

FAIR：應用於部落格文章之資訊攫取模式

陳隆昇
朝陽科技大學資訊管理系
助理教授
lschen@cyut.edu.tw

林姿呈
朝陽科技大學資訊管理系
研究生
s9614638@cyut.edu.tw

摘要

愈來愈多的企業希望能從大量的部落格文章中擷取出有用的資訊，從中了解消費者需求及市場導向，以幫助企業改善商品及服務品質，亦或評估企業本身或同業競爭者的優劣勢。因此，本研究針對部落格中的產品使用心得，提出一套 FAIR 模組，希望藉由該模組得以達到在短時間內有效地分析產品評價，以利於企業或消費者在掌握商品重點特色及整體評價時。FAIR 模組為模糊自適應共振理論 (Fuzzy ART) 類神經網路結合隱含語意索引 (LSI) 的特性，將文章集予以分群並從中擷取出代表性關鍵詞，以達到資訊檢索的目的，最終再透過關聯法則 (AR) 提升關鍵詞彙的解釋性。最後，我們以美容保養品之部落格文章作為實驗對象，以驗證所提出的 FAIR 模組之有效性。

關鍵詞：部落格、資訊檢索、模糊自適應共振理論類神經網路、隱含語意索引、關聯法則

Abstract

A lot of companies attempt to discover useful knowledge from that huge amount of bloggers' comments for business purposes. Therefore, this research aims to propose a Fuzzy Adaptive Resonance Theory (Fuzzy ART) network based Information Retrieval (FAIR) scheme by integrating Fuzzy ART network, Latent Semantic Indexing (LSI), and Association Rules (AR) discovery to extract knowledge from blog. In FAIR, Fuzzy ART network first has been employed to segment bloggers. For each customer segment, we use LSI technique to retrieve important keywords. Then, in order to make the extracted keywords to be understandable, AR is presented to organize these to form concepts. Finally, a real case of cosmetics products evaluation has been provided to demonstrate the effectiveness of the proposed FAIR scheme

Keywords: Blogs, Information Retrieval, Fuzzy

Adaptive Resonance Theory Neural Network, Latent Semantic Indexing, Association Rule.

1. 前言

部落格 (Blogs) 已成為時下流行的資訊交流媒介之一，其普遍性可根據 Technorati 於 2008 年的統計數據，全球網路的部落格讀者已達 1 億 8890 萬人口之多，其成長的速度由 2002 年開始，平均每天至少有 900,000 筆新文章被發佈。使用者可藉由部落格撰寫文章，為自己的生活做記錄，亦可分享本身經驗及對事物的評價和觀感 [3]。

相關文獻指出，人們透過部落格傳播個人的意見及想法，可能將導致互相影響的效果，甚至因而產生被洗腦的現象 [4]，這意味著部落格儼然已成為日常生活中，不可或缺的溝通管道之一。近年來，許多企業已掌握住部落格快速傳播訊息的特性，適時的將其應用於企業內外部或服務顧客，甚至藉由部落格導正企業形象，避免及減少負面評價的產生。許多廠商亦開始將部落格視為一種經營產品的管道，提供部落客免費試用及體驗商品，藉由部落格的高度感染力，進而提高產品的曝光程度、可信度、亦或對該商品作市場調查 [15]。故公司企業如果能成功地運用部落格，依賴其快速且容易使人信服的口耳相傳特性，必定能達到有效的行銷效果 [16]。以 Nissan 為例，該公司於 2005 年為新車 Tiida 成立了專屬部落格，提供車主們分享駕駛心得、經驗、以及影音照片的空間，使車主以及潛在消費者得以產生交流和互動，而部落格中所討論到的問題及評價，也幫助該公司作為參考，從中發掘對企業有用的資訊 [17]。

過去也曾有諸位學者對部落格進行網路探勘的研究，Todoroki *et al.* 建立一套以部落格為主的電子化系統，讓使用者可以透過簡易的操作平台，對部落格加以管理、編輯及搜尋 [14]；Chau and Xu 以半自動化的機制，針對部落格當中的特殊族群（如種族主義者）進行分析及監控，避免危害社會的可能性 [4]；Lin and Huang 從分析個案中發現，部落格擁有強大的

影響力，觀光業可藉由部落格之特性，開創市場商機[11]；相關應用之文獻皆以 web usage mining[9]及 web structure minin[4]為主要導向，而近年來有關於部落格之 web content mining 的應用，則主要透過資料之蒐集，以分析當前部落格主要的議題及討論趨勢[6]。其中以 Latent Semantic Analysis (LSA) 最常被應用於部落格 content mining 之中。此外，針對部落格搜尋引擎的部份，亦常使用 Probabilistic Latent Semantic Analysis (PLSA) 於追蹤最熱門之部落格標題及關鍵字。然而，透過此類資訊檢索技術所擷取出來的關鍵字，並不易於解讀其代表之含意，故本研究提出一套 FAIR(Fuzzy Adaptive Resonance Theory based Information Retrieval scheme)模組，結合自適應模糊理論 (Fuzzy ART) 類神經網路[7]及隱含語意索引 (LSI) [8]達到資訊檢索之功效，並透過關聯法則 (Association Rules) [1,2]技術，以提昇其關鍵字之可讀性。最後，我們以美容保養品之部落格文章作為實驗對象，以驗證所提出的 FAIR 模組之有效性。

2. FAIR 模組

本節將針對所提出之 FAIR 模組進行詳細的方法及流程介紹，將部落格之產品使用心得分享做內容探勘，利用分群技術將文章集劃分為若干群組，並透過探勘技術擷取出文章的代表性詞彙，以達到短時間內即可掌握相關評價資訊。FAIR 之研究流程如圖 1 所示，可分為五個步驟：

- (1) 蒐集部落格文章
- (2) 資料之前置處理
- (3) 文章分群
- (4) 語意檢索
- (5) 提升解釋性

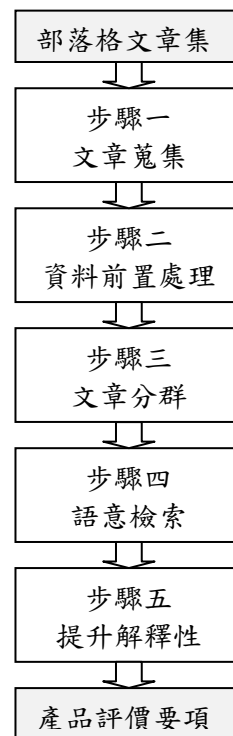


圖 1 FAIR 模組之流程

各步驟之詳細介紹請參考 2.1-2.5 節之內容。

2.1 部落格之文章蒐集

由於使用者從部落格參考特定內容之文章，勢必耗費大量的時間瀏覽閱讀，而文章內容及評價皆不一，使得獲得資訊的過程變得冗長、無效率，進而使得參考的結果較為模糊化，讓部落客在擷取產品資訊而言，成為一門耗時且耗力的工作。故本研究針對部落格文章進行內容探勘，故首要步驟為蒐集部落格文章，而文章的篩選條件則以美容保養品為例，希望藉由資料探勘的過程，擷取出該產品的重點資訊，以具代表性之關鍵字表達該產品的優缺點，亦或備受關切的要項及特點。

2.2 資料之前置處理

在進行資料探勘之前，必須對資料做前置處理的動作，為避免太多雜訊的干擾，進而影響探勘的結果及效率，故於文章蒐集後，逐篇針對顏文字及注音文進行修改或刪除的動作。再者，我們利用中央研究院的中文斷字系統進行斷詞，該系統依據其定義的詞典，進而將特定詞彙切割為一個獨立字詞，斷詞後的文章，依名詞及動詞作為篩選條件，將符合其詞性之詞彙擷取出，以作為文章之關鍵字。

詞彙－文件向量矩陣常被使用於資訊檢

索中，由於它能表達詞彙與文件間的關聯性，向量的結果亦方便進行群聚及分類的處理，故本研究根據斷詞後的篩選結果，進行建立詞彙—文件向量矩陣的步驟，以欄作為關鍵字的向量表示，以列作為文章編號的向量表示，依據每篇文章的關鍵字集合，在該文章編號所對應的關鍵字向量當中，該關鍵字未出現於文章中則填入 0 值，反之，該關鍵字出現於文章中則填入 1 值。其流程大致上如圖 2 所示：

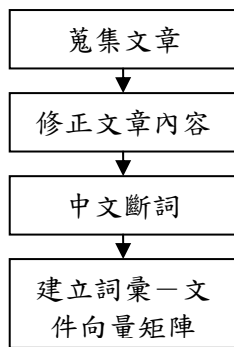


圖 2 資料前置處理之流程

2.3 文章分群

根據步驟二所建立的詞彙—文件向量矩陣，我們將其作為文章分群的依據，藉由模糊自適應共振理論 (Fuzzy ART) 類神經網路將文章區分為若干個集合，將文章內容近似的文章視為同一個群組，希望從分群的結果中，觀察顧客對該產品的心得主要可劃分為哪些類別要項。

Fuzzy ART 類神經網路於 1991 年由 Carpenter et al. 所提出，屬於非監督式學習的網路模式，其結合模糊集合理論 (Fuzzy Set Theory) 與自適應共振理論 (Adaptive Resonance Theory, ART)，以改善過去 ART 所面臨的輸入格式之限制。相較於 Fuzzy ART，ART1 僅接受二元值作為輸入資料的格式，不支援介於 0 至 1 之間的連續值；而整體執行架構方面，Fuzzy ART 亦比 ART2 簡單得多，不須要經過資料正規化的程序，即可達到快速並穩定的收斂分群[12]。此外，Fuzzy ART 不需要事先定義分群數，故可避免人為因素的干擾，進而影響分群結果的準確性。

Fuzzy ART 的運作方式與人類的記憶系統相似[18]，在學習新物件特性的同時 (具可塑性)，又必須保留舊物件特性的記錄 (穩定性)，但此兩項特性存在著衝突關係，而解決此項問題的方法即採用「共振測試」或稱「警

戒值測試」(Vigilance test)。共振測試即透過警戒值判斷資料相似與否[5, 13]。當向量資料輸入時，首先必須設定其相關參數： ρ (Vigilance parameter)，定義群集相似度的標準； α (Choice parameter)，在快速學習模式中，通常建議趨近於 0。若 $\alpha=1$ ，則表示無條件通過共振測試； β (Learning rate)，更新權重時所使用，其大小影響學習速度的快慢；

再者，透過計算選擇函數(1)，確定元素執行共振測試時之順序，以最大值為優先(2)，其計算方式如下列公式所示[8]：

$$T_j(I) = \frac{|I \wedge w_j|}{\alpha + |w_j|} \quad (1)$$

$$T_j = \text{MAX} \{T_j : j = 1..n\} \quad (2)$$

其中， W_j 為輸入向量資料 I 之權重； \wedge 的定義為 $(X \wedge Y) = \text{MIN}(X_i, Y_i)$ ； $\|$ 之定義為 $|X| = \sum_{i=1}^m x_i$ 。確定優先順序後，依公式(3)計算該值與群組之間的相似度，若大於等於 ρ 值，則表示通過共振測試，其條件符合該群組之特色，可歸納為同一群組。

$$\frac{|I \wedge W_j|}{|I|} \geq \rho \quad (3)$$

2.4 資訊攫取

文章經由分群後，必須擷取出每個群組所表達的主旨，因此本研究利用隱含語意索引 (Latent Semantic Index, LSI) 進行文章內容分析，而經由奇異值分解 (Singular value decomposition, SVD) 後的結果，將可擷取出每篇文章最具代表性的關鍵字，每個群組根據所屬的關鍵字集合，便可輕易地觀察出哪些文章關切著什麼重點，例如若干人認為該產品售價太貴、若干人認為該產品的味道很刺鼻等。

學者 Deerwester et al. 於 1990 年提出了隱含語意索引 (Latent Semantic Index, LSI) 的概念，為資訊檢索領域中常被予以應用的一項技術，用以發掘文件之隱含語意。隱含語意索引利用向量空間的概念，將文件及索引詞以投影至向量空間的方式呈現，可表示出索引詞與索引詞、索引詞與文件、及文件與文件之間的關聯，其主要目的為有效縮減資料維度，減少資訊檢索所必需耗費的時間，並可解決一字多義 (polysemy) 及多字一義 (synonymy) 之問題。

隱含語意索引主要仰賴奇異值分解 (Singular value decomposition, SVD) 的執行，其目的為將資料維度縮減 100 至 300 大小的維度空間[19]。假設 A 為 m x n 之矩陣，其 rank(A)=r 並滿足 m>n 之條件，對 A 做奇異值分解則可表示為 A=USV^T，其中 S=diag(σ₁,...,σ_n)，σ_i>0 for 1<=i<=r，σ_j=0 for σ_j>r+1，且 U^TU=V^TV=I_n。U 為詞彙之向量集合，其行向量稱之為左奇異向量 (Left singular value)。V 為文件之向量集合，其行向量稱之為右奇異向量 (Right singular value)，S 為對角矩陣，其示意圖如圖 3 所示[20]。

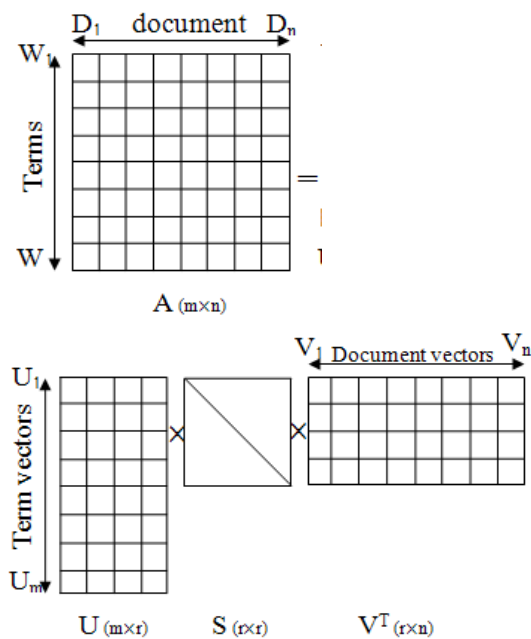


圖 3 奇異值分解之示意圖

2.5 提升解釋性

由上個步驟可得知每個群組的代表性關鍵字，然而，關鍵字的可讀性不甚理想，故本研究透過關聯法則，將詞彙做組合的動作，使得結果更具邏輯且精確，而使用者也能更輕易地解讀其最終結果。

關聯法則 (Association rules) 係由學者 Agrawal *et al.* 於 1993 年提出，其主要目的為找出項目之間所隱含的關聯性[1]。關聯法則應用於交易系統中，可輕易地從大量的顧客交易資料記錄裡，依顧客的消費習性及購買商品清單，發掘出人為不易發現地隱藏資訊。譬如購買了某產品的同時，通常也會伴隨另一項產品的購買慾，此類資訊即可提供決策者擬定有效的產品組合或促銷活動，並有助於賣場動線規

劃及控管進貨量的安排。

假設 I 為所有項目的集合，而 X 和 Y 各代表某些項目的集合，並且滿足 X 及 Y 皆包含於 I、X∩Y 不為空集合之條件，如欲觀察消費者購買 X 與 Y 項目的關聯性，則可透過關聯法則之表示式 X→Y[support%, confidence%]得知，而該法則成立與否則是依據事先定義的支持度 (support) 與信心度 (confidence) 之門檻值來判斷，支持度與信心度必須等於或大於門檻值，該法則才予以有效性及參考價值。支持度與信心度為關聯法則的主要評量指標，支持度可顯示特定的項目集在所有項目集中所佔有的比率，其值愈高即表示該項目集出現的次數愈頻繁，計算方法如公式(4)所示；信心度則可顯示觸發 X 項目的情況下，Y 項目亦伴隨發生的條件機率，計算方法如公式(5)所示。

$$Support(X \rightarrow Y) = \frac{X \cap Y}{I} \quad (4)$$

$$Confidence(X \rightarrow Y) = \frac{Support(X \cap Y)}{Support(X)} \quad (5)$$

3. 實驗

3.1 部落格之文章蒐集

本研究以美容保養產品—「雪肌精化妝水」作為實驗範例，蒐集該產品之相關部落格文章，以 FAIR 模組實證研究目的之效益。資料來源以「無名小站部落格」以及「UrCosm 化妝品使用心得分享網站」為主，以雪肌精化妝水為搜尋主題，蒐集該產品之使用心得文章，文章筆數共三百則，以下為文章內容之範例：

最近愛上雪肌精，我還滿愛它那個味道的~香香的，靠櫃當天用完也沒有紅 or 其他事情發生，那天櫃姊也有幫我上乳液，本來想說會很油，但是我個人覺得還 ok 耶，而且也是香香的~，真的是買對了(感動~~~)，我很 enjoy 那個味道~呵，都是用化妝棉沾濕上全臉，大家說用完會很乾，不過我不會耶~，不黏不油，good~!!!

資料來源及文章範例：

- http://www.urcosme.com/index.htm
- http://www.wretch.cc/blog
- http://www.wretch.cc/blog/sophia5646&article_id=26457973

3.2 資料之前置處理

文章內容經由適當的修正後，即進行中文斷詞之處理，其結果範例如圖 4 所示。依斷詞後之詞性，擷取出符合名詞及動詞之詞彙作為文章關鍵字之代表。三百則文章之斷詞結果共產生了 736 筆非重覆詞彙，經由詞性篩選關鍵字之步驟後，詞彙數則降至為 466 筆。最後依關鍵字及文件之分佈關係，建立詞彙-文件向量矩陣，其大小為 300 x 466，該矩陣之部份範例結果如表 1 所示。

最近(Nd) 愛上(VJ) 雪肌精(Na) , (COMMATEGORY)

 我(Nh) 還(D) 滿(Dfa) 愛(VL) 它(Nh) 那(Nep) 惟

 靠(VJ) 櫃(Na) 當天(Nd) 用完(VC) 也(D) 沒有(VJ)

 那(Nep) 天(Nf) 櫃姊(Na) 也(D) 有(V_2) 幫(P) 理

 本來(D) 想(VE) 說(VE) 會(D) 很(Dfa) 油(VH) ,

 但是(Cbb) 我(Nh) 個人(Na) 覺得(VK) 還(D) ok(FI)

圖 4 中文斷詞之範例

表 1 詞彙-文件向量矩陣之範例結果

	美白	透亮	嗆鼻	酒精	效用	...
Doc1	1	1	0	1	1	...
Doc2	0	1	1	0	0	...
Doc3	0	1	0	1	0	...
...

3.3 文章分群之結果

執行 Fuzzy ART 之前之相關參數設定如下： α 設置為 0.0001、 β 設置為 1。而由於無法事先定義最終之分群數，故採取實驗不同 ρ 值之分群，其範圍從 0.9 至 0.5，以 0.05 作為間隔單位，結果如圖 5 所示。

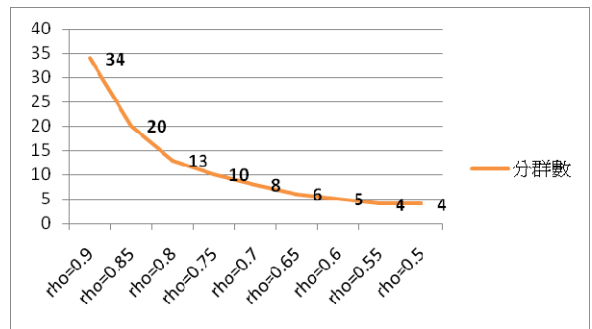


圖 5 不同 ρ 值之分群數比較

當 $\rho=0.55$ 時，其分群數開始達到收斂之效果，與其他分群數之差距甚小，故本實驗選擇分群數為 4 之結果作為依據。而根據 Fuzzy ART 之執行結果，文件與各分群之間的分佈關係如表 2 所示。 ρ 值主要影響著分群數的多寡，對於各種資料類別之內容而言，皆無差異。

表 2 各群組之文件數

	#1	#2	#3	#4
文件數	84	47	66	103
累加文件數	84	131	197	300

3.4 語意檢索

根據 SVD 之文件向量 V 矩陣（右奇異向量），於每個文件所對應到的關鍵字集合中，取其最大值所對應之關鍵字作為該文件之代表性詞彙。而根據 Fuzzy ART 之分群結果，可匯整出每個群組所對應之文件集合，而經由 LSI 之萃取文件代表性詞彙的過程，各群組則可表示為若干個代表性詞彙的集合，如表 3 所示，其結果為部份範例。

表 3 各群組之代表性詞彙

群組一		群組二	
刺鼻	乾乾	刺刺	毛孔
消炎	反應	傷口	變白
刺激	仔細	暗沉	皮膚
普通	價錢	質地	酒精味
嗆死	整體	日曬	天生
會長	濕度	舒服	太多
缺點	...	濕	...

群組三		群組四	
涼涼	濃	有錢	消炎
淡淡	好險	作用	不錯
清爽	地方	味道	酒精味
透	有限	明亮	透明
普通	淺	明顯	痘痘
浪費	明顯	長	黏
刺激	...	重點	...

3.5 關聯法則之提昇解釋性

從表 3 中可發現，逐筆解讀每個關鍵字之意義頗為困難，由於字彙缺乏該程度之描述，譬如以關鍵字「功效」而言，意味著該產品具有功效？亦或該產品具有何種功效？其模糊性質導致無法從關鍵字揣測其代表之意涵。

為解決此一問題，本研究導入關聯法則以提昇關鍵字之可讀性。以表 4 為實驗結果範例，從群組三的關聯結果可推測出，該群組之使用者著重於二個要項—「美白」及「效果」。此外，並能解讀成該產品為使用於皮膚，主要功能為美白肌膚之用，其效果明顯，並感覺清爽。

表 4 導入關聯法則之結果

群組三			
概念	項目	關聯法則	Sup% times
概念一：美白	Item1	酒精味→美白	6.154% 4
	Item2	效果→美白	20% 13
	Item3	皮膚→美白	7.692% 5
	Item4	臉→美白	6.154% 4
	Item5	白→皮膚	6.154% 4
概念二：效果	Item6	效果→酒精味	9.231% 6
	Item7	效果→清爽	6.154% 4
	Item8	效果→明顯	7.692% 5
	Item9	效果→酒精味→ 美白	6.154% 4

由分析結果可發現，關聯法則有效提升了關鍵字之可讀性，比較於單一詞彙之結果，關聯法則以詞彙組合的方式呈現，有助於降低其模糊性質，使企業決策者或消費者更易於明白其意涵，對於改善產品或購買決策而言，該分析結果能提供更為精確之參考方針。

4. 結論

隨著部落格使用性的普及，其儼然已成為一項不可或缺之溝通媒介，使用者藉由部落格與其他部落客分享彼此的生活及想法，導至輕易地便可從部落格當中，查詢到豐富的資訊。然而，在資訊過載的情況下，使用者及企業如何從中有效地發掘重點資訊為其關鍵要項。本研究提出一套 FAIR 模組，並以美容保養品—雪肌精為例，以探勘部落格之關鍵資訊為目的。

本研究嘗試以 FAIR 模組，利用 Fuzzy ART 將關於雪肌精化妝水之部落格文章做分群，內容性質相似的文章得以群聚，從各群組中也可觀察出消費者對於該產品有哪些主要的特徵及認知，並以 LSI 及 AR 幫助解讀各群組之意涵。於實際執行該模組流程後發現，FAIR 模組可確實擷取出該產品之主要特徵，於未事先了解該產品資訊之前，透過 FAIR 之分析結果，即可大約掌握其產品之特徵資訊，對於企業決策者而言，能以即時且最有效之方法，掌握消費者對於該產品之重點認知及產品動向。

此外，本實驗於關聯法則之執行結果發現，其有效之規則數隨著詞彙數量的增加，有明顯遞減的情況，將造成提升詞彙解釋性時，可能發生無關聯規則的情形。根據推測，由於類似含意之詞彙為數眾多，譬如意旨皮膚透亮之詞彙，即包含了透明、透亮、透白、變透、變白、不黯沉等，大量近義之詞彙將分散文件之關聯性，使得規則數大為減少。應擬定將此類近義之詞彙視為同一詞彙，以解決規則數過少之情形。

5. 致謝

本研究受到國科會計畫(契約編號 NSC 96-2416-H-324 -003 -MY2)部分贊助，作者在此表達感謝之意。

參考文獻

- [1] R. Agrawal, T. Imielinski, and A. Swami, "Mining Association Rules between Sets of Items in Large Databases," *ACM SIGMOD International Conference on Management of Data*, Vol. 22, No. 22, pp. 207-216, 1993.
- [2] R. Agrawal and R. Srikant, "Fast Algorithm for Mining Association Rules," *20th*

- International Conference on Very Large Data Bases*, pp. 487-499, 1994.
- [3] A. N. Bonnie and J. S. Diane and G. Michelle, and S. Luke, "Why We Blog," *Communications of the ACM*, Vol. 47, No. 12, pp. 41-46, 2004.
- [4] M. Chau and J. Xu, "Mining Communities and Their Relationships in Blogs: A Study of Online Hate Groups," *International Journal of Human - Computer Studies*, Vol. 65, No. 1, pp. 57-70, 2007.
- [5] L.-S. Chen and C.-T. Su, "Using granular computing model to induce scheduling knowledge in dynamic manufacturing environments," *International Journal of Computer Integrated Manufacturing*, Vol. 21, No. 5, pp. 569-583, 2008.
- [6] Y. Chen, F. S. Tsai, and K. L. Chan, "Machine learning techniques for business blog search and mining," *Expert Systems with Applications*, Vol. 35, pp.581-590, 2008.
- [7] G.A. Carpenter, Grossberg, S. and Rosen, D.B., "Fuzzy ART: fast stable learning and categorization of analog patterns by an adaptive resonance system," *Neural Networks*, Vol. 4, pp.759-771, 1991.
- [8] S. Deerwester, S. T. Dumais, T. K. Landauer, G. W. Furnas and R. A. Harshman, "Indexing by latent semantic analysis," *Journal of the Society for Information Science*, Vol. 41, No. 6, pp. 391-407, 1990.
- [9] F. M. Facca and P. L. Lanzi, "Mining interesting knowledge from weblogs: a survey," *Data & Knowledge Engineering*, Vol. 53, pp. 225-241, 2005.
- [10] S. Kamal and L. I. Burke, "FACT : A New Neural Network-Based Clustering Algorithm for Group Technology," *International journal of production research*, Vol. 34, No. 4, pp. 919-946, 1996.
- [11] Y.-S. Lin, and J.-Y. Huang, "Internet Blogs As A Tourism Marketing Medium: A Case Study," *Journal of Business Research*, Vol. 59, pp. 1201-1205, 2006.
- [12] P. Sungjine, "Neural Networks and Customer Grouping in E-commerce:A Framework Using Fuzzy ART," *Proceedings of Academia/Industry Working Conference*, pp. 331-336, 2000.
- [13] C.-T. Su, L.-S. Chen and Y. Yih, "Knowledge acquisition through information granulation for imbalanced data," *Expert System with Applications*, Vol. 31, No. 3, pp. 531-541, 2006.
- [14] S. Todoroki, T. Konishi, and S. Inoue, "Blog-Based Research Notebook: Personal Informatics Workbench for Hightthroughput Experimentation," *Applied Surface Science*, Vol. 252, No. 7, pp. 2640-2645, 2006.
- [15] 安德魯·基恩著，尤傳莉譯，*你在看誰的部落格？*，早安財經文化有限公司，2008年。
- [16] 陳穆臻，青蛙變身術 Blog 行銷有多行？！，*管理雜誌*，第 376 期，第 114-117 頁，2005 年。
- [17] 蔡孟珈，個人部落格內容出版之探討，碩士論文，世新大學圖文傳播暨數位出版系，2007 年。
- [18] 謝正嘉，應用 Fuzzy ART 與啟發式規則偵測資料庫異常連線，朝陽科技大學資訊管理系，2006 年。
- [19] 彭華瑞，應用潛在式語意分析於語言模型之研究，碩士論文，成功大學資訊工程系，2002 年。
- [20] 吳佳昇，使用貝氏潛在語意分析於文件分類及資訊檢索，碩士論文，成功大學資訊工程系，2005 年。