{A \rightarrow C\}$ 、

{A→D}。故在矩陣內[A,C]、[A,D]填入 1，其它仍為 0。

表十八：使用者 X 相似矩陣表示

	A	B	C	D	E
A		0	1	1	0
B			0	0	0
C				0	0
D					0
E					

4.2. 相似向量之計算

相似矩陣轉換成相似向量 (Similar vector)，相似向量的定義如下：

$$SV(X) = [sm_{12} \ sm_{13} \dots sm_{1n} \quad sm_{23} \ sm_{24} \ sm_{2n} \dots sm_{(m-1)n}] \quad (4)$$

經由公式 (4)，可從表十八轉換使用者 X 的相似向量為 [0110 000 00 0]，求得相似向量後，就可進行每一使用者之間的相似度比對。

4.3. 相似使用者聚集

上述方法是為了完成使用者的相似向量為目的，而現在我們要進行每一使用者之間的相似度比對，以便相似聚集 (Similayity) 分類的動作。

$$Similar_{xy} = (X_1 \oplus Y_1) + (X_2 \oplus Y_2) + \dots (X_l \oplus Y_l) \quad (5)$$

$$Similar_{set}(X) = \{ U_y \mid Similar_{xy} \leq \theta \} \quad (6)$$

公式 (5) 是計算使用者 X 與使用者 Y 之間相似值 ($Similar_{xy}$)，其中 X_l 及 Y_l 在這裡是指相似向量 X 或 Y 的第 l 個元素， $\oplus$ 為互斥或閘 (相同為零，不相同為一)。公式 (6) 設定分類門檻值 θ ，分類門檻值得大小將影響相似使用者的變動，造成推薦程度的不同。將相似值小於等於門檻值的使用者 $Y(U_y)$ 歸為使用者 X 的相似集合 ($Similar_{set}(X)$)。上述相似矩陣建置及聚集的方式，可以更清楚的將顧客們做相似的聚集，不同於之前的 RFM 分群。

4.4. 相似類別使用者聚集範例

假設類別項目 i 分別為 A、B、C、D、E 為代表，門檻值 $\theta = 2$ 。使用者 U_x 最近習慣的類別項目為 {A→C}、{A→D}。表十九將 U_x 最近習慣的類別項目表示為向量矩陣方式。接著將 U_x 量矩陣轉換成相似向量 [0110 000 00 0]，然後利用公式 (5)，以 U_x 為基準，與各個 Users 兩兩比對並計算相似值。見表二十，求出相似值後，透過分類門檻值 $\theta = 2$ 來區分相似同好，並聚集為相似同類。因此清楚看出 U_x 與 U_y 、 U_z 為相似同類，而 U_w 則視為不相似類。

表十九： 向量矩陣表

	A	B	C	D	E
A		0	1	1	0
B			0	0	0
C				0	0
D					0
E					

表二十： U_x 與所有使用者之相似值

使用者	相似向量	相似值
U_x	0110 000 00 0	0
U_y	0110 001 00 0	1
U_z	0110 001 10 1	2
U_w	0110 000 01 1	3

5. 協力式推薦模組描述

5.1. 針對高價值顧客之推薦

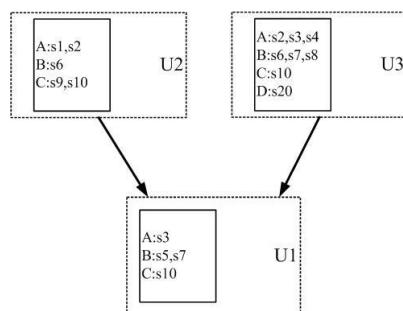
本論文在分群模組中，利用 RFM 分析顧客價值，並分成八個顧客群組，而這種方式，可提供網站針對價值度不同的顧客做策略性推薦。而本推薦方法將所有顧客選擇的產品做為推薦的產品，如果使用者是屬於高價值群組內，那推薦的產品就以探勘結果作協力式推薦；反之，使用者是屬於低價值群組內，那推薦的產品就依探勘結果及隨機推薦不相關產品。

例如：使用者 U_1 是屬於高價值群組內，其探勘結果為 $\{A \rightarrow C\}$ 、 $\{A \rightarrow D\}$ 兩種規則，我們將所有顧客選擇有關 A、C、D 的產品，推薦給使用者 U_1 ，達到精確個人化產品推薦。如果使用者 U_1 是屬於低價值群組內，我們不只推薦相關的產品，還要增加其它類別產品，因為我們必須對低價值群組的顧客，多做行銷廣告的策略，好讓這一群組的顧客能夠有機會光臨網站並提高消費。

5.2. 使用者相似類別同好之推薦

第二種方式為相似向量分類的結果，我們聚集了同一群組內與其他使用者最相近風格的 Users。因此本方法是利用 RFM 分類後在相同群的其他 Users 所點選的歌曲來進行相互推薦。這種方式的目的是經由第二次的分類聚集，可以得出更接近使用者習慣的分類，因此作法就是將其他使用者所選擇的歌曲，依據之前使用者的喜好類別不重複推薦給使用者，達到協力式資訊分享的結果。

圖三範例說明：假設 U_1 的其他相似同好 U_2 、 U_3 ，可以知道三位 Users 所選擇的歌曲編號（其中 s_t 代表第 t 首歌曲）。首先將 U_1 與 U_2 在相同類別中，不重複的推薦給 U_1 ，例如： U_2 將 A:s1,s2、B:s6、C:s9,s10 推薦給 U_1 。而 U_3 相同的類別也推薦給 U_1 ，例如： U_3 將 A:s2,s4、B:s6,s8、C:s10、D:s20 推薦給 U_1 。如此一來就會綜合 U_2 與 U_3 兩者的結果不重複的推薦給 U_1 ，即 A:s1,s2,s4、B:s6,s8、C:s9。

圖三： 使用者 U_1 與相似同好推薦

四、實驗設計與結果

1. 實驗環境與工具

本研究系統，使用 Windows 2000 Server 為系統建構之平台，而網站伺服器採用 Apache Tomcat 5.0，以可提供多名使用者同時上線且保持穩定為目的，主要負責 Client 端連上本系統時網頁的正常顯現與資料的正確連接；網頁設計以及資料庫存取方面，以 JSP (Java Server Page) 負責 Client 端網頁的顯示，而 Java Servlet 負責伺服器端資料的計算與採用 JDBC 協定連結後端資料庫並將結果呈現給 Client 端，資料庫則採用 Microsoft Access 2000。

2. 實驗系統架構說明

系統架構如圖四所示，共包含四個子系統，五個資料庫。

後端管理子系統主要負責新增會員註冊、登入與登出、管理者登入與查詢、產生音樂資料及紀錄交易情況控制。而這些功能，全部透過 JSP Engine，連接 Client 端與 Server 端之間的回應請求。權重探勘子系統主要是將交易紀錄，進行探勘使用者的最近喜好類別項目，並紀錄於探勘結果資料庫中。分群管理子系統，則是將顧客歷史交易資料庫依照 RFM 模組分析進行分群的運算。然後，利用探勘結果資料庫中的資料進行相似矩陣的計算，進一步的找出與使用者擁有相同喜好的人，並紀錄在分群資料庫中。

最後的推薦流程子系統包含一個推薦產品管理控制元件，將產品的推薦方法分成四種控制，分別為：

推薦 1. 使用權重漸進與 RFM 模組 (IMW+RFM)。

推薦 2. 使用權重漸進與 RFM 模組及相似聚集 (IMW+RFM+Similarity)。

推薦 3. 不使用權重漸進但加入 RFM 模組及相似聚集 (Non-IMW +RFM + Similarity)。

推薦 4. 使用權重漸進但是不加入任何分群技巧 (IMW+Non-Cluster)。

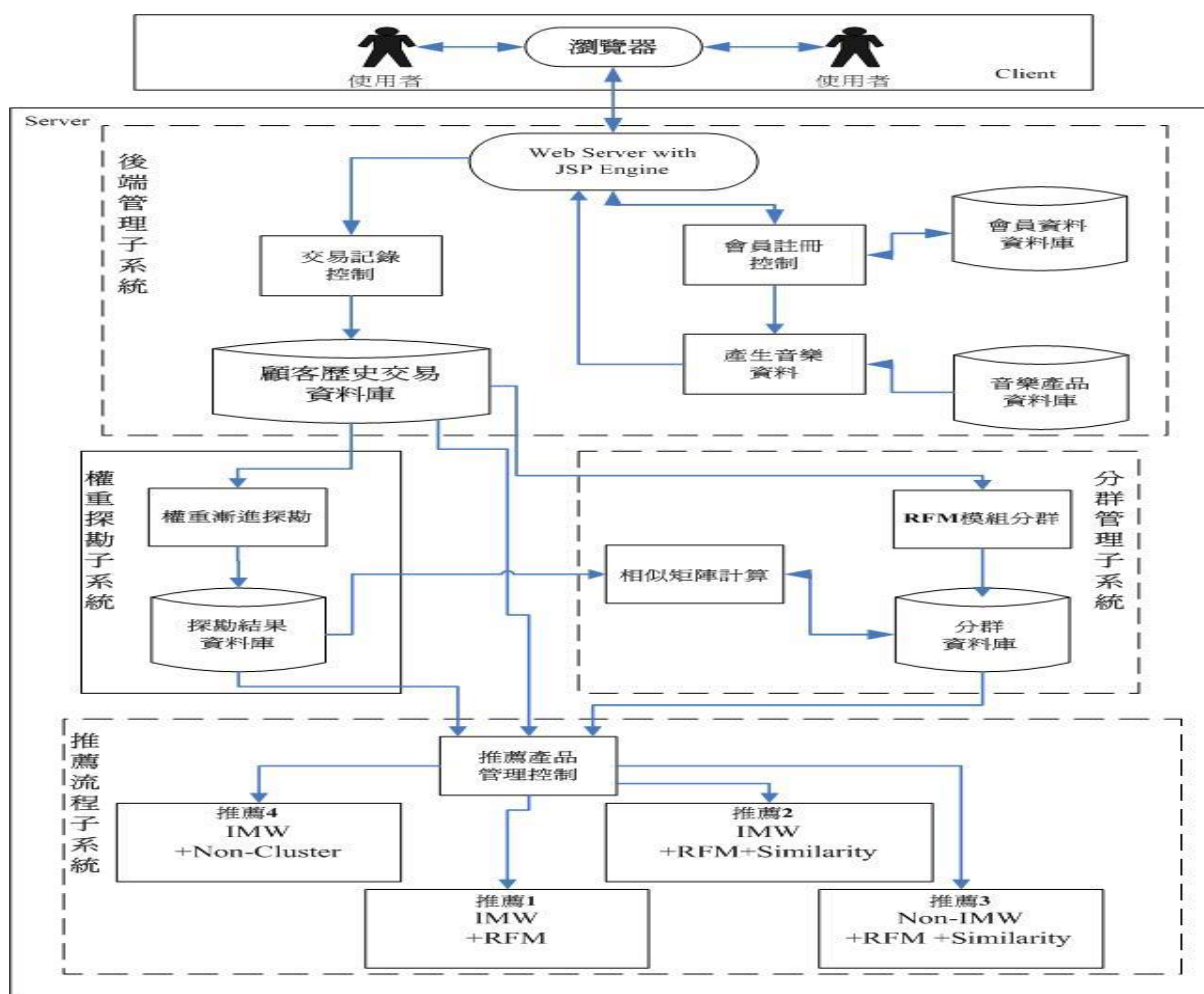
3. 實驗資料內容與執行畫面

本系統在實驗設計的部份，將以音樂風格為主，建購一個線上音樂推薦系統，來進行本論文之實驗及後續的實驗評估。

本系統在資料的整理上，主要是推薦之音樂資料內容，及顧客的所有歷史交易資料，分別說明如下：

音樂資料—本論文從網路 Amazon.com 網站或一些中文音樂網站，收集各類風格音樂，共有 165 首，其中國語歌曲類型數量較多，其次西洋流行，目的是讓本地受測者能夠接受推薦音樂的內容，所以本實驗資料有較多的國語、西洋流行音樂。見表二十一。系統在瀏覽器上會呈現每首音樂的歌唱者、歌名、價錢、專輯圖片，因此，使用者可以觀看這些資訊來點選有興趣的音樂，而這些特徵也是每位使用者會考慮的因素。

顧客資料—本實驗共徵得 178 位同學，年齡層分佈約在 20~30 歲之間進行，實驗受測者，並要求每一位受測者必須交易超過 2 次以上，共取得 1787 筆交易資料，這些交易紀錄將成為本實驗進行權重探勘與 RFM 分群或相似聚集之推薦依據。



圖四：音樂系統推薦架構圖

表二十一：音樂分類一覽表

編號	風格類別	數量
100~130	藍調音樂	30
200~230	爵士音樂	30
300~330	西洋流行音樂	35
400~435	國語流行音樂	40
500~530	搖滾金屬音樂	30

本論文所開發的音樂推薦系統，如圖五的執行畫面所示，左上畫面顯示，首頁將介紹整個推薦網站的流程，及一些重要的研究說明。由於本系統採用會員制，所以新會員必須先由『會員專區』進入。

右上畫面為，會員登入後，有提供五種音樂類別的歌曲，使用者可以依自己的喜好點選購物。而本畫面將介紹每一類風格音樂的價格、歌名、演唱者、專輯封面，可讓使用者做為參考購買的依據。這些音樂歌曲都由音樂產品資料庫提供，並展示於瀏覽器上。

左下畫面說明當使用者購買產品完畢後，就可結束交易，進行下一步驟，因此系統將透過顧客歷史交易資料庫進行權重探勘及 RFM 分群分析。如圖 4.6，本畫面上方先說明目前使用者是屬於哪一個 RFM 群組與代表意義，接著下方再說明與目前使用者同樣屬於同一群組的顧客名單。最後點選『請點選下一步驟（相似喜愛風格）』。

右下畫面，進行相似分類，目的是將 RFM 分群中的顧客，利用相似向量矩陣模組，再進行相似分群，分別說明與目前使用者擁有相同喜好的顧客及相同 RFM 分群但沒有相同喜好的顧客，因此在下方會出現四種推薦方法，分別為：『使用權重漸進與 RFM 模組』(IMW+RFM)、『使用權重漸進與 RFM 模組及相似聚集』(IMW+RFM+Similarity)、『不使用權重漸進但卻加入 RFM 模組及相似聚集』(Non-IMW+RFM+Similarity)、『使用權重漸進但是不加入任何分群動作』(IMW+Non-Cluster)，而每一個推薦都有其差別在，所以使用者必須一一進入，觀察推薦的結果。

最後本系統在推薦的音樂中建立了『是否購買』產品的選項，並收集這些資料，儲存於資料庫中，以便分析推薦的準確率。



圖五：網站執行畫面

4. 實驗評估

在完成整個系統建置後，最後必須進行系統的測試與評估，評估的目的在於瞭解使用者是否有依正常程序來操作系統，以作為修正系統或瞭解系統的貢獻度的參考。也同時試著瞭解使用者對於本系統的滿意度，瞭解推薦是否符合使用者的需求。本實驗評估利用兩種方式：

- (1). 在推薦系統中最常被用來衡量的指標包括準確率 (Precision) [20]。準確率指的是所找出的項目中，有多少是相關的。以下將說明之：

$$precision = \frac{\text{No. of purchased items} \cap \text{recommended items}}{\text{No. of recommend items}} \quad (7)$$

在準確率的計算上，主要是以系統推薦且顧客有購買之產品數量相對於所有推薦產品的數量，可以用來衡量顧客對所推薦產品的接受程度。也就是在我們所推薦的產品項目中有多少比例是顧客有興趣的且願意購買的，因此準確率越高，表示系統推薦的產品越被顧客接受。

(2). 問卷調查法：利用事先設計妥善的問卷，用於調查使用者或管理階層對系統使用的分析與困難問題。

本實驗目的主要是依據系統特性修訂而成，因此資訊推薦系統如果要填答過多的問卷時，會造成使用者的不耐煩或填選不正確的答案，所以我們將相關研究的問題濃縮，僅設計了 10 題的問項。問卷的內容主要在詢問使用者，『衡量權重漸進探勘音樂推薦系統的滿意度』，問卷採李克特五點尺度，從（一分，非常不滿意）到（五分，非常滿意），用此來了解使用者對本推薦系統所抱持的期望與態度。

5. 實驗結果與討論

本系統共收集了 181 位音樂推薦網站的瀏覽者，扣除填寫不完全的無效樣本，有效樣本為 178 位。以上每種推薦方法，都有『是否購買』協力推薦所產生的產品，目的是為了收集相關的準確率指標。準確率的計算方式為使用者從該推薦區點選的音樂個數除以推薦的音樂個數，見公式（7），例如，若推薦 10 首音樂，使用者願意購買 5 首音樂，則準確率為 0.5。而系統為實際線上實驗，實驗時間由 2005 年 6 月 24 日到 2005 年 8 月 26 日，並觀察這段期間使用者們之行為。為了考量 RFM 模組分群的效果，時間的計算採用每日批次的進行，讓 RFM 模組得到最佳的群集效果。由於每個人所登入操作的時間不同，因此在 RFM 模組分群中，我們將每一個群組的人數做了下列統計，如下表二十二。

表二十二: RFM 群組分佈人數

群組型式	使用者登入人數
群1. R ↓ F ↓ M ↓	45
群2. R ↓ F ↓ M ↑	30
群3. R ↓ F ↑ M ↓	26
群4. R ↓ F ↑ M ↑	24
群5. R ↑ F ↓ M ↓	15
群6. R ↑ F ↓ M ↑	10
群7. R ↑ F ↑ M ↓	12
群8. R ↑ F ↑ M ↑	16
總人數	178

群組 7 與群組 8，由於是剛開始系統實際測試，所以人數會比較少，而一般都集中在群組 1、群組 2、群組 3 與群組 4，因為測試的使用者大都是在最後的日期才進入本系

統操作。而各群族人數的不同是依照受測者的 RFM 之下所決定的，因此會出現某些群族出現較少的情形造成推薦人數的不同，所以當群族人數多時，推薦的產品也會出現多、雜的情況，所以，會影響推薦準確率的高低，而人數少時，也有可能造成推薦不符合需求的可能。接下來，我們將分析每一群組的使用者在每一個推薦過程的準確率與問卷討論。實驗比較總共分為四個推薦方法，而實驗將討論比較推薦 1 與推薦 2、推薦 3 與推薦 2、推薦 4 與推薦 2 的準確率，最後分析問卷調查的滿意度。以下我們分成兩個實驗結果的討論，分別為準確率討論與問卷討論。

5.1 準確率討論

表二十三是經由實際線上系統操作，將 178 位使用者之瀏覽行為，由推薦的準確率來看，推薦 1：單純使用 RFM 模組來分群，其平均準確率為 0.66；推薦 2：使用相似聚集做相似分群，其平均準確率為 0.78；推薦 3：不使用漸進探勘來推薦，其平均準確率為 0.53；推薦 4：不加入分群方式，其平均準確率為 0.57。可見，推薦 2 的平均準確率相較於推薦 1 高出 15%，較推薦 3 高出 32%，較推薦 4 高出 26%。

表二十三： 權重漸進+RFM+相似聚集與其他方法的平均準確率

	IMW	IMW	Non-IMW	IMW
	+RFM	+RFM+Similarity	+RFM+Similarity	+Non-Cluster
Precision 總平均	0.66	0.78	0.53	0.57

我們由推薦 2 與其他推薦方法的比較，可以說明單純只考慮 RFM 分群，在推薦的過程中可能會出現較多其他顧客的推薦，因此可能推薦一些非使用者喜好的產品，所以，本實驗設計證明兩階段分群效果優於單純的一項分群。

而推薦 3 不使用漸進探勘技巧的方法，在推薦上會不理會使用者的喜好，完全任意的推薦其他使用者所點選的音樂，造成推薦上會比較雜亂，所以使用者對這種混亂的推薦，可能感到不喜歡，因此推薦後的準確率明顯低於使用探勘的方式，證明權重漸進探勘的重要性。

另外，推薦 4 不使用分群技術，其平均準確率為 0.57，明顯低於推薦 2 使用分群技術的準確率，因此本研究的分群技術讓使用者更可以得到想要的，而不是一堆音樂產品，讓使用者無可適從。

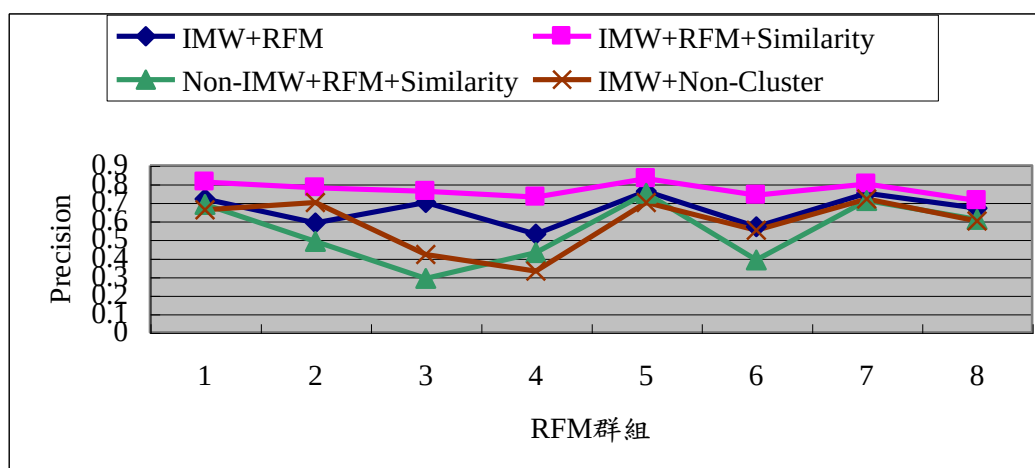
圖六，顯示 8 個 RFM 分群的準確率的分佈。我們可以看出群組 5 的推薦 2 準確率最高，達 0.83，而推薦 2 的群組 1 的結果也有 0.81，綜合這兩項結果，可能原因是這一群低價值的顧客們潛藏著未來高消費的購買力。

在群組 2、群組 3、群組 4 與群組 6 中，可以明顯的看出推薦 2 的準確率優於其他推薦方法，可能是高購買力與高忠誠度的顧客們，相對於其他推薦方法感到不適用，無法準確的推薦這些顧客們的要求，造成準確率的落差相當大。

最後由這些數據可說明推薦 2 利用權重漸進探勘與 RFM 模組及相似聚集的研究方法，能夠推薦使用者喜好的音樂，得到高準確率的成果，有效的達到個人化推薦的效果。

5.2. 問卷討論

經由受測者實際使用本系統後，對受測者進行系統功能滿意度問卷調查，如表二十四所示，Q1、Q2、Q3、Q4、Q5、Q6 等六項問題，主要針對受測者對系統整體的感受。就個別項目而言，Q2 分數最高，達 3.848 顯示使用者接受協力式推薦的方式。而 Q1、Q3、Q4、Q6，也高於總平均滿意度 3.477 也顯示了整個網站介面讓受測者可以輕鬆的瀏覽及測試，並且認同本論文利用最近習慣的探勘與推薦恰當的音樂資訊以符合個人化。Q5 的問項分數 3.184 滿意度，原因可能是使用者對音樂本身的認知較弱，所以對音樂推薦系統還是無法完全接受。



圖六： 權重漸進+RFM+相似聚集與其他方法的準確率比較

表二十四： 使用者對推薦音樂系統平均滿意度分析

問卷項目	平均數	變異數	標準差
Q1 本音樂推薦系統的流程，是否感到滿意	3.832	0.323	0.583
Q2 你是否認同協力式推薦的方式	3.848	0.386	0.612
Q3 我可以從本系統的協力推薦得知其他使用者與自己相同喜好的音樂	3.736	0.431	0.664
Q4 你是否認同由最近的交易紀錄為推薦依據	3.723	0.348	0.589
Q5 你是否喜歡以音樂為推薦的系統功能	3.184	0.371	0.608
Q6 這個系統給于我適當的資訊與推薦符合個人的產品	3.520	0.631	0.786
Q7 對於『IMW+RFM』是否滿意	3.384	0.343	0.581
Q8 對於『IMW+RFM+Similarity』是否滿意	3.781	0.590	0.767
Q9 對於『Non-IMW+RFM+Similarity』是否滿意	2.970	0.294	0.541
Q10 對於『IMW+Non-Cluster』是否滿意	2.802	0.24	0.704
平均滿意度	3.477		

此外，Q7、Q8、Q9、Q10 是針對四種推薦方法實用性的滿意度調查，滿意度最高的為 Q8，達 3.781，高於平均滿意度，顯示使用者相當滿意『IMW+RFM+Similarity』，這一項推薦的方法。相較之下，『IMW+RFM』這一項推薦方法，Q7 滿意度為 3.38，稍微低於平均滿意度，顯示使用者對單純使用 RFM 模組分群，所推薦的產品，滿意度還

是感到可以接受，原因可能是使用者點購音樂產品不多，造成推薦上無太大的變化。Q9、Q10 這兩個問項分數，明顯低於平均滿意度，因此可以看出使用者對不使用權重漸進探勘與無分群方式，造成無法恰當的推薦使用者喜好音樂，所以使用者對這兩項方法，都感到不滿意。

綜合上述實驗結果，由本論文提出的研究方法，可以證明其推薦上的效果，的確優於一般不使用權重漸進探勘與無任何分群技巧，而且對系統所提供的推薦資訊或協力式方法都是感到滿意的，所以我們提出的以權重漸進探勘使用者最近習慣之音樂推薦，達到預期的效果。

五、 結論與未來研究方向

1. 結論與研究貢獻

現今電子商務的快速發展，使得許多實體公司紛紛架起電子商務網站，做起網路生意。如果當使用者的人數及瀏覽記錄也日益增加，那按照傳統探勘技術，必定造成相當大的負擔。而此時，為了節省顧客搜尋的時間同時亦提高顧客忠誠度，推薦系統的應用不失為一種有效及個人化的手段。所以本論文為了減少探勘資料庫內的紀錄但又不失其準確度，因此以三個推薦模組流程作說明，分別是顧客分析模組：透過歷史交易紀錄的行為將資料進行顧客價值分群（Clustering）動作，先達到資料整合及往後對高價值顧客做適當的推薦。使用者輪廓模組：主要提出以權重漸進探勘（IMW）並結合 Apriori 演算法，將使用者交易資料進行權重漸進分析並取得使用者高頻交易項目的關聯規則，因此可探勘出使用者在最近習慣中的規則。協力式推薦模組：基於相似向量比對使用者的相似度後，將可再一次聚集相似使用者，最後進行協力式推薦方法。

研究結果顯示，結合 RFM 模組及相似聚集推薦較單純只使用 RFM 分群方式為佳。此外，利用權重漸進與分群方式，更能夠推薦使用者喜好的音樂。而整體上，本論文的推薦準確率高達 0.78，比其他推薦方法高出 15%~32%，有效的達到個人化推薦的效果。本論文的貢獻主要有三點：

(1). 兩階段分群效果

本論文為了達到有效的分群效果，採取了兩階段分群方法，並且從實驗結果也證實兩階段分群比單純使用 RFM 模組分群高出 14%，更讓受測者感到相當滿意的準確率，因此利用 RFM 模組與相似聚集，確實可以有效的將顧客群聚在一起，使推薦的項目更貼近使用者喜好。

(2). 系統權重動態的配置

由文獻探討中，可以得知演進式或時間式兩種探勘在分配權重上的缺點，造成管理者必須不斷的依照時間的變化給予權重值，因此系統負擔很大。而本論文加入權重的動態分配計算，是依照漸進的次數來遞減權重值，以符合使用者探勘習慣規則的權重配置。所以本論文為了解決其它研究上的不足，以固定動態的權重配置，得知相同的習慣規則。此外，權重的計算，更可以清楚的將喜好類別做有排序的分別，好讓推薦產品有差別性的運作。

(3). 以顧客為導向來進行個人化行銷

本論文最終的想法，是希望達到個人化的推薦並得到高度的滿意度，讓系統真正的應用在企業上，所以我們將 RFM 模組分析，應用於顧客分群上，提供銷售人員可以

即時的反應顧客的需求與良好的互動關係。因此，也可以藉由分析顧客消費行為之差異，來針對使用者個人進行不同的行銷策略，以提高獲利利益。

2. 未來研究方向

目前在推薦系統的相關研究上雖然很多，但大部分仍侷限於研究之領域上，而受限在研究資料的取得不易，往往很少有真正應用在實務產業上，可能原因是其成本高，而且執行效果也未達預期的滿意，因此在推薦系統的研究上仍有許多的改進與研究的地方。而未來的研究方向有以下幾點：

(1). 不限定產品類別推薦

目前，本論文還是以單純的類別來說明，並無更深入類別內容的應用，因此推薦的效果就會有差別。往後，希望能夠以類別階層來區別產品，讓產品多樣化。使推薦產品能夠更接近個人化選擇。

(2). 長期的資料收集

推薦系統在進行效果的評估時，往往都是以少數測試人員與模擬資料來進行，而本論文雖然以實際的線上實驗，不過由於時間的限制，無法收集更多的使用者資料，因此在相似推薦的過程及數目，都會有些許的變化，所以如果能夠進行一段較長時間的資料收集，相信能夠推薦讓使用者更多滿意的音樂。

(3). 朝使用者更新

一般的推薦系統研究大多是敘述推薦研究的過程，所以存在著是使用者歷史資料更新的問題。使用者不斷的進行點選動作，因此資料庫也不斷的記錄，當要分析推薦時，必須再重新的進行探勘，而造成效率的低落。如果能夠有一機制，以上一次的探勘結果與新的資料做一結合，必定能減少資料庫分析的負擔，達到更有效率及節省時間。

(4). 結合其它的推薦模型

每一種推薦方法都有其優缺點，本論文嘗試以相似分群及協力式推薦結果的方式，來加強推薦的滿意度，未來的研究希望結合更多其它推薦模型，整合其中的優缺點，以提升推薦系統整體的準確性。

六、 參考文獻

- [1] 汪軒楷、王台平，”策略式資料探勘在個人化推薦上之研究 ”，私立真理大學資訊管理學系碩士論文，2002。
- [2] Agrawal, R. Imielinski, T. and Swami, A. “Mining Association Rules between Sets of Items in Large Databases,” *Proceedings of the ACM SIGMOD Conference on Management of Data*, 1993, pp.207-216.
- [3] Agrawal, R. and Srikant, R. “Fast Algorithms for Mining Association Rules,” *Proceedings of the 20th International Conference on Very Large Database*, Sep. 1994, pp.487-499.
- [4] Basu, C. Haym, H. and Cohen, W.-W. “Recommendation as Classification: Using Social and Content-Based Information in Recommendation,” *Proceedings of National Conference on Artificial Intelligence*, 1998.

- [5] Balabanovic, M. and Shoham, Y. "Fab: Content-based Collaborative Filtering Recommendation," *Communications of the ACM*, Vol. 40, March 1997, pp. 66-72.
- [6] Cho, Y. H. and Kim, J. K. "Application of Web usage mining and product taxonomy to collaborative recommendations in e-commerce," *Expert Systems with Applications*, Vol: 26, 2004, pp. 233-246.
- [7] Cho Y. B., Cho, Y. H. and Kim S. H., "Mining changes in customer buying behavior for collaborative recommendations," *Expert Systems with Applications*, Vol: 28, 2005, pp. 359-369.
- [8] Chen, H.-C. and Chen, Arbee L. P. "Collaborative Filtering and Algorithms: A Music Recommendation System Based on Music Data Grouping and User Interests," *Proc. of ACM international Conference on Information and Knowledge Management*, 2001.
- [9] Goldberg, D. Nichols, D. Oki, B.M. and Terry, D. "Using collaborative filtering to weave an information tapestry," *Communications of the ACM*, Vol. 35, December 1992.
- [10] Han Jiawei, Kamber, Micheline ,*Data Mining: concepts and techniques*, Morgan Kaufmann, USA, 2001.
- [11] Hughes, Arthur , *Strategic Database Marketing*, Chicago : Probus Publishing, 1994.
- [12] Kaymak "Fuzzy target selection using RFM variables" *IFSA World Congress and 20th NAFIPS International Conference*, Vol.2, 2001, pp. 1038 –1043.
- [13] Kuo, F.-F. and Shan, M.-K. "A Personalized Music Filtering System Based on Melody Style Classification," *2002 IEEE International Conference on Multimedia and Expo ICME*, pp.9-12,Dec. 2002.
- [14] Kim, J. K., Cho, Y. H., Kim, W. J., Kim, J. R. and Suh, J. H. "A personalized recommendation procedure for Internet shopping support," *Electronic Commerce Research and Applications*, Vol: 1,2004, pp. 301-313.
- [15] Lang, K. "NewsWeeder: Learning to filter netnews," *Proceedings of the Twelfth International Conference on Machine Learning*, Tahoe City, pp. 331-339, 1995.
- [16] Liu, D.-R. and Shih, Y.-Y. "Integrating AHP and data mining for product recommendation based on customer lifetime value," *Information & Management*, Vol. 42, Issue: 3, March, 2005, pp. 387-400.
- [17] Shan, M.-K. Kuo, F.-F. and Chen,M.-F. "Music Style Mining and Classification by Melody," *2002 IEEE International Conference on Multimedia and Expo ICME*, Vol. 1, Aug. 2002, pp.26-29.
- [18] Schafer, J. B., Konstan, J. and Riedl, J. "Recommender Systems in E-Commerce," *ACM Conference on Electronic Commerce*, 1999, pp.158-166.
- [19] Schafer, J. B., Konstan, J. and Riedl, J. "E-Commerce Recommendation Applications," *Data Mining and Knowledge Discovery*, Vol. 5, 2001, pp. 11-32.
- [20] Saracevic, T., Kantor, P., Chamis, A.Y., and Trivision, D. "A Study in Information Seeking and Retrieving. II User, Questions and Effectiveness," *Journal of the American Society for Information Science*, Vol.39, 1998, pp.176-195.
- [21] Sharadanand, U. and Maes, P. "Social Information Filtering: Algorithms for Automating 'Word of mouth'," *Proceedings of CHI'95 Conference on Human Factors in Computing System*, 1995.

- [22] Weng, S.-S. and Liu, M.-J. "Feature-based recommendations for one-to-one marketing," *Expert Systems with Applications*, Vol: 26, 2004, pp. 493-508.
- [23] Wu, Y.-H. and Chen, Arbee L. P. "Enabling Personalized Recommendation on the Web Based on User Interests and Behaviors," *Eleventh International Workshop on Research Issues in Data Engineering*, April 2001.