

複合式關聯之內容導向過濾推薦系統

Using Relational Compound for Content-based Filtering Recommendation System

洪啟富
C.-F. Hung
國立臺北科技大學
電腦與通訊研究所
t6419003@ntut.edu.tw

段裘慶
C.-C. Tuan
國立臺北科技大學
電腦與通訊研究所
cctuan@ntut.edu.tw

曾冠惟
K.-W. Tseng
國立臺北科技大學
電腦與通訊研究所
seng.puppet@gmail.com

摘要

近年來，隨著無線網路與行動計算技術日益進步，所提供的查詢服務也趨近多樣性與多元化。行動通訊環境之行動用戶端(Mobile User, MU)向通訊基地台(Base Station)發出服務請求時，而產生重複之查詢結果，若能將 MU 間重複之查詢視為感興趣之行為，並有效推薦於其他使用者，則能減少發出相同之查詢需求。考量 MU 連續性位置移動與查詢服務之特性，本論文提出以行為鏈結(Event Link)為基礎之推薦評比系統，以內容過濾導向，進行行為價值推薦，依據熱門喜好幫助使用者篩選出重要訊息，提供更貼近個人需求之推薦系統。

關鍵詞：行動計算、資料預取、內容過濾推薦、事件樹

Abstract

Due to wireless networks and mobile computing technology advances, the services provide are close to diversity and pluralism. In mobile environments, most users will be repeated in some cases demand inquiry, condition servers excessive overload. In this paper, for a similar situation, associated overlap of mobile location and query requirements, possess of common behavior patterns for user to execute event clustering, and import content-based filtering event recommender, according to top preference assist users perk junk information and filter out important messages.

Keywords: Mobile computing、Data prefetch、Content filtering recommend、Event tree

1. 前言

行動用戶端透過行動裝置(Mobile Device, MD)，在無任何時間及空間的限制下，透過無線通訊傳輸技術，進行位置及查詢需求之資訊存取，例如：路況查詢、鄰近餐廳資訊與停車位等，遠端伺服器依 MD 所在座標位置回覆所需之資訊內容。亦即該資訊只有在某一特定區域內有效，且查詢結果與 MU 現況位置有密切關聯[1]。這類之服務雖然對生活帶來便利，仍有許多由無線網路所產生的問題須克服，例如：資料延遲傳送及行動用戶端重複發出查詢請求，所造成遠端伺服器過度負荷及網路頻寬的浪費，為讓資訊存取更具效率，過去相關研究，如運用行動代理人(Mobile Agent)、預取機制(Prefetching)[2][3]等，但其缺乏考量連續性查詢之使用者行為，造成遠端伺服器頻繁回覆 MU 之位置相依資料問題。本研究以分析行為事件鏈結為基礎，佐以探勘事件間隱含關聯價值，藉事件鏈結評比推薦(Event Link Rating Recommended, ELRR)掌握 MU 更多相關特徵(Profile)，以位置與查詢需求預估使用者之後續行為模式，將個人經常查詢類別納入大眾評比條件，亦更準確預估個人喜好項目，推薦更貼近於 MU 在現實生活中之預取行為。

2. 文獻探討

2.1 關聯式規則

透過歸納方式分析 MU 之行為模式，探勘事件間隱含價值資訊，以預測其未來可能的觸發行為之方法，最常見係以資料探勘[4]之關聯式規則(Mining Association Rule, MAR)[5]作分析預測方式，以將大量潛在雜亂無章之資料，歸納成少量易於辨識的可用數據，例如：A 資

料與 B 資料彼此具相對關聯，當 A 資料產生，通常會有較高可能性產生 B 資料。在[5]中提及購物籃分析(Market Basket Analysis)之應用，能在使用者交易記錄中，找出產品項目間的關聯性，以分析出顧客之購買模式與行為。MAR 研究範疇中多根據 Apriori 演算法[6]延伸，由下往上(Bottom-up)循序漸進找尋最佳項目集，制定支持度成為高頻項目集之門檻值，探勘出符合支持度設定的一階高頻項目集(Frequent 1-itemset)，經反複測試而產生候選項目集，直到高頻項目集符合信賴度的門檻值，則該項目集就成為關聯法則。

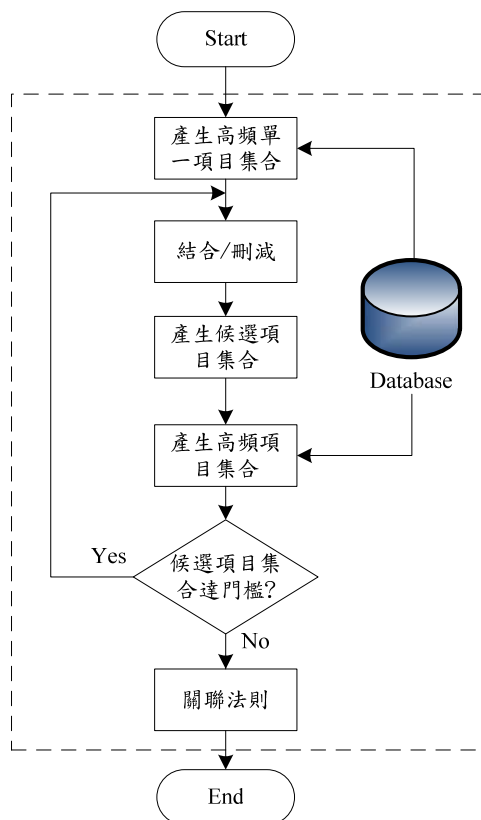


圖 1 Apriori 演算法流程圖

2.2 使用者行為探勘

現實生活環境中，每個使用者為完成一個特定目標，所做的一連串需求與位置變化稱之行為事件，例如：林先生從家裡到巷口買早點，買早點代表需求，家到巷口則代表個人之位置變化。個人行為以動態目標作導向，且易受到外在環境及生活習慣所影響，造成行為過程中產生頻繁變化，無固定規則或模式可依循。行為鏈結以不同型態之關聯行為組成，因此，於動態行動環境計算中具高度彈性，一般研究行為探勘多著重於移動方向或需求條件

擇一進行探討[7]，並無考慮同時存在移動及需求之情況。研究[8]針對 MU 移動的位置及查詢服務具有相依關聯時，能夠用二維資料探勘方式，擷取相互間具相依關聯之行為模式。如圖 2 所示，其中 User A、User B 及 User C 分別進行不同之行為事件，實際應用以 User A 為例子，User A 由通訊細胞(Cell) A 起始點出發，於 Cell A 底下作查詢超商之服務，接著至 Cell E 醫院查詢需求，移動到 Cell F 的加油站，然後到 Cell F 的銀行領錢，最後停留在 Cell I 的辦公大樓，構成具連續性行為之鏈結。而 UserB 及 UserC 同樣也擁有一套屬於個人的 EL，發現不同使用者行為事件中，有部份的位置移動或服務作重複之動作請求，本系統亦藉著 MU 所產生的行為鏈結以 top-N 方式，推薦評比較高之熱門鏈結，如此能夠更精確掌握 MU 之位置及需求服務，利用這些行為則預估未來使用者可能之行為情境。

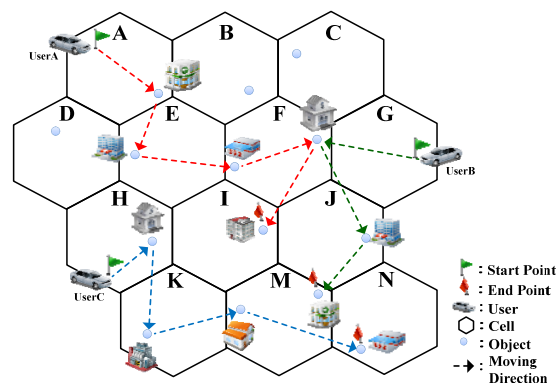


圖 2 行為事件示意圖

2.3 內容導向過濾

為了幫助使用者在面對大量未知資料時，能依據機制的推薦及過濾，預測商品的喜好程度，篩選出使用者可能感興趣之資訊，進而將喜好程度較高的商品推薦給使用者做選擇，利用資料特性分析技術，其中個人評價與演算效率在推薦系統扮演極為重要的角色。推薦系統(Recommender System)[9]應用層面相當廣泛，例如：大型網路書店亞馬遜(Amazon.com)，讓使用者能對個別商品(Item)給予評價，也普及應用於旅遊景點、餐廳、旅館及書籍等。通常推薦方式可分為以使用者為基礎(User-based)、以內容為基礎(Item-based)及關係過濾(Social Filtering)之協同過濾推薦，在眾多的推薦技術中，大多透過個人或其喜好近似之群體，以 top-N 方式推薦給使用者感興趣項目，同時該使用者也藉由評比機制給予一

定程度之評價，系統紀錄其回饋之動作，將回應評比作為其他使用者喜好相似度之參考值。本論文選擇以內容過濾推薦方式，其喜好知識為使用者對商品的歷史喜好程度及物品間所挖掘出的關聯度，能夠準確預估使用者諸多未知評比，推薦給予使用者較高評比之物品，解決推薦系統所衍伸之商品稀疏性問題。

	I_1	I_2	I_3	I_4	...	I_{m-1}	I_m	Item
U_1	+	-	+	+		+	-	
U_2	-	+	-	-		?	+	
U_3	+	-	-	-		+	+	
U_4	+	?	+	+		+	+	
$\vdots$								
U_{n-1}	+	-	-	+		+	?	
U_n	-	-	+	?		-	-	
User								

+ : Interest Rates - : Non-interest Rates ? : Unknown Rates

圖 3 Item-based CF

如圖 3 所示，使用者對商品之滿意度評比陣列(User-item Matrix)，能比較商品間之相似程度，利用距離公式計算相關性以產生基礎模型。於內容導向過濾推薦應用上，相似度演算法扮演極為重要的角色，以往研究除考量其使用者位置與商品位置相對距離外，亦同時藉由相關係數計算商品間相似程度，進而預測該使用者對未給予評價商品的感興趣程度。相關公式如(1)，並以 $\bar{r}_u = \{r_1, r_2, r_3, \dots, r_u\}$ 表示使用者 u 之平均評價紀錄， U 為 $Item_i$ 與比較 $Item_j$ 之共同評比之使用者集合；即 $U \in I_i \cap I_j$ 。

$$Sim(i, j) = \frac{\sum_{u \in U} (r_{u,i} - \bar{r}_u)(r_{u,j} - \bar{r}_u)}{\sqrt{\sum_{u \in U} (r_{u,i} - \bar{r}_u)^2} \sqrt{\sum_{u \in U} (r_{u,j} - \bar{r}_u)^2}} \quad (1)$$

其中， $Sim(i, j)$ ： $Item_i$ 與 $Item_j$ 之相似度。

$R_{u,i}$ ：使用者 u 對 $Item_i$ 之評比。

$R_{u,j}$ ：使用者 u 對 $Item_j$ 之評比。

$\bar{r}_u$ ：使用者 u 之平均評比。

相似度取得後計算預測商品的滿意度，預估使用者對商品評比，過濾關聯性較低的商品相關公式如(2)。

$$P_{u,i} = \frac{\sum_{x \in I} (S_{i,x} \times R_{u,x})}{\sum_{x \in I} (|S_{i,x}|)} \quad (2)$$

其中， $P_{u,i}$ ：使用者 u 對 $Item_i$ 之預測評比。

$S_{i,x}$ ： $Item_i$ 與 $Item_x$ 間之相似度。

3.事件樹與推薦評比預估

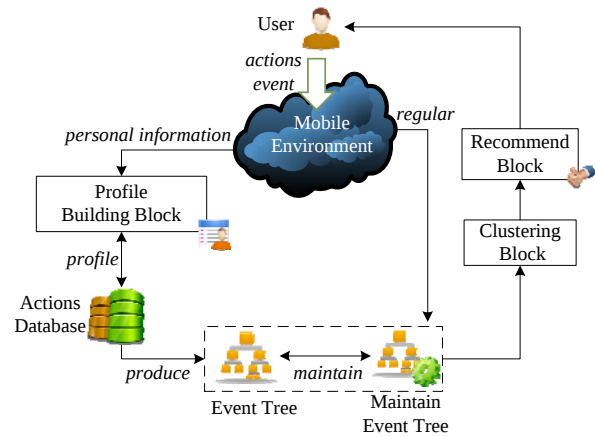


圖 4 基礎系統架構

本論文主要由兩方面進行探討，第一從眾多複雜且具多重關聯規則的行為中，根據其關聯與行為喜好，建置索引式事件樹。第二是藉由分析使用者之相關特性，根據熱門行為事件與使用者間相似度，推薦具高相似度的行為事件。基礎架構如圖 4 所示，由使用者行為事件建置索引事件樹，當事件樹隨著時間及節點增加日益龐大，將影響到資訊品質與系統負荷，所以事件樹需經動態維護，確保事件節點間的行為強度及關聯度，透過內容過濾評估出的評價與實際的評價間平均之絕對誤差。ELRR 架構問題描述，如圖 5 所示，區分四個區塊，下列分別對區塊作定義及描述：

- (1) 使用者(User)：包含一般進行消費顧客或瀏覽者，令 $U = \{u_1, u_2, \dots, u_{n-1}, u_n\}$ ，代表所有使用者之集合。
- (2) 商品(Item)：包含各種產品或服務，本架構以 MU 所產生之行為做評比對象，令 $I = \{i_1, i_2, \dots, i_{n-1}, i_n\}$ ，代表所有商品之集合。
- (3) 群體(Group)：泛指由二個或以上經節點關聯之行為鏈結，群體成員間制定強弱關

係。其中 $G \subseteq U$, G 表示預推薦之群體, $|G|$ 代表 G 之成員個數。

- (4) 總評比(Rating): 每位 MU 之喜好程度, 讓推薦資料更具可信度及說服力, 方便使用者作預測評估。總評比為大眾評比及個人評比搭配, 若無個人之參考評比情況取大眾評比為依據。

$R_{u,t,i}^i$ 定義為使用者 u 在時間點 t 於 Cell l 對某行為事件 i 之評比, i 可以為單體或表示群體行為鏈結。

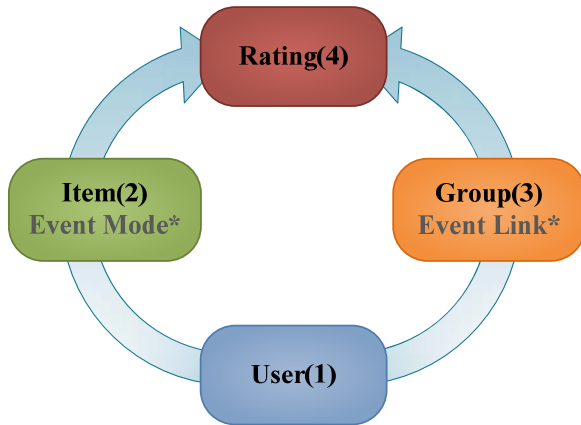


圖 5 架構問題定義與描述

3.1 事件樹建構

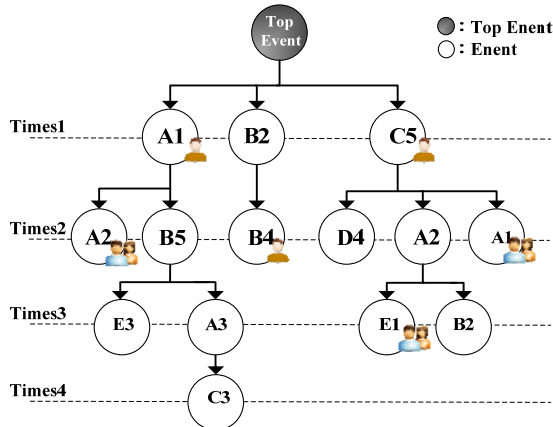


圖 6 索引事件樹

考慮到使用者行為具有目標導向，完整的行為事件包含關聯規則與組合特性，以樹狀結構做行為動態調整，事件樹亦有效快速存取使用者之行為為模式。如圖 6 中 MU 於 Cell A 發出項目 1 的查詢請求，接著移動到 Cell B 做項目 5 之查詢動作，伺服器記錄 MU 所在位置及查詢服務項目，能夠做後續預估該使用者相關行為資訊。在進行事件樹有效維護上，若單純對

結構所有已知事件節點進行調整，忽略關聯事件時效性及可靠度，會造成樹狀結構額外沒有必要的維護成本，因此妥善修剪事件樹結構顯得相對重要。如圖 7 所示，插入位置索引標籤 (Location Index) 與項目索引標籤 (List Index)，其意義在於大部分事件節點彼此間皆有鏈結關係，某父事件節點觸發後才能觸發執行子事件，如果在新增事件節點時插入索引標籤，如此於每次樹結構做調整時，能快速搜尋某事件節點相對應之鏈結，從對應鏈結取得該關聯之 MU 連續行為，亦透過位置標籤索引，幫助辨識 MU 所在地理位置。行動運算環境中，使用者行為觸發是經由位置及需求要素構成，項目索引編號對應該通訊細胞內相關查詢服務，以圖 7 行為 2 為例：編號 1 表示便利超商，編號 5 代表加油站，編號 3 代表車站，就構成使用者完整的行為鏈結 (便利超商 → 加油站 → 車站)。

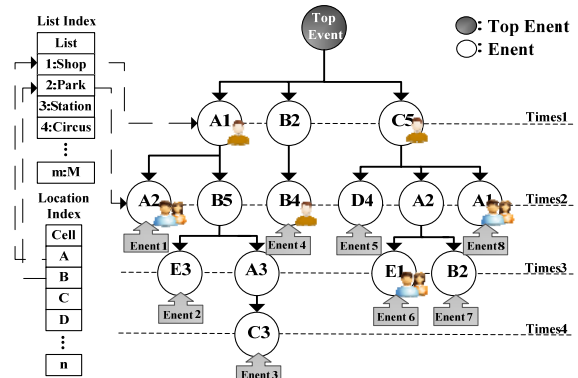


圖 7 階層式事件索引標籤

3.2 階層結構調整

樹狀結構建置越龐大則事件節點數就越多，造成整個架構演算效率大幅降低，造成多餘的維護成本，因此初步做事件結構調整，針對較弱環節進行改善，提高行為可靠度。如圖 8、圖 9 所示，步驟是將行為事件列為矩陣型式，由頂端事件逐次往下展開，找出該事件鏈結關聯性，每次輸入事件以矩陣一行中增加一元素表示，增加事件群數目的延伸。由頂端事件開始，第一次搜尋之事件 A1、B2 及 C5 於矩陣排序成行，接著圈選第二次搜尋行為鏈結 A1 的 A2、B5，B2 之鏈結事件 B4，亦使成行；圈選第三次搜尋之事件，成第三個矩陣，以此類推。最後列成以二行為事件為一組之事件組合，優點是在調整過程找尋重複觸發事件，例如：用戶由 Cell A 查詢便利超商 (編號 1)，接著移動至 Cell B 的車站 (編號 3)，與從 Cell B

車站移動到 Cell 便利超商,其實兩者是重複觸發事件,未調整前的樹狀結構會儲存兩筆數據,調整過後由樹狀矩陣中額外做熱門度統計,熱門度是依據該節點有多少用戶人數從事

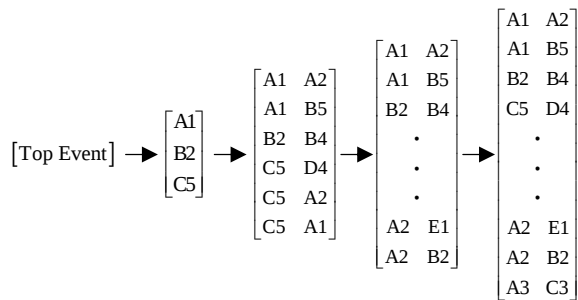


圖 8 樹狀矩陣之事件群延伸例圖

該行為事件而定,因此越熱門行為節點代表越多人進行過,反之亦然,重複事件使該行為事件增加熱門程度。以行為鏈結中父與子節點之熱門度,做兩者間關聯程度計算,取得最後將矩陣中所得結果構成為樹狀階層,如圖 10 所示。

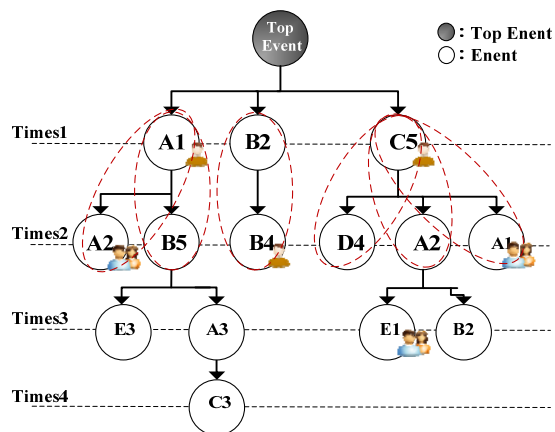


圖 9 索引事件樹階層搜尋圖選

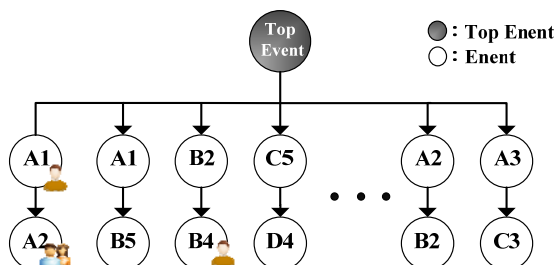


圖 10 調整後索引事件樹

3.3 行為鏈結分享機制

考慮到使用者進行同樣行為節點,卻會因不同組合方式造成相異的行為鏈結,為了避免遠端伺服器重複計算關聯度,系統需以階層索

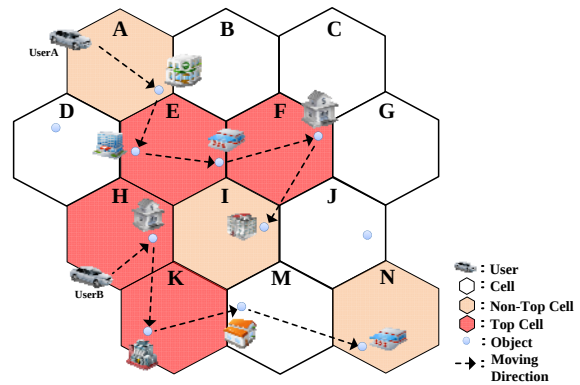


圖 11 熱門鏈結行為推薦

引方式架構,熱門度較高之行為鏈結進行內容過濾時,會將前後關聯之行為分享推薦給使用者,如圖 11 所示,UserA 之 Cell E、F 為高度熱門度行為事件的區域,進行分享事件鏈結推薦時,將具鏈結的 Cell A 與 Cell I 區域內相關需求服務,推薦給其他使用者查詢,並思考導入行為狀態及個人喜好為考量。

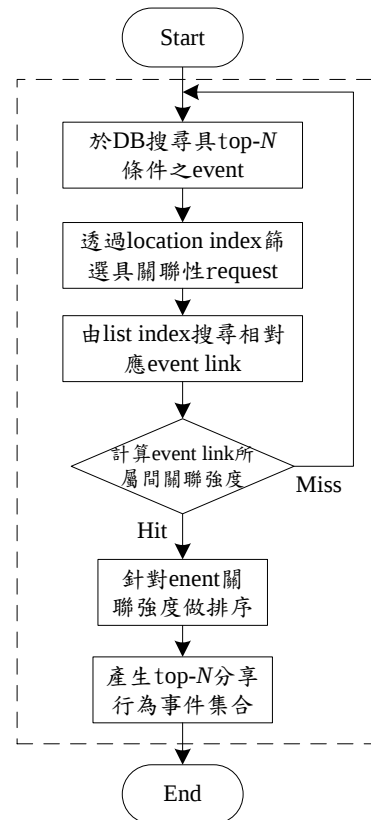


圖 12 行為鏈結分享機制流程圖

3.4 事件群推薦

由連續事件所組成之行為鏈結，在計算關聯程度價值時，會比降低其關聯行為強度，以圖11為例，UserA完整的行為鏈結是由5項事件所組合，若依照傳統方式去評估該鏈結，其中一項成員(便利商店)之查詢服務，反而忽略掉事件對於其餘成員關聯強度。鑒於此關聯程度問題，重新思考以群體推薦作為預估評比。本論文中提出ELRR推薦方法，一般的個人化推薦無法滿足所需的期望，以群體評比推薦是為了達到多項類別之推薦目的。因群體成員每項事件能夠反應出不同的關聯強度及影響力。由相似行為中所篩選出較理想之同好群(Neighbors)，藉此彌補內容過濾中資料稀疏的問題。如圖13所示，為了預測 K_1 之評比分別進行以下步驟：

- (1) 搜尋群體 G 中尚未評比過事件集合 K_G 。
- (2) 對每樣事件 $k \in K_G$ ，預測評比 $R_{G,k}$ 。
- (3) 列出與預測評比對應相似度，搜尋與 k_1 近相似之事件。
- (4) 由相似商品中篩選出較理想作為 k 的同好群 $C_G = \{c_{i1}, c_{i3}, c_{i7}\}$ 。
- (5) 由群體 G 中具高度相似度之同好群 C_G ，對 K_1 預估未知評比，能獲得較高可靠度。

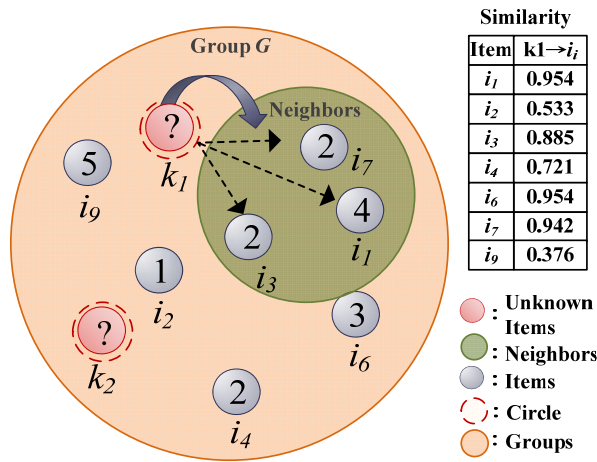


圖13 搜尋未評比之商品群示意圖

若發生同好群內成員有未知評比狀況，如圖14所示，進行以下步驟：

- (1) 當同好群 C_G 無評比， $c \in C_G$ 。
- (2) 取現有評分之同好群成員作樣本，以個體推薦之演算法對其他未知評比個別做評比預估。

- (3) 其中發生成員 c_{i7} 相似關聯程度偏低情況，是故不取 i_7 之相似度做評分樣本。
- (4) 最後將 i_1, i_4 視為 k_1 之同好群事件，依此推估 k_1 之評比 $R_{G,k}$ 。

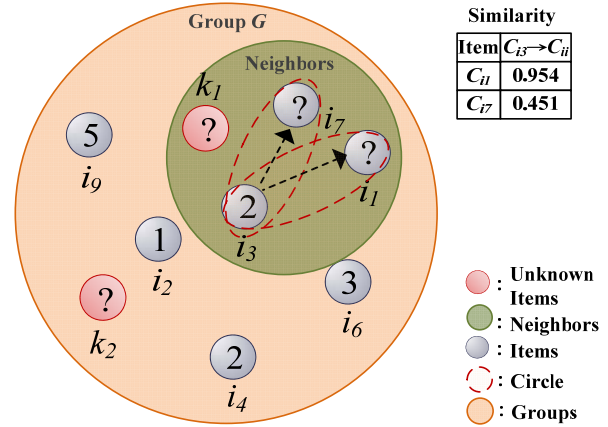


圖14 同好群成員無評比示意圖

4. 模擬與分析

本論文提出之行為鏈結推薦策略與其他相關策略做比較，模擬分析以JAVA程式語言進行開發，行動環境假設為 $N \times M$ 個細胞群所構成的典型個人通訊網路，環境中細胞類型定義為 $\leq 1\text{Km}$ 。系統中預測推薦項目評比部分，以10位評估者所產生之連續性行為樣本為依據，擷取評估者所給予的評比，驗證與模擬總評比之平均絕對誤差(Mean Absolute Error, MAE)及準確率(Precision)，如公式(3)(4)所示。評估本研究在預測使用者個人評比時之準確度。

4.1 Mean Absolute Error

$$MAE = \frac{\sum_{i=1}^N |p_i - r_i|}{N} \quad (3)$$

其中， MAE ：預測評比與實際評比之離散程度。

p_i ：Item $_i$ 之預測評比。

r_i ：Item $_i$ 之實際評比。

N ：總Item的數量。

由MAE值之大小，可瞭解預估評比及實際評比之離散程度，其值越小代表誤差越小，其結果較佳，反之亦然。在許多不同應用領域上皆採用平均絕對誤差做評估準確度的指標。

4.2 Precision

$$Pre = 1 - \frac{\sum_{i=1}^N |p_i - r_i| / \max(|R_{\max} - r_i|, |R_{\min} - r_i|)}{N} \quad (4)$$

其中， Pre ：預估評比之正確率。

P_i ：Item_i之預測評比。

r_i ：Item_i之實際評比。

N ：總Item的數量。

$R_{\max}$ ：最大可能的評比值。

$R_{\min}$ ：最小可能的評比值。

Precision用來評估預測評比相對於實際評比之正確率，並經過正規化步驟，因此能夠具體瞭解實際的正確率多寡，其值介於0到1之間，該值越低代表正確率越低，反之亦然。本論文設定最大可能評比為5，最小可能評比值即為0，評比0不代表低評比，而是使用者未對Item給予評價。

5. 結論及未來工作

索引事件樹不會破壞完整行為規則之特性與彼此間關連性，雖樹結構可能因調整後而忽略其重要行為關聯，但以兩項事件節點為基礎的索引事件樹，較調整前更能突顯行為間事件強度，後續做預測評比亦能避免演算效率降低之情況，限制樹狀結構深度讓分支點向外擴增，對搜尋分享節點與樹結構維護及新增皆較以往快速。ELRR考量MU連續性位置移動與查詢服務之特性，探勘事件間隱含關聯價值，掌握更多個人行為特徵，依據預估評比之熱門喜好幫助使用者篩選適合個人化項目，提供更貼近個人需求之系統。

後續研究會深入探討當樹狀結構越大情況下，會造成系統演算效率降低，需建置妥當的分群機制，以限制事件樹的結構，雖藉內容導向過濾解決推薦系統原本存在的稀少性(Sparsity)及擴充性(Scalability)問題，但相對產生評比陣列的模糊預測評比，探討進行行為資訊分類，找到具有價值可利用之樣式(Pattern)，在能接受範圍內提升該數據精確度。

參考文獻

- [1] D. -L. Lee, J. Xu, B. Zheng, W.-C. Lee, "Data Management in Location-dependent Information Services," *IEEE Pervasive Computing*, vol. 1, Issue 3, pp. 65- 72, 2002
- [2] V.-D. Personè and V. Grassi, "Performance Analysis of Caching and Prefetching Strategies for Palmtop-based Navigational Tools," *IEEE Trans. on Intelligent Transportation Systems*, vol. 4, Issue 1, pp. 23-34, Mar. 2003.
- [3] K. -Y. Huang, C. -L. Chen, and C. C. Tuan, "Fuzzy-based Routing for Wireless Sensor Networks," in *Proc. of the 11th Mobile Computing Workshop 2005 (MC'05)*, Taoyuan, Taiwan, pp. 688-692, Mar. 2005.
- [4] Agrawal, R., T. Imielinski, and A. Swami, "Mining association rules between set of items in large databases," in *the Proc. ACM SIGMOD Conference on Management of data*, pp. 207-261, May 1993.
- [5] Agrawal, R. and R. Srikant, "Fast Algorithms for Mining Association Rules in Large Databases," in *Proc. of the 20th International Conference on Very Large Data Bases*, pp. 487-499, 1994.
- [6] W. Yu, X. Wang, F. Wang, E. Wang, and B. Chen, "The Research of Improved Apriori Algorithm for Mining Association Rules," in *Proc. Of the ICCT 11th IEEE International Conference on Communication Technology*, Hangzhou, China, pp. 513-516, Nov. 2008.
- [7] C.-H. Yun and M.-S. Chen, "Mining Sequential Patterns in A Mobile Commerce Environment," *IEEE Transactions on Systems, Man, and Cybernetics Part C: Applications and Reviews*, vol. 37, Issue 2, pp. 278-295, Mar. 2007.
- [8] V.-S.-M. Tseng, K.-W.-C. Lin, "Mining Sequential Mobile Access Patterns Efficiently in Mobile Web Systems," in *Proc. of the AINA 19th International Conference on Advanced Information Networking and Applications*, vol. 2, pp. 762 - 767, Mar. 2005.
- [9] J.-L. Herlocker, J.-A. Konstan, L.-G. Terveen, and J.T. Riedle, "Evaluating Collaborative Filtering Recommender Systems," *ACM Transactions on Information Systems*, Vol. 22, Issue 1, pp. 5-53, Jan 2004.