

以學術部落格為主之個人化推薦系統

(The Personalized Recommendation for Academic Blog)

李麗華
朝陽科技大學
資訊管理系教授
lhli@cyut.edu.tw

李富民
朝陽科技大學
資訊管理系副教授
fmlee@cyut.edu.tw

詹尚驥
朝陽科技大學
資訊管理系研究生
s9514627@cyut.edu.tw

周裕健
朝陽科技大學
資訊管理系研究生
s9614644@cyut.edu.tw

摘要

根據 Technorati[29]於 2007 年報告指出，截至 2007 年全世界部落格總數量已增至 7200 萬，此一現象表現出部落格已受大眾普遍的接受。然而，快速增長的部落格也造成了部落格文章數量的暴增，亦即增加了瀏覽者對於文章資訊過載及資訊查找不易的困擾。除此之外，部落格文章的個人化推薦服務也是現今部落格所缺乏的。

有鑑於此，本研究提出一個學術文獻推薦系統(BARS)，首先依本體論階層概念建構瀏覽者偏好興趣階層圖，再運用自適應共振理論網路，將相似偏好興趣瀏覽者進行分群，分群後透過推薦系統(Recommendation System, RS)中協同過濾法及內容導向法，將具有及非具有鄰居群的瀏覽者予以推薦。

本研究欲達成的目標及貢獻為：(1) 提升文章資訊搜尋效率，(2) 降低文章推薦冷啟始問題，(3) 提昇個人化文章推薦效能。

關鍵詞：類神經網路、自適應共振理論網路、推薦系統、本體論、部落格

Abstract

According to Technorati's 2007 report, the blog sites is up rising to 72 millions and the popularity of bloggers has drawn many attention. The vast amount of blog information has brings the phenomenon of information overloading which is not handled by the blog function yet. In addition, the personalized recommendation service, which should be provided, is also not incorporated in the blog function now.

To better service the bloggers and to overcome the above problems, this research proposes a Blog Article Recommendation System (BARS) which provides personalized article recommendation based on blogger's interests. This research adapts the ontology technique in

BARS to construct a personal preference tree for understanding blogger's interests. The ART (Adaptive Resonance Theory) network is also utilized to cluster the group with similar interests. In order to find the similar preference between target blogger and the corresponding neighbors, we apply the Collaborative Filtering (CF) technique to generate the recommendation. The "cold-start" problem, i.e. lacking of blogger's usage data at the very beginning, is handled by combining ontology and Content-Based (CB) filtering method to infer the potential preference in BARS. The purpose of this research is as followings.

- (1) To solve the problem of information overloading.
- (2) To propose the cold-start problem when making the recommendation.
- (3) To build the BARS for personalized blog-article recommendation.

Keyword: Artificial Neural Network(ANN), Adaptive Resonance Theory(ART), Recommendation Systems(RS), Ontology, Blog

1. 緒論

微軟董事長比爾·蓋茲指出[2]，部落格是繼 E-mail、BBS、即時訊息(如 MSN Messenger)之後，第四個改變世界的網路殺手級應用。部落格是一種新興的個人網路日誌，它提供一個線上個人平台，讓每位網路瀏覽者都可擁有自己的部落格空間，每個部落格皆可分享私人的知識、經驗、甚至是影像等，除此之外，不同部落格間資訊是可彼此共享的。

現今大眾型的入口部落格網站如 MySpace[23]與 Blogger[8]等，這些部落格多應用於個人的生活記事、心情寫真，較少有學術研究論文資料閱讀心得的部落格。此外，部落格內的網站分類也過多且無一明確公定好的

類別名稱[17]，如此容易造成瀏覽者的混淆、不易查找所需的文章。除此之外，根據美國部落格搜尋網站公司 Technorati[29]在 2007 年報告指出，自部落格萌芽起到現在雖不足 5 年，但部落格文化已經在全球開花。2005 年 3 月的部落格數量還僅為 800 萬，而至 2007 年 3 月底，全球部落格數量已攀升到 7200 萬。此外，部落格文章的數量也持續增加，2005 年 3 月時僅 50 萬份量，而至 2007 年 3 月份的統計數量已高達 140 萬。

在如此龐大的部落格資訊流中，產生了文章資訊過載(Overloading)，自然也讓網路瀏覽者產生搜尋及查找文章日益複雜的問題[31]，現今有許多學者針對網路文章議題已提出許多相關的推薦系統(Recommendation System)之技術與方法[18][22][31]，推薦系統除廣泛用於電子商務上，針對有形商品推薦外，非實體商品(如文章、電影、音樂等)也經常運用推薦技術，如 Amazon[7]、Group Lens[15]均有利用。

針對推薦系統在不同領域上的應用，Sarwar[28]學者將推薦系統的技術分為二種方法，分別為協同過濾法(Collaborative Filtering method)與內容導向法(Content-based method)。這二種方法以協同過濾法較常被推薦系統所使用[32]，它的方法主要是先找出所有使用者的興趣偏好，將這一群使用者進行分群，之後透過同群者的鄰居群概念將與具有相似偏好興趣的鄰居進行分析、預測，進而推薦出使用者可能產生興趣的項目(Item)。

然而進行協同過濾法前，需先進行分群的動作，有關分群的研究已有許多學者提出，其中類神經網路(Artificial Neural Network, ANN)[5]中非監督式學習(Unsupervised learning)的自組織映射圖理論(Self-Organization Map, SOM)及自適應共振理論(Adaptive Resonance Theory, ART)，此種學習法通常用於分群聚類用。目前已有許多研究針對推薦系統及 SOM 作進一步的分析與應用，如 Y. B. Cho 等學者提出探勘顧客購買行為的變化[11]，並且利用 SOM 來幫助系統為顧客分群，之後依分群結果來顧客作推薦。

目前協同過濾法已成為重要且受歡迎的個人化(Personalized)推薦方法[19]。然而在制定個人化推薦方法前，系統需事先收集所有使用者的個人輪廓(Profile)，透過個人輪廓資料可讓系統得知一個使用者過去的歷史偏好及興趣。而使用者偏好通常包含了個人身份資料、網頁瀏覽習慣及興趣的個人資料[6]。使用者個人輪廓

有多種不同表達方式，如 Middleton 等人[22]提出以本體論(Ontology)技術來建構使用者的個人輪廓，他們透過一套名為 Quickstep 的論文推薦系統，此系統透過網頁圖形化介面，讓使用者在此介面上接收、提供資訊，之後利用本體論技術來收集使用者的偏好，最後並依本體論概念推論法來推薦學術文章給使用者。

現今推薦系統針對文章推薦方面已有許多研究，然而就大眾型入口部落格網站文章的個人化推薦來看，目前尚未有人進行研究，原因大致可歸納至下列二點：

- (1) 部落格入口網站文章類別與其底下個人部落格私下定義的文章類別不一致[17]，導致同義異名無法根據其類別來作推薦。
- (2) 部落格內個人網站眾多，文章類型大多偏屬心情記事、經驗分享等。較屬短暫抒發、較無收藏價值之文章，因此也較無推薦的需求。

為了能讓有意義且有價值的知識進行分享、傳播、且主動地提供所需的瀏覽者，故本研究提出以推薦系統技術來減少部落格文章之資訊過載、文章查找不易的困境，且利用個人化方法主動地將文章推薦給使用者，建構一個學術文章部落格及其回饋系統，用來接收使用者的回饋，本研究已驗證本研究之推薦效果。本研究以學術文章部落格為平台引入推薦系統的技術，其原因有下列二項優點：

- (1) 提高瀏覽者與部落客、部落格網站互動性，
- (2) 吸引更多部落客前來進駐，提供更專業、更有價值性的文章。

在本研究中，提出針對部落格建制一個能提供個人化的部落格文章推薦系統(BARS)。此系統讓瀏覽者可更便利閱讀其喜好及相關文章。其建立的系統首先會收集瀏覽者的過往閱讀資訊，依本體論方法建構使用者個人輪廓資料並建立本體論個人偏好階層圖，再透過類神經網路的 ART 分群法為每位瀏覽者進行分群，最後依分群的結果來判定採用協同過濾法或內容導向法(本體論推論為基底)，進行文章推薦。

而本研究所建構的部落格文章推薦系統(BARS)是以類別為主，加以運用推薦系統、自適應共振理論、本體論等方法，以此方法所建立的文章推薦機制，其最大的預期綜效在於希望能提供一個符合瀏覽者需求的文章推薦機制；以下為本研究之目的及說明。

- (1) 提供瀏覽者個人化部落格文章推薦，解決資訊過載及搜尋文章耗時之問題。
- (2) 降低瀏覽者冷啟始(Cold start)問題：利用本體論概念推論方法來協助解決無法找到相似鄰居群的瀏覽者之推薦問題。
- (3) 以本體論個人偏好樹來表達瀏覽者個人輪廓(Profile)，並透過類神經網路的自適應共振理論網路的分群結果來挖掘出瀏覽者的潛在偏好，並推薦合適的文章給予瀏覽者。

2. 文獻探討

2.1 推薦系統

A. M. Rashid et al.[25]等學者對推薦系統定義為是一種可以在複雜的資訊環境下幫助使用者制定決策，且可依據系統對使用者及項目的瞭解來產生推薦給使用者。推薦系統依其推薦機制可分二種方法[10]，分別為內容導向式推薦(Content-based Filtering)與協同過濾式推薦(Collaborative Filtering)，以下將分別詳述其定義、方法及其限制。

2.1.1 內容導向式(Content-based Filtering, CB)

內容導向式推薦[10]主要是從資訊擷取(Information Retrieval)技術中延伸而來，此種方法主要是針對使用者偏好的項目進行分析項目之屬性與特徵，並且比對使用者的個人輪廓，配對出合適的項目推薦給使用者，例如可應用於數位內容相關資訊[20](客製化新聞服務及知識管理文件等)。雖然，此方法可以不用依賴其它瀏覽者也能產生推薦效果，但也存在著一些限制的範圍[30]，如下：

- (1) 內容導向式推薦僅侷限於機器可讀的文字(Text)格式，其他如：聲音(Sound)、影像(Video)、圖片(Photographs)、藝術類(Art)等多媒體項目無法經由系統做自動化分析及推薦，
- (2) 內容導向式推薦僅能依使用者過去的歷史偏好記錄與目標項目間的關聯性來進行推薦，無法有彈性地找出其潛在偏好，
- (3) 內容導向式推薦無法過濾相同項目間的品質、風格及觀點。例如，同時有兩篇具有相同名稱的文章，一篇品質較佳，另一篇則較差。此時，系統仍視這兩篇文章為可推薦文章，無法辨識其好壞。

為改善上述推薦方式所帶來成效不彰的影響，S. Puntheeranurak 等學者[24]提出利用內容導向式推薦來預測瀏覽者的個人偏好，並結

合以瀏覽者及項目為基礎的協同過濾式推薦(Collaborative Filtering)方法來加強內容導向式推薦能力不足的問題。

2.1.2 協同過濾式(Collaborative Filtering, CF)

協同過濾式推薦[21][27][30]是目前較為廣泛的推薦技術，主要是依據使用者間對項目的偏好程度來進行相似度的分群，並將每個群聚中目標使用者找出最相似偏好的鄰居使用者依其偏好來予以推薦。此方法與內容導向式推薦的差異在於無需分析使用者偏好下的項目內容，只需找出具有一定相似程度偏好的使用者即可進行分析、推薦。根據 Sarwar et al.[27]針對協同過濾式推薦內的基本推薦程序，共分為三個步驟，分別為輸入資料表示(Representation of input data)、鄰居群建立(Neighborhood formation)、推薦產生(Recommendation generation)等。

雖然協同過濾式推薦為目前較受歡迎且較成功的推薦系統，但它仍然有些限制存在[22][27]，分別為稀疏性(Sparsity)及冷啟始(Cold start)，詳述如下。

- (1) 稀疏性：在推薦系統進行推薦時，如果存在的資料量很大，但是被使用過的可能只是部份少數時，就會較難分析及推薦。例如在電子報網站中，若是被瀏覽者閱覽過的文章不到百分之一，其產生的交易量過少，將無法計算出相似的鄰居群且推薦效果會非常差。
- (2) 冷啟始：通常是針對新使用者或新項目而言，當一個新使用進入後，因為系統不了解使用者的偏好，所以無法推薦項目，而在項目中則是因為若未有人閱讀過文章的文章，將無法被系統分析推薦。

2.2 類神經網路(Artificial Neural Network, ANN)

類神經網路依照學習策略之網路型態[5]可分作監督式學習(Supervised Learning)、非監督式學習(Unsupervised Learning)、聯想式學習(Associate Learning)，詳述如下。

- (1) 監督式學習(Supervised Learning)：從問題領域中取得輸入變數及輸出變數的值作為訓練範例，並從中學習輸入變數與輸出變數的內在對映規則，以應用於新的案例。
- (2) 非監督式學習(Unsupervised Learning)：從問題領域中僅取得輸入變數的值作為訓練範例，並從中學習範例的內在聚類規則，以應用於新的案例。

- (3) 聯想式學習(Associate Learning Network)：從問題領域中取得狀態變數的值作為訓練範例，並從中學習範例的內在記憶規則，以應用於新的案例。

本研究採用非監督式學習的網路作為本研究的分群工作角色。非監督式學習包含自組織映射圖網路及自適應共振理論網路，而自組織映射圖網路需透過拓撲圖形的觀察來決定群數，較難以明確判斷分群結果；自適應共振理論網路則不需透過圖形識別即可決定出明確的分群結果，故本研究採用自適應共振理論網路作為分群的工具。

2.2.1 自適應共振理論網路 (Adaptive Resonance Theory, ART)

自適應共振理論網路(Adaptive Resonance Theory, ART)[5][12]其基本原理可追溯自認知學(Cognitive)。它是一種非監督式學習網路模式，最早是由 1976 年葛羅斯伯(S. Grossberg)[14]將此網路模式作了重要的整合與擴充，之後經由卡本特(G. A. Carpenter)[9]加以發揚光大。而在 ART 中具有下列二項特性。

- (1) 穩定性：當新的事物輸入時，舊的事物應適當地保留。
- (2) 可塑性：當新的事物輸入時，應迅速地學習。

自適應共振理論網路架構包括輸入層、輸出層及網路連結層，此架構功能、圖示(圖 1)如下所示：

- (1) 輸入層(Input layer)：用以表現網路的輸入變數，即訓練範例的輸入向量，或稱特徵向量，其處理單元數目則依問題而定。
- (2) 輸出層(Output layer)：用以表現網路的輸出變數，即訓練範例的聚類，其處理單元數目在學習過程之初只有一個，隨著學習的進展會逐漸增加，最後會穩定在一定的數目，學習過程即結束。

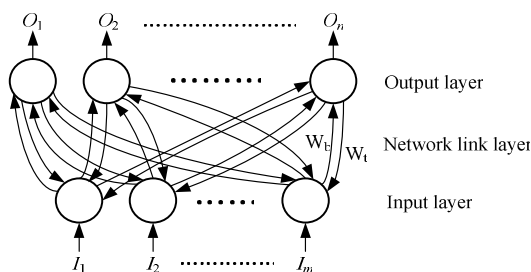


圖 1 自適應共振理論網路架構(資料來源：[5])

- (3) 網路連結層(Network link layer)：自適應共振理論網路的每一個輸入層單元與輸出層單元間有二根網路連結，此一特性與其它類神經網路模式較為不同。

2.3 本體論(Ontology)

透過本體論中的基本元素，即概念與概念間的關連，作為描述真實世界的知識模型。根據 Grouber[13]於 1993 年所提出的定義，指出「本體論為一種共享概念化且正式的、明確的表示形式」(An ontology is a formal, explicit specification of a shared conceptualization.)。針對此段定義，說明如下：

- (1) 正式的(Formal)：本體論是機器可讀、可理解的。
- (2) 明確的(Explicit)：是以明確的方式表示出本體論的概念形態及限制。
- (3) 共享(Shared)：本體論為一群體性共享之概念。
- (4) 概念化(Conceptualization)：是指對現存某個現象或領域的確定現象之相關概念抽象模型。

現今已有多位學者[3][6][22]結合本體論概念推論方法來強化推薦系統之推薦效能，這些學者大多是利用本體論階層概念應用至使用者的個人輪廓中，將使用者原有的個人偏好加以延伸、探索，以其推測出使用者可能的「潛在偏好」，而這些潛在偏好通常是推薦系統中相當重要的指標[6]。此外，本體論也可有效改善內容導向式推薦的一些限制[16]。

2.4 詞彙分析相關技術及方法

基於文章分類方法，本研究欲透過文章關鍵字擷取的技術及重要性的評量來決定每篇文章隸屬於何種類別。本研究採用中央研究院資訊科學詞庫小組的 CKIP 中文斷詞系統作為文章關鍵字擷取的方法，而關鍵字的重要性則透過 TF(Term Frequency) 來評量。依關鍵字擷取的技術，曾元顯[4]將此區分為詞典、文法分析及統計分析三種。

2.4.1 CKIP 中文斷詞系統

CKIP 中文斷詞系統為一具有新詞辨識能力並附加詞類標記的選擇性功能之中文分詞系統。此一系統包含約 10 萬詞的詞彙庫及附加詞類、詞頻、詞類頻率及雙連詞類頻率等資料[1]。本研究採用中央研究院詞庫小組所研發的 CKIP 中文斷詞系統，來處理本研究的文章

資料。

2.4.2 TF Operation

TF 全名為 Term Frequency。它是一種統計分析方法，用以評估一詞彙(Term)在一篇文章中出現的「詞彙次數」，其值越高代表該詞彙在文章中越具有重要性[26]。

3. 研究方法

本研究提出一個「學術領域文章」的部落格網站推薦系統 BARS(Blog Article Recommendation System)，主要是以資訊科技(Information Technology, IT)領域的文章為網站主題與內容，BARS 系統的主要目的是在於提供瀏覽者(Browser)對於資訊科技相關之文章資訊，且推薦給具有此需求的瀏覽者。本研究結合類神經網路的自適應共振理論(Adaptive Resonance Theory, ART)網路與本體論(Ontology)的概念來推論瀏覽者的興趣偏好，最

後制定出 Top-N 推薦清單供瀏覽者參考閱覽，以下就本研究的方法逐一介紹。

3.1 BARS 系統架構

本研究所提的 BARS 系統如圖 2 所示，主要可分為三個模組，第一個模組為文章分類模組(Article Classifier Module, ACM)，此模組目的在於自動的將網站內所有文章進行分類，以利於下一個模組進行比對動作；第二個模組為本體論建構模組(Ontology Construction Module, OCM)，此模組的目的在於將所有的瀏覽者的個人輪廓(Profile)建構成個人偏好樹形式，用此個人偏好樹來描述一個瀏覽者的興趣偏好；第三個模組為文章推薦模組(Article Recommendation Module, ARM)，此模組負責將所有建構好的個人偏好樹進行分群之動作，並透過鄰居群的偏好來進行偏好推薦及利用本體論概念推論將文章推薦給使用者。

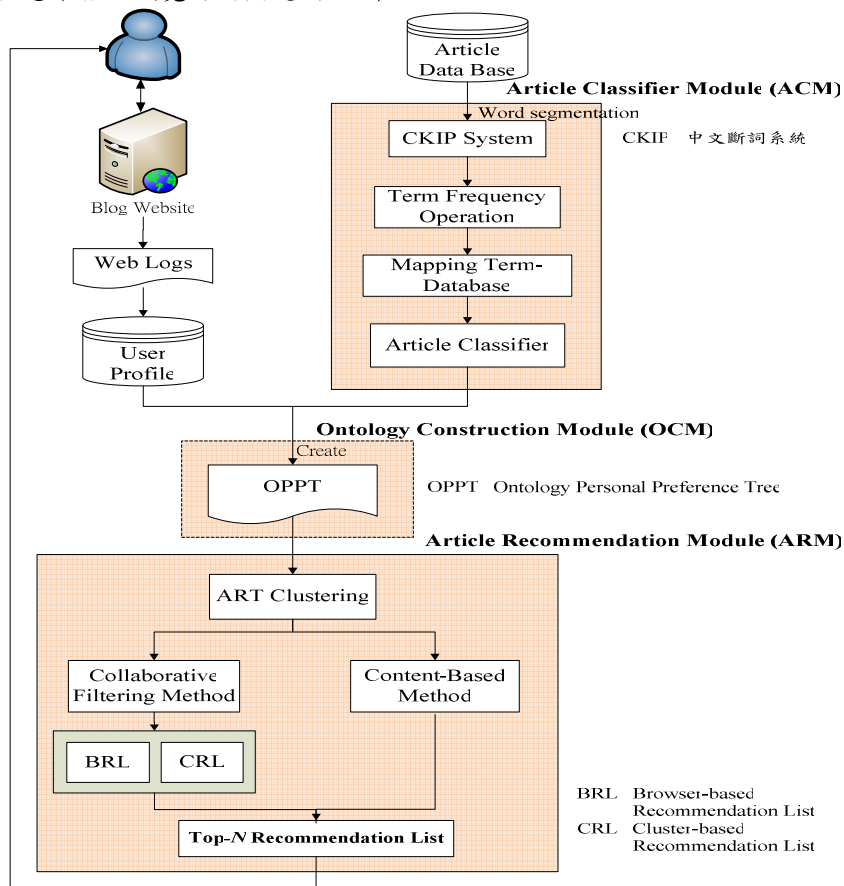


圖 2 部落格文章推薦系統(BARS)架構

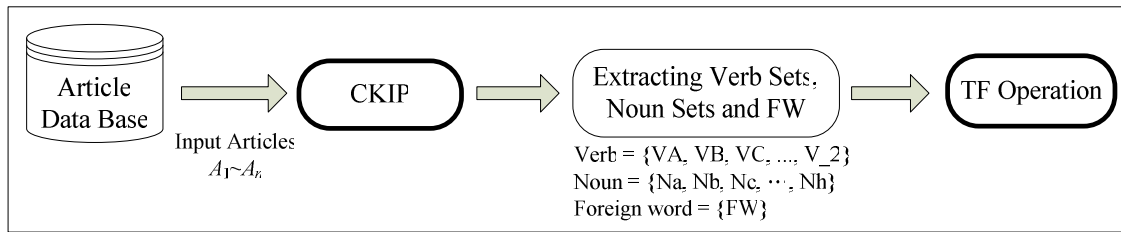


圖 3 CKIP system 文章斷字流程

3.2 文章分類模組(Article Classifier Module, ACM)

文章分類模組可分四個處理步驟，第一個步驟需先從文章資料庫中擷取出本網站系統內所有的文章，接著透過 CKIP 中文斷詞系統來為每篇文章進行斷詞的動作；第二步驟則是利用 TF (Term Frequency) 來計算出每個詞彙的頻率，藉此找出具有意義性的詞彙；第三步驟將找出具有重要性的詞彙與本研究事先定義好的詞彙庫進行比對，找出真正具有代表性的詞彙；第四步驟開始依每個文章所代表的詞彙來找出符合的類別。以下將詳述此模組各個步驟的功能。

步驟 1：CKIP System，在此步驟中採用了中央研究院資訊科學所詞庫小組的 CKIP 中文斷詞系統來為本研究網站文章進行斷字(Word Segmentation)處理。其流程如圖 3 所示，此外，CKIP 系統斷字處理後，亦會對每個詞彙進行詞類分析，並給予適當的詞類作為標記識別。因此，本研究在此步驟進行完斷字後，將帶有動詞集、名詞集及外文標記之字詞擷取出來作為下一步驟詞頻計算的依據，而其他詞類如連接詞、感嘆詞、介詞、語助詞等一律移除，不予以計算。

步驟 2：TF operation，進行完斷字處理並取出動、名詞及外語標記的詞類後，此步驟即會利用 TF(Term Frequency)來計算出每個關鍵字出現的頻率及關鍵字出現在文章的重要性與代表性程度為何。

步驟 3：Mapping Term-Database & Article Classifier，詞彙庫(Term-Database)包含了本實驗網站系統內所有文章類別的關鍵字及該類別經常出現的詞彙，並將之整理成一詞彙庫。

步驟 4：找出具有最大詞彙出現次數的類別及文章 m 的隸屬類別，如公式(1)。

$$M A = M \underset{l}{A} X \{ K C_m(l) \} \quad (1)$$

$$l^* = l, \text{ if } M \underset{l}{A} X \{ K C_m(l) \}$$

$M \underset{l}{A} X \{ K C_m(l) \}$ 為對所有 l 類篩選出具有最大詞彙出現次數

$l^* =$ 若 $K C_m(l)$ 為最大值，則令 l 為 l^*

3.3 本體論建構模組(Ontology Construction Module, OCM)

在本體論模組中，主要的功能在於為每位瀏覽者過去閱讀過文章的歷史記錄檔與已分類好的文章類別進行相對映，並且以樹狀的方式來呈現出各別瀏覽者的瀏覽記錄，此方法即稱為本體論個人偏好樹(Ontology Personal Preference Tree, OPPT)。其結構如圖 4 所示，而有關 OPPT 建構步驟說明如下。

步驟 1：讀取瀏覽者個人輪廓(UP)來建立本體論個人偏好樹。

步驟 2：初始化第 u 個 OPPT，並判斷此 OPPT 是否為空集合(新加入的瀏覽者，代表此 OPPT 無任何新瀏覽者相關資訊)。

步驟 3：以遞迴的方法由上至下將其他節點(瀏覽者所閱讀過的文章)依文章類別順序加入至

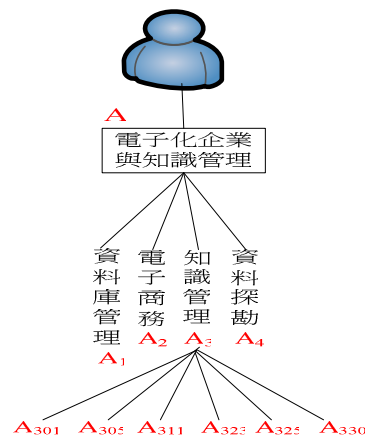


圖 4 OPPT (列舉範例)

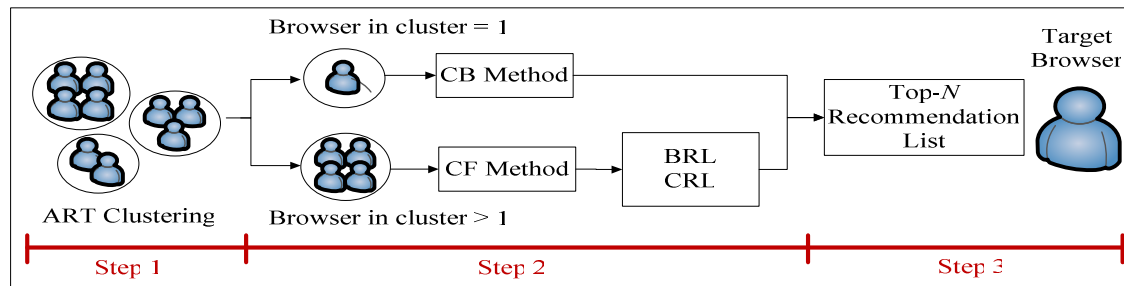


圖 5 ARM 推薦機制流程圖

OPPT，並下達一個判斷式來判斷是否有新進的重覆節點(瀏覽者重覆閱讀過的文章)。

步驟 4: 再次判斷是否所有新節點已加入完畢。

3.4 文章推薦模組(Article Recommendation Module, ARM)

文章推薦模組(ARM)可分為三個步驟，第一個步驟使用類神經網路的自適應共振理論網路(Adaptive Resonance Theory, ART)將瀏覽者的個人偏好樹進行分群的動作。

第二個步驟則為推薦機制的運作，已分群後可細分成二部份，首先針對具有鄰居群的瀏覽者，可透過推薦系統中的協同過濾法(Collaborative Filtering method, CF)，找到具的鄰居群將瀏覽者可能產生的興趣來進行推薦，其推薦方法採用本研究所提出的BRL(Browser-based Recommendation List)及CRL(Cluster-based Recommendation List)來產生推薦清單，其次針對不具有鄰居群的瀏覽者，因多數是興趣偏好與其他的瀏覽者不同，或者是新加入系統的瀏覽者，系統缺乏瀏覽者的相關資訊(即冷啟始問題)，因此這類型的瀏覽者，可利用推薦系統中的內容導向式法(Content-based method, CB)來挖掘出瀏覽者在過去的歷史記錄偏好中，哪些文章是瀏覽者所感興趣卻尚未閱讀過的文章。

第三個步驟則是為每位目標瀏覽者產生一份 Top-N 的推薦清單給予推薦。ARM 各個詳細步驟說明如下(如圖 5 所示)。

步驟 1: ART Clustering

此步驟是針對在瀏覽者分群之前，必須將所有的資料進行編碼轉換。而 ART 是採用 0 跟 1 兩種位元作為資料的輸入(Input)，因此本研究將瀏覽者於 OPPT 上的資料依閱讀後賦予屬性值，1 的位元代表該瀏覽者有閱讀過此文章類別，而 0 的位元則代表該瀏覽者沒有閱讀過該文章類別。基於上述的描述，每位瀏覽者的 ART 編碼表示法如公式(2)所示。

$$BRR = \{IC^l \mid 1 \leq l \leq L, L \in \mathbb{N}\}$$

$$IC^l = \{R_1^l, R_2^l, \dots, R_f^l\}, R_f^l \in [0, 1], R_f^l = \begin{cases} 1 & \text{if } R_f^l \text{ is read} \\ 0 & \text{otherwise} \end{cases} \quad (2)$$

BRR(Browser Reading Record)代表瀏覽者於各文章類別的瀏覽記錄； IC^l 為第 l 類文章主類別， R^l 為第 l 類文章子類別， f 則代表文章子類別的閱讀次數範圍(如 2~4、5~9、10 以上)。

透過 ART 編碼後，透過警戒值($0 \leq \rho \leq 1$)的測試可產生適當的分群結果，再依警戒值 ρ 的設定來找出所要之群數，分群結果公式如公式(3)所示， G^y 代表第 y 個瀏覽者群組， B_x 則代表群組內的瀏覽者。

$$G^y = \{B_1^y, B_2^y, B_3^y, \dots, B_x^y\}, 1 \leq x \leq X \quad (3)$$

步驟 2: CF and CB method

此模組的主要核心分別為協同過濾法(Collaborative Filtering method, CF)及內容導向法(Content-based method, CB)，此兩種推薦機制是目前推薦系統中的主要核心技術，在這步驟主要是接收上一步驟 ART 已分群完成的結果，接著依上述兩種推薦方法的需求各自取所需的資料進行文章推薦的動作。協同過濾法會去接收各個分群結果，並找出欲推薦的目標瀏覽者，找尋出該瀏覽者隸屬於哪一群中，並依據其周圍最相似鄰居群偏好給予文章推薦，進而運用本研究所設計的 BRL (Browser-based Recommendation List) 及 CRL (Cluster-based Recommendation List) 產生推薦清單。

為解決分群階段中所產生不具相鄰居群的問題，本研究提出另一項推薦方法為利用本體論概念推論的內容導向法，此方法即依據瀏覽者過去的文章閱讀記錄，找尋出過去閱讀次數較多的文章類別，再依點閱次數較高的文章且未被目標瀏覽者閱覽過的文章推薦給瀏覽者。而 BRL 及 CRL 方法的差異在於後者是在相似鄰居群中找出最相似的 Top-N 偏好文章類別，而 BRL 則是相反，其個別步驟如下說明。

步驟 2.1: BRL 處理流程是以瀏覽者本身偏好為基底的推薦清單，其處理流程如下。

步驟 2.1.1: 根據瀏覽者的 OPPT，搜尋出瀏覽者所有文章類別的閱覽記錄，並從中計算出瀏覽頻率較高的文章類別，其計算可透過門檻值的設定來決定，門檻值計算方式如公式(4)所示。

$$t = \frac{BRC}{BRCC} \quad (4)$$

t : 瀏覽者於該文章類別的門檻值。

BRC(Browser Reading Count): 瀏覽者過去所閱覽過的文章數量

BRCC(Browser Reading Category Count): 瀏覽者過去所閱覽過的文章子類別數量。

步驟 2.1.2: 從步驟 2.1.1 計算出瀏覽者閱覽頻率較高的文章類別中，找出該文章類別每篇文章的總閱覽次數(所有瀏覽者對該類別文章的總閱覽次數)，接著依據閱覽次數較高且未被目標瀏覽者瀏覽過的文章以 Top- N_w 的方式作推薦，Top- N_w 計算方式如公式(5)。

$$Top-N_w = \frac{CA}{TT} \quad (5)$$

Top- N_w 指的是 Top- N 可推薦的 w 篇文章數百分比； CA 則是該類別的所有文章數； TT 代表所有 Top- N 的總文章數。

步驟 2.1.3: 利用公式(6)來搜尋出瀏覽者與鄰近瀏覽者的共同偏好(Common Preference with threshold, CPT^y)文章類別集合。

$$CPT^y = \{NC^y - (NC^y - BC(t))\} \quad (6)$$

CPT^y (Common Preference with threshold): 瀏覽者與鄰近瀏覽者的共同偏好文章類別集合

NC^y (Neighbor Category): 鄰近瀏覽者 y 的主要偏好文章類別集合 $BC(t)$ (Browser Category)

步驟 2.1.4: 配對 $Article_{HC}\{Max_{top-N}\}$ ，將瀏覽者與鄰近瀏覽者的共同偏好文章類別集合中，取出瀏覽者未曾閱讀過且具有較高閱覽率(即點擊率)的文章配對給瀏覽者(閱覽率排名前 n 篇文章)。

步驟 2.2: CRL 處理流程是以群內相似偏好為基底的推薦清單，其處理流程如下。

步驟 2.2.1: 利用公式(7)計算出前幾項瀏覽者與鄰近瀏覽者最大的共同偏好文章類別集合。

$$CP = \{NC^y - (NC^y - BC(I))\} \quad (7)$$

CP -Top- N is the set of Top- N reading category

$BC(I)$: 瀏覽者所閱讀過的文章類別。

NC^y (Neighbor Category): 鄰近瀏覽者 y 的主要偏好文章類別集合 $BC(t)$ (Browser Category)。

步驟 2.2.2: 透過公式(5) Top- N_w 來決定瀏覽者偏好的文章類別中，每一個文章類別分別可以推薦多少篇文章。

步驟 2.2.3: 配對 $Article_{HC}\{Max_{top-N}\}$ ，將瀏覽者與鄰近瀏覽者的共同偏好文章類別集合中，取出瀏覽者未曾閱讀過且具有較高閱覽率的文章配對給瀏覽者。

提出 CRL 方法的原因在於由於 BRL 方法需先找出目標瀏覽者的 N 項喜好程度較高的文章類別再與相近鄰居瀏覽者的文章偏好類別進行比對，透過此種方式可能會產生在相近鄰居瀏覽者的文章偏好類別中，難以完全找出與目標瀏覽者 N 項偏好的類別。目標瀏覽者的 N 項偏好類別中，有可能其中一項是鄰居瀏覽者沒有的，也有可能是二項，因此在部份偏好文章類別缺少的情況下，本研究加入本體論概念推論的方式來將目標瀏覽者可能產生偏好的文章加以推薦。而 CRL 方法則直接與鄰居群中 N 項最相似的文章偏好類別取出目標瀏覽者可能產生偏好的文章。

此外，為驗證個人化文章推薦之精確性及有效性，本研究另提出一項非以個人偏好為量化的非個人化文章推薦 NPRL(Non-personalized Recommendation List) 方法，其方法係推薦所有被瀏覽者共同點閱之點閱率最高的類別文章。

3.5 推薦效能評估

本實驗為能驗證其推薦方法的預測效能，首先會透過由瀏覽者所提供的推薦回饋分數，即 Likert 五等地評分法，再利用準確率(Precision)、回應率(Recall)及 F1 指標(F1-measure)評估方法來衡量本研究之 BRL、CRL 及 NPRL(Non-personalized Recommendation List)三種文章推薦方法的效能。其表示如公式(8)-(10)所示。

$$\text{Precision} = \frac{A_{\text{browser's scores}}}{A_{\text{total scores of all recommended}}} \quad (8)$$

$A_{\text{total scores of all recommended}}$ 代表所有推薦給瀏覽者 Top- N 文章的加總分數， $A_{\text{browser's scores}}$ 代表瀏覽者給予所有 Top- N 文章分數的總和，此公式用意在於計算出推薦給每位瀏覽者的推薦清單中滿意程度有多高。

$$\text{Recall} = \frac{A_{\text{accept}}}{A_{\text{numbers of all recommended}}} \quad (9)$$

$A_{\text{numbers of all recommended}}$ 代表推薦給瀏覽者 Top- N 的文章數量， A_{accept} 代表在 Top- N 的推薦清單中，瀏覽者給予評分大於或等於 3 分的文章以 1 的單位加總起來，若該篇文章僅有 1 或 2 分，該單位則設為 0。(Likert 五等地評分法)，此公式用意在於測量瀏覽者對於文章的接受度如何。

$$F1 = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (10)$$

F1 指標的目的在於減緩準確率與回應率兩者之間的不平衡，舉例來說，通常推薦的文章量增加，則會出現回應率值增加，而準確率值下降的問題。因此，於本研究中將結合準確率與回應率，以求得更精確且更公正的推薦效能。

4. 實驗結果

4.1 實驗資料與環境設定

在現今的部落格網站如無名小站 (<http://www.wretch.cc/>)、yam 天空部落格 (<http://blog.yam.com/>)、樂多日誌 (<http://blog.roodo.com/>) 等，大多缺乏了一個完整主題式的、較有意義的、具有保留價值的文章。為了能有效的將文章推薦給瀏覽者，因此本研究實作並設立一個「學術研究日誌」部落格入口網站 (<http://fblog.no-ip.org/>)，如圖 6 所示。

本網站鎖定單一主題且文章內容大都屬有意義、有價值的學術類、技術類文章。本研究網站內容包含 50 個有效的個人部落格，660 筆文章資料，67 位具有資訊科技專業領域背景的瀏覽者。學術研究日誌以資訊科技 (Information Technology) 為主要建構的領域。本網站主要可分成三大主題類別及 25 項子類別。學術文獻日誌入口網站畫面如圖 6 所示。部落格的實作

是採用 PHP 5 程式語言、My-SQL 4.1.8 資料庫及 Life Type 1.1 部落格建構模組來架設本研究實驗用網站，同時利用 Matlab 7.0 加上外掛的 ART-1 工具來執行瀏覽者個人偏好樹分群之工作，在前置文章分類處理方面則是採用 JAVA (JDK 6.0) 程式語言來負責文章分類，而本體論個人偏好樹則利用 protégé 3.2.1 來呈現出瀏覽者的個人偏好。

4.2 資料研析

本實驗網站主題類別之選用是採專家建議法，而參考之依據則取自 2007 資訊科技國際研討會 (International Conference on Advanced Information Technologies, AIT) 之徵文類別。由於該研討會類別主題係經由專家事先所定義好的，因此較適用於本研究所針對類別。類別與各主題範例如表 1。

在瀏覽者的文章偏好記錄方面，則是經由瀏覽者過去所閱覽過的文章中的 Web logs 內取出來，以方便之後在本體論模組建立 OPPT 時可對映 (Mapping) 至文章類別。瀏覽者的個人輪廓表範例如表 2。Web logs 中包含了瀏覽者所有閱覽過的文章網址，其網址所代表的意義如表 3 所示。

4.3 實驗流程與成果

依本實驗之架構與流程可分為三個模組程序，依順序分別為文章分類模組 (Article Classifier Module, ACM)、本體論建構模組 (Ontology Construction Module, OCM)、文章推薦模組 (Article Recommender Module, ARM)。以下分別就每個模組的實驗內容逐一介紹。

4.3.1 文章分類模組 (Article Classifier Module, ACM)

本模組實驗一開始需先將學術研究日誌網站內所有文章計有 660 篇來進行斷字 (Word Segmentation) 處理，這是要用來判斷該文章較屬於何種類別的主題以便於在下一階段來計算每個斷字的出現頻率。在斷字處理階段本研究採用中央研究院資訊科學所詞庫小組的 CKIP 中文斷詞系統。在斷字的前置處理上本研究只採用動詞集及名詞集及外文標記。圖 7 及圖 8 分別表示為全部詞類詞彙及本研究所需的動、名詞詞彙的結果範例。

圖 9 為 Top-10 最受歡迎的文章類別，此類資料乃是收集於所有瀏覽者於所有文章類別中總閱覽文章的次數。從下圖可看出閱覽率最高且

最受歡迎的文章類別依序為 Top-1 資訊隱藏、Top-2 資料探勘、Top-3 RFID、Top-4 電子商務、Top-5 網路行銷、Top-6 網路安全、Top-7 行動商務應用、Top-8 類神經網路、Top-9 無線感測網路及 Top-10 供應鏈管理。



圖 6 學術研究日誌入口網站畫面(<http://fblog.no-ip.org/>)

表 1 學術研究日誌類別主題

類別碼	類別名稱 (共 3 類)	子類別碼與子類別名稱 (共 25 類)			
A	電子化企業與知識管理	A ₁	資料庫管理	A ₆	軟體品質管理
		A ₂	企業資源規劃	A ₇	資料探勘
		A ₃	供應鏈管理	A ₈	知識管理
		A ₄	行動商務應用	A ₉	電子商務
		A ₅	顧客關係管理		
B	網路通訊技術與應用	B ₁	網路行銷	B ₆	網路管理
		B ₂	行動通訊及網路	B ₇	語音辨識
		B ₃	網路安全		
		B ₄	衛星通訊		
		B ₅	無線感測網路		
C	資訊系統理論與應用	C ₁	RFID	C ₆	電子教學平台
		C ₂	基因演算法	C ₇	物件導向技術
		C ₃	網路多媒體運用	C ₈	模糊邏輯理論與應用
		C ₄	類神經網路	C ₉	生物資訊應用
		C ₅	資訊隱藏		

表 2 瀏覽者瀏覽歷史記錄 (列舉)

Browser no.	Web logs
099503	http://163.17.9.89/index.php?op=ViewArticle&articleId=178&blogId=1 http://163.17.9.89/index.php?op=ViewArticle&articleId=18&blogId=1 http://163.17.9.89/index.php?op=ViewArticle&articleId=188&blogId=17 http://163.17.9.89/index.php?op=ViewArticle&articleId=20&blogId=1 http://163.17.9.89/index.php?op=ViewArticle&articleId=294&blogId=60 http://163.17.9.89/index.php?op=ViewArticle&articleId=311&blogId=24 http://163.17.9.89/index.php?op=ViewArticle&articleId=315&blogId=41 http://163.17.9.89/index.php?op=ViewArticle&articleId=442&blogId=17 http://163.17.9.89/index.php?op=ViewArticle&articleId=556&blogId=16 http://163.17.9.89/index.php?op=ViewArticle&articleId=559&blogId=16

表 3 瀏覽者歷史記錄描述

分段網址	分段網址定義
163.17.9.89	本實驗網站網址
ViewArticle&articleId=178	代表瀏覽者已閱覽過的文章識別號碼為第 178 篇
blogId=1	代表瀏覽者已閱覽過的部落格代號為第 1 個部落格



圖 7 斷詞後包含全部詞類標記結果 (列舉)

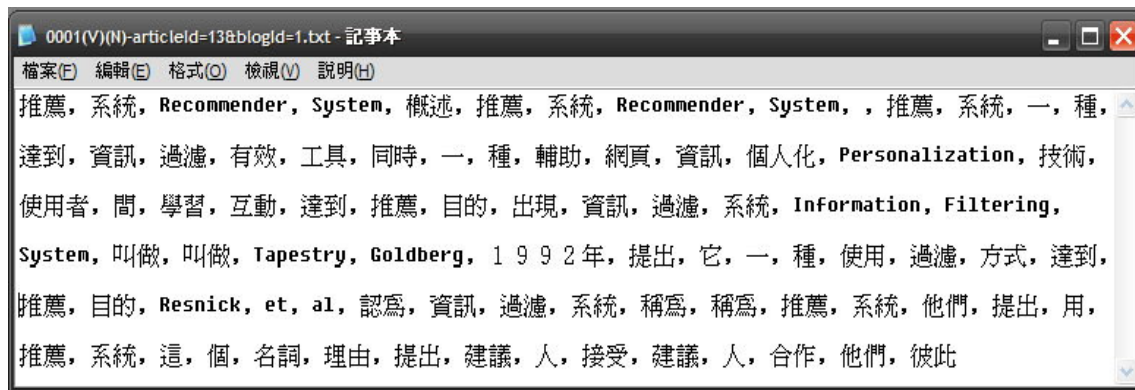


圖 8 斷詞後僅包含本研究所需詞類結果 (列舉)

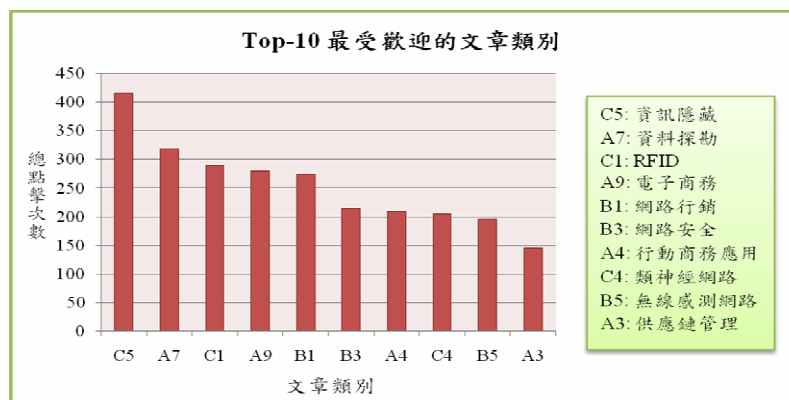


圖 9 Top-10 最受歡迎的文章類別

4.3.2 本體論建構模組(Ontology Construction Module, OCM)

在本體論建構模組部份主要負責將所有瀏覽者的個人輪廓(Profile)也就是他們過去文章閱覽的歷史記錄檔建構成樹狀形式的本體論個人偏好樹(Ontology Personal Preference Tree, OPPT)。建構 OPPT 首先需讀取瀏覽者的個人輪廓(User Profile, UP), 接著初始化每個瀏覽者的 OPPT, 然後判斷 OPPT 是否為空集合,

若此 OPPT 為空集合則在它的 root 上建立一個新的節點, 若非空集合, 可直接在原先的 OPPT 上加入新的節點即可。在加入新的節點至 OPPT 時需透過判斷式來注意是否加入了重覆的節點, 避免有重覆的節點加入, 圖 10 為本體論個人偏好樹(OPPT)列舉範例。

4.3.3 文章推薦模組(Article Recommendation Module, ARM)

本研究 ART 分群實驗是採用 Matlab ART1 來進行，而 ART 是採用 0 跟 1 兩種位元作為資料的輸入(Input)，因此本研究將瀏覽者於 OPPT 上的資料依閱讀後賦予屬性值，1 的位元代表該瀏覽者有閱讀過此文章類別，而 0 的位元則代表該瀏覽者沒有閱讀過該文章類別。

圖 11 為分群實驗後部份的矩陣編碼圖。圖中最左側的編號為瀏覽者編號(例如，019307)，最上層的編號為文章子類別(例如，A7、A8、A9、B1、B2)，在文章子類別的底部則為文章的閱讀次數(例如，2~4、5~9、10 以上)，接著透過瀏覽者與文章類別閱讀次數的矩陣計算，即可表示為 ART 輸入的矩陣編碼。

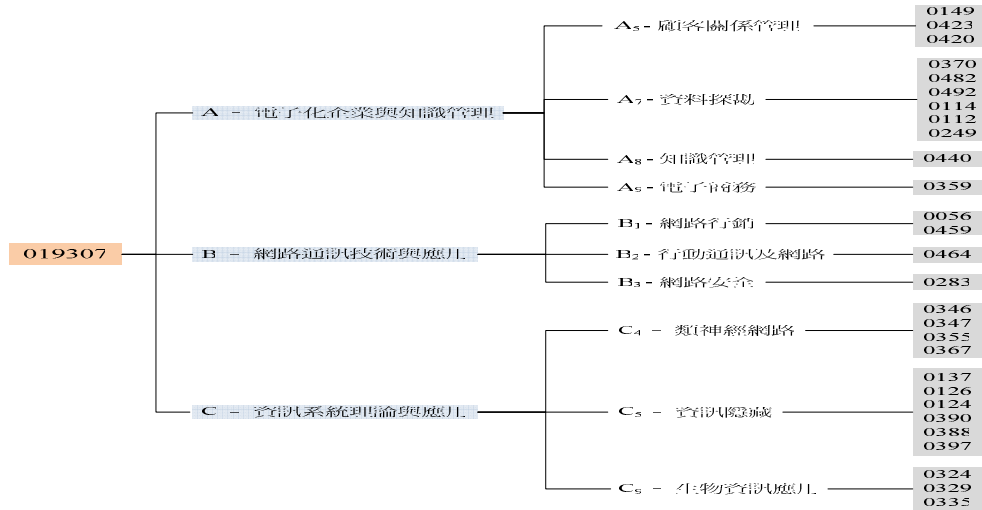


圖 10 本體論個人偏好樹(OPPT) (列舉範例)

	A ₇			A ₈			A ₉			B ₁			B ₂		
	2~4	5~9	10↑	2~4	5~9	10↑	2~4	5~9	10↑	2~4	5~9	10↑	2~4	5~9	10↑
019307	1	1	0	0	0	0	0	0	0	1	0	0	0	0	0
029332	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
039403	0	0	0	0	0	0	0	0	0	1	1	1	1	1	0
049406	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0
059413	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0
069420	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0
079441	1	1	1	0	0	0	0	0	0	1	1	0	0	0	0
089445	0	0	0	0	0	0	0	0	0	1	1	1	0	0	0
099503	1	1	1	0	0	0	0	0	0	1	1	0	0	0	0
109504	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0
119511	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0
129514	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
139515	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0
149518	1	1	0	0	0	0	0	1	1	0	1	0	0	0	0
159501	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
169509	1	1	1	0	0	0	0	1	1	0	0	0	0	0	0
179519	1	0	0	1	0	0	0	1	0	0	0	0	1	0	0
189520	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

圖 11 ART 矩陣編碼 (列舉範例)

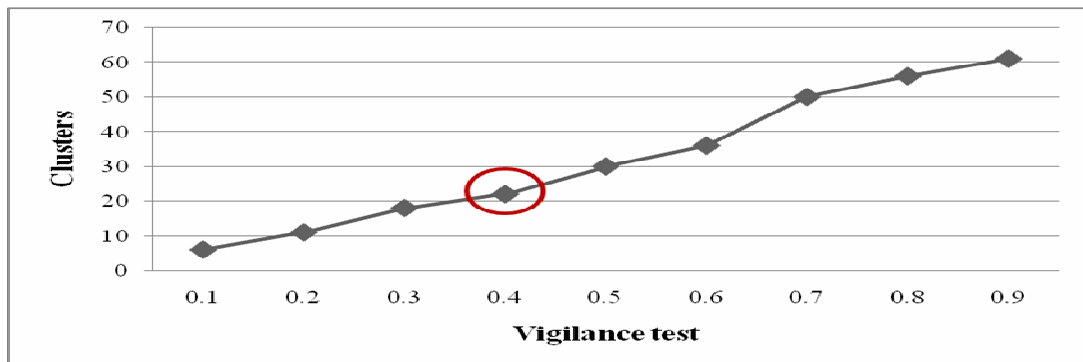


圖 12 警戒值與分群數測試

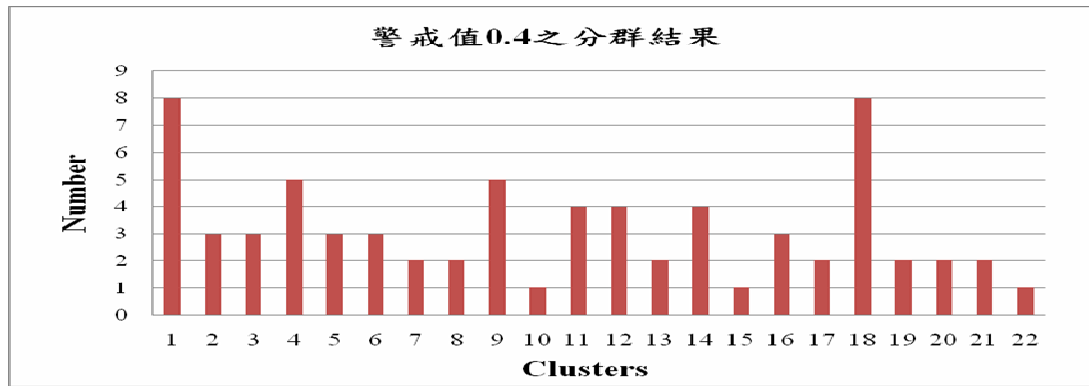


圖 13 警戒值 0.4 之分群結果

ART 編碼完成後緊接著可帶入至 Matlab 程式中進行分群動作，由於 ART 需要經由調整警戒值 ρ 方能決定出群聚的數目，因此本研究設定以 0.1 為一個區間來找出適當的警戒值。圖 12 為警戒值測試後的結果，在實驗中將取警戒值 0.4 作為本研究的警戒值，其原因在於經比對過後發現每一群均能平均分配出偏好相似的瀏覽者，這也代表者瀏覽者間具有一定相似程度的偏好，且落於 0.4 的警戒值中有三群的分群結果都只有單獨一個瀏覽者屬於單一群中，這就代表了這些的瀏覽者的閱讀偏好較屬奇特，與其他瀏覽者的偏好較為不同。

圖 13 可看出同一群中具有兩個以上瀏覽者的群組會被分至協同過濾法，依鄰居群概念，找出與目標瀏覽者最相似的鄰居來進行文章推薦；若該群組只有一個瀏覽者的話則被分至內容導向法進行推薦。因此，第 1、2、3、4、5、6、7、8、9、11、12、13、14、16、17、18、19、20、21 群的瀏覽者採用協同過濾法推薦；第 10、15、22 的瀏覽者則採用內容導向法推薦。

4.4 推薦效能評估

在本實驗最後部份，為有效評估其推薦效能為何，採用了推薦系統較廣泛使用的評估指標，分別為準確率(Precision)、回應率(Recall)及 F1 指標(F1-measure)來評估本研究提出的 BRL 及 CRL，於本實驗中，準確率(Precision)被定義為依據 Likert 五等地分法將瀏覽者對所有推薦文章的分數(1~5 分)加總起來，除上 Top-N 的總分，即可得知文章推薦的準確率；回應率(Recall)也同樣依據 Likert 五等地分法，從推薦的文章中，將被瀏覽者所給予 3~5 分的文章以 1 的單位加總起來，除上 Top-N 總數的單位(例如，若 Top-N 為 10，則總數的單位為

10)；最後為了讓本研究的推薦效能評估更具有公正性與有效性，則採用 F1 指標(F1-measure)做為本研究的推薦效能衡量指標。

此外，針對本研究提出三種的推薦方法並加以比較其效能，分別是非個人化的 NPRL(Non-personalized Recommendation List)及以同群偏好為主的 CRL(Cluster-based Recommendation List)及以個人偏好為主的 BRL(Browser-based Recommendation List)。NPRL 方法係依據全部文章類別中，找出前幾項 Top-N 閱覽率最高且最受歡迎的類別並予以推薦。圖 14 實驗結果中可得知，在非個人化情況下，系統無法了解目標瀏覽者實際偏好，更無法根據瀏覽者偏好去挖掘出其潛在的偏好。因此，針對 NPRL 的實驗結果，準確率、回應率及 F1 指標，分別可得到 64.06%、75% 及 69% 的文章推薦命中率。

CRL 係先與相似鄰居群中找出與他們最相似的 Top-N 偏好文章類別，然後再從中篩選出目標瀏覽者未曾閱覽過且點閱率最高的文章來進行推薦。圖 15 為 CRL 分別以三種評估指標衡量的實驗結果。從圖 14 所得的實驗結果，準確率為 72.51%、回應率為 90% 及 F1 指標為 80%。

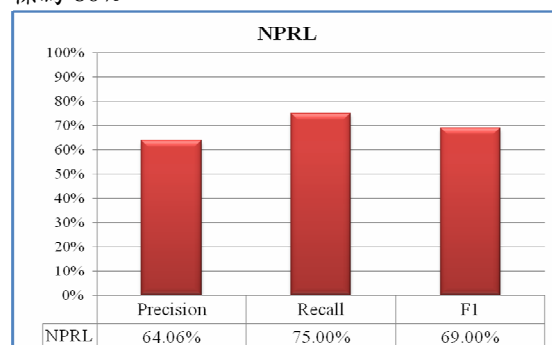


圖 14 NPRL 三種效能評估實驗結果

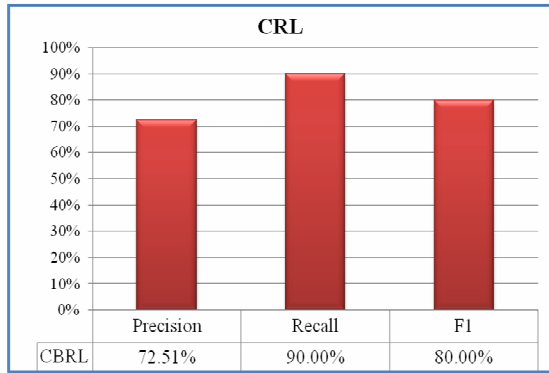


圖 15 CRL 三種效能評估實驗結果

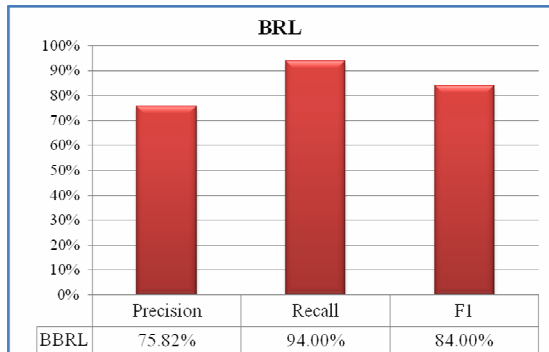


圖 16 BRL 三種效能評估實驗結果

由本實驗中，CRL 所得數值均略低於 BRL 實驗結果(請參照圖 16)，原因在於 CRL 係先考量同群間偏好來產生出可能的潛在偏好，並非先以瀏覽者個人偏好為主。因此，在 CRL 的實驗中，可能某些推薦出來的潛在性偏好文章並非目標瀏覽者所偏好的，造成了推薦命中率的低落。然而，就該實驗的 F1 指標也達到 80% 的文章推薦命中率，僅低於 BRL 四個百分點，由此可知，藉由 CRL 潛在偏好的推論於推薦系統的領域上可具有參考的價值。

BRL 與 CRL 恰好相反，BRL 依據目標瀏覽者 Top-N 的偏好文章類別，找出前幾項最具偏好的文章類別，再從相似鄰居群中找尋其他瀏覽者與目標瀏覽者相似偏好的文章類別，之後從其他瀏覽者相似偏好的文章類別中篩選出目標瀏覽者未曾閱覽且點閱率最高的文章，將這些類別的文章以清單方式推薦給目標瀏覽者。

簡而言之，BRL 先以個人偏好為基礎，再參考群組內共同偏好的文章類別；而 CRL 則是先以群組內共同偏好的文章類別為基礎，再從個人偏好中找出瀏覽者的偏好。針對 BRL 依準確率、回應率及 F1 指標三種評估方法，如圖 16 實驗結果，準確率為 75.82%、回應率為 94% 及 F1 指標為 84% 的文章推薦命中率。

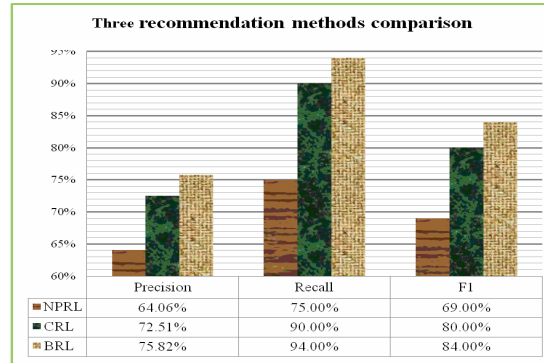


圖 17 三種推薦方法比較圖

根據上述所提之三種推薦方法實驗結果，為更清楚了解 BRL、CRL 與 NPRL 在三種效能評估方法之差異，本研究綜合整理如圖 17 所示。

圖 18-20 呈現出 Top-1 到 Top-10 三種推薦方法的比較圖，由於 Top-N 是依據瀏覽者的偏好來給予一個順序性的推薦，因此最在前面的序號(即 Top-1)通常代表著瀏覽者對於某文章的偏好程度最高。此外，本研究於 Top-N 推薦時係依據事先篩選出瀏覽者所偏好的文章類別，再從各類別中依偏好程度擷取出瀏覽者偏好的文章作為 Top-N 推薦，因此在實驗結果中可能產生 Top-1 至 Top-10 非直線性的遞減結果，而會出現波浪狀、不具太大明顯性的起伏，但仍可輕易看出於 Top-1 至 Top-10 是具有遞減、下降的結果，換句話說，越接近 Top-1 的文章，瀏覽者接受度越高，給予的回饋分數也越高。

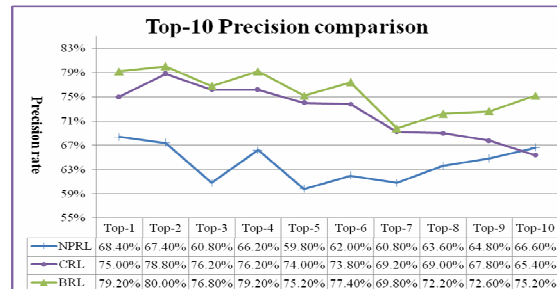


圖 18 Top-10 Precision 比較圖

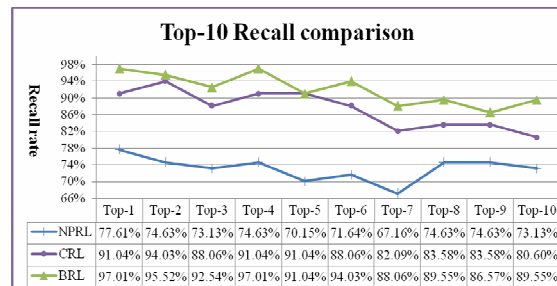


圖 19 Top-10 Recall 比較圖

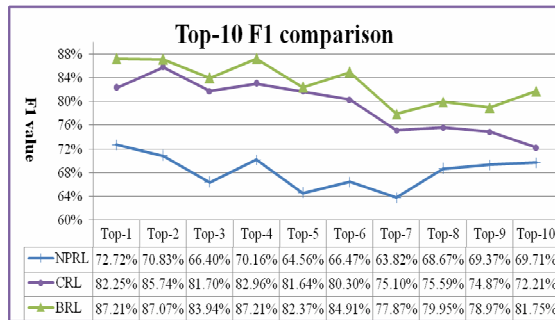


圖 20 Top-10 F1 比較圖

5. 結論與未來研究

本研究提出了個人化的部落格文章推薦系統(BARS)，利用過去瀏覽者於部落格網站中所留下的歷史記錄產生個人偏好樹，並利用類神經網路中的自適應共振理論網路(ART)將瀏覽者進行分群的動作，藉此分群的動作可從瀏覽者的鄰居挖掘出可能的潛在偏好，之後並依此潛在偏好來給予推薦；此外，本研究另提出一項以本體論概念為基礎的內容導向式推薦，來幫助與他人興趣均不同亦或者是系統的新進瀏覽者缺乏足夠的瀏覽資訊而無法造成推薦的問題(即冷啟始問題)。

在推薦系統中，冷啟始為普遍常見之問題。冷啟始即代表系統缺乏瀏覽者相關資訊，因此較難去找出較高相似的鄰近瀏覽者來做出推薦。本研究提出一個以本體論概念的內容導向式推薦方法，來推論出這類瀏覽者的潛在偏好，藉以降低冷啟始問題帶來推薦成效不彰的問題。

本研究提供一個部落格平台，並收集瀏覽者在網站上相關歷史記錄，依使用者偏好來提供個人化的文章推薦，幫助瀏覽者能夠從大量的文章中，從過多繁雜的資訊推為更合宜有效的文章，藉此來增加瀏覽者對網站的向心力及忠誠度。

由本研究的方法與實驗後可解決了下列問題：(1) 降低冷啟始問題，(2) 提昇個人化文章推薦，(3) 應用於部落格之先導研究。本研究的貢獻為提出二種推薦清單產生的方法，分別為 BRL 與 CRL，BRL 係先以瀏覽者的個人偏好為基底，再以群組內的共同偏好來找出潛在偏好，此方法經實驗後，其推薦命中率(F1 指標)可達到 84%；另一個方法 CRL 則是與 BRL 相反，它係以群組的共同偏好為出發點，再去尋找其個人偏好，其產生的推薦準確率為 80%。

在未來研究方面，本研究所提出的方法可

加強方向如下：

- (1) 多主題式(Multi-topic)部落格網站推薦：現今部落格趨勢已漸漸走向多主題式的網站，由於本研究僅採用資訊科技類的主題，因此較難以實現多主題式部落格網站的推薦。相信在未來上若加入多主題式部落格網站推薦結合本研究的部落格文章推薦，將可使部落格推薦系統更加的完善。
- (2) 行動式(Mobile)部落格文章推薦：目前行動式部落格大多流行於影音的即時分享，較少於進行部落格文章的分享，而本研究所提之部落格文章推薦也僅侷限於一般的 Web。因此，若利用本研究的個人化推薦方法於行動推薦上，對手機用戶而言，將會是一個新穎且便利性的應用。

誌謝

本研究感謝行政院國家科學委員會專案研計劃 (編號：NSC-95-2416-H-324-011、NSC-96-2416-H-324-006)之補助

參考文獻

- [1] 中央研究院資訊科學研究所詞庫小組 (CKIP)(2007)<http://ckipsvr.iis.sinica.edu.tw/>。
- [2] 李欣岳，Blog Inc. 趨勢探索，*天下雜誌*，2005 年 7 月號，55 期，82~89 頁，2005。
- [3] 張慧玲，應用本體論網路分析於資訊推薦之研究，*碩士論文*，天主教輔仁大學資訊管理學系，臺北，2005。
- [4] 曾元顯，“關鍵詞自動擷取技術與相關詞回饋”，*中國圖書館學會會報*，59 期，59~64 頁，台北，1997。
- [5] 葉怡成，*類神經網路模式應用與實作*，第八版，儒林圖書有限公司，2003。
- [6] 歐仁德(2006)，結合本體論與通用個人輪廓於個人化推薦之研究，*碩士論文*，朝陽科技大學資訊管理學系，臺中。
- [7] Amazon, Available at: <https://www.amazon.com/>, 2007.
- [8] Blogger, Available at: <https://www.blogger.com/>, 2007.
- [9] G. A. Carpenter, and S. Grossberg(1987), "ART2: Self-organization of Stable Category Recognition Codes for Analog Input Patterns," *Applied Optics*, pp. 4919-4930.
- [10] F. Chesani(2002), "Recommendation Systems," *Corso di laurea in Ingegneria*

- Informatica*, pp. 1-32.
- [11] Y. B. Cho, Y. H. Cho, and S. H. Kim(2005), "Mining Changes in Customer Buying Behavior for Collaborative Recommendations," *Expert Systems with Applications*, Vol. 28, pp. 359-369.
 - [12] R. C. Chen, J. Y. Liang, and R. H. Pan(2008), "Using Recursive ART Network to Construction Domain Ontology Based on Term Frequency and Inverse Document Frequency," *Expert Systems with Applications*, Vol. 34, pp. 488-501.
 - [13] T. R. Grouber(1993), "A Translation Approach to Portable Ontology Specifications," *Knowledge Acquisition*, Vol. 5, pp. 199-220.
 - [14] S. Grossberg(1976), "Adaptive Pattern Classification and Universal Recoding: I. Parallel Development and Coding of Neural Feature Detectors," *Biological Cybernetics*, Vol. 23, pp. 121-134.
 - [15] GroupLens(2007), Retrieve at: <http://www.grouplens.org/index.html>.
 - [16] N. Guarino, C. Masolo, and G. Vetere(1999), "OntoSeek : Content-Based Access to the Web," *IEEE Intelligent Systems*, Vol. 14, pp. 70-80.
 - [17] P. L. Hsu, P. C. Liu, and Y. S. Chen(2006), "Using Ontology to Map Categories in Blog," *Proceedings of the International Workshop on Integrating AI and Data Mining(AIDM'06)*, pp. 65-72.
 - [18] H. J. Lee and S. J. Park(2007), "MONERS: A News Recommender for The Mobile Web," *Expert Systems with Applications*, Vol. 32, Issue 1, pp. 143-150.
 - [19] Y. Li, L. Lu and L. Xuefeng(2005), "A Hybrid Collaborative Filtering Method for Multiple-interests and Multiple-content Recommendation in E-Commerce," *Expert Systems with Applications*, Vol. 28, Issue 1, pp. 67-77.
 - [20] T. P. Liang, Y. F. Yang, D. N. Chen, and Y. C. Ku(2007), "A Semantic-expansion Approach to Personalized Knowledge Recommendation," *Decision Support Systems*.
 - [21] B. Mobasher, R. Burke, R. Bhaumik, and J. J. Sandvig(2007), "Attacks and Remedies in Collaborative Recommendation," *Intelligent Systems*, Vol. 22, pp. 56-63.
 - [22] S. E. Middleton, N. R. Shadbolt, and D. C. De Roure(2004), "Ontological User Profiling in Recommender System," *ACM Transactions on Information Systems (TOIS)*, Vol. 22, No. 1, pp. 54-88.
 - [23] MySpace(2007), Retrieve at: <http://blog.myspace.com/>.
 - [24] S. Puntheeranurak and H. Tsuji(2007), "A Multi-clustering Hybrid Recommender System," *Computer and Information Technology, 2007. CIT 2007. 7th IEEE International Conference*, pp. 223-228.
 - [25] A. M. Rashid, I. Albert, D. Cosley, S. k. Lam, S. M. McNee, J. A. Konstan, and J. Riedl(2002), "Getting to Know you: Learning New User Preferences in Recommender Systems," *Proceedings of the 7th International Conference on Intelligent User Interfaces*, pp. 127-134.
 - [26] G. Salton and C. Buckley(1988), "Term-weighting approaches in automatic text retrieval," *Information Processing & Management*, Vol.24, No.5, pp. 513-523.
 - [27] B. Sarwar, Karypis, G. Konstan, and J. Riedl(2000), "Analysis of Recommendation Algorithms for E-commerce," *Proceedings of ACM E-commerce 2000 Conference*, pp. 158-167.
 - [28] B. Sarwar(2001), "Sparsity, scalability, and distribution in recommender systems," *Ph.D. Dissertation*, University of Minnesota.
 - [29] Technorati(2007), Retrieve at: <http://www.technorati.com/>.
 - [30] S. Upendra and M. Pattie(1995), "Social Information Filtering: Algorithms for Automating "Work of Mouth"," *Proceedings of the SIGCHI conference on Human factors in computing systems*, pp. 210-217.
 - [31] S. S. Weng and H. L. Chang(2008), "Using Ontology Network Analysis for Research Document Recommendation," *Expert Systems with Applications*, Vol. 35, Issue 3.
 - [32] Y. Zhang and J. Jiao(2007), "An Associative Classification-based Recommendation System for Personalization in B2C E-commerce Applications," *Expert System with Applications*, Vol. 33, pp. 357-367.