

# 利用模糊推論探討瀏覽行為的擷取與分析 在網路廣告上的應用

## An Application of Capturing and Analysing Browsing Behavior on the Web Advertising Using Fuzzy Inference

吳憲忠

國立暨南大學資訊管理研究所

545 南投縣埔里鎮大學路 1 號

hcwu@im.ncnu.edu.tw

陳宏泰

國立暨南大學資訊管理研究所

545 南投縣埔里鎮大學路 1 號

s9213509@ncnu.edu.tw

### 摘要

網路廣告一直是網站經營者最重要的收入財源，雖然目前網路廣告的技術與機制日漸完善，但是廣告主對於網路廣告所帶來的效益一直也是抱持著懷疑的態度，主要的原因在於廣告商無法掌握使用者的特性及偏好以致於提供不適切的廣告訊息。在此本研究提出了一個以模糊推論來分析使用者瀏覽網頁行為的方法，藉此來獲得其偏好，同時也提出了一個個人化廣告資訊推薦架構，此架構是以網路廣告聯播制為基礎，利用模糊推論以及學習使用者輪廓 (profile) 與過濾資訊的方法來推薦個人化的廣告資訊，以達到一對一的行銷，並提升網路廣告的效益。

關鍵詞：網路廣告、個人化、模糊推論

### ABSTRACT

Web Advertising is the most import finance source to web site. Even though current techniques and mechanisms of web advertising are getting maturity, advertisers always doubt the benefit of web advertising. The main reason is that the advertisement dealer can not understand the characters and preferences about user so that they provide inappropriate advertisement information to users. Therefore, this paper proposes a mechanism that can analyze the web-browsing behavior of user by using fuzzy inference to capture the user's preference. On the other hand, we also propose a recommended architecture of personalized advertisement. This architecture is based on web advertising association mechanism and applies fuzzy inference, learning user's profile, and information filtering techniques to recommend personalized advertisement information that arrives at one-to-one marketing and improves

benefit of web advertising.

Keywords : Web Advertising,  
Personalization, Fuzzy  
Inference

### 一、緒論

根據資策會市場情報中心 (MIC) 結合多家網路公司與廣告公司，最新完成的 2000 年網路廣告調查[3]，2000 年台灣網路廣告市場為 8.7 億元，比起 1999 年的 3.5 億元成長了 149%，2001 年預估將可成長 105%，達到 17.8 億元。另外，根據 MIC 這項調查，台灣網友最常接觸的媒體，除電視(占 90%)外，其次就是網路(79%)，顯示網路在台灣儼然已成為第 2 大媒體。2004 年台灣網路廣告市場則預估可突破 60 億元，網路廣告金額占整體廣告市場的比率 2000 年為 1%，預期 2004 年將可突破整體廣告量的 5%。

在網路廣告服務的機制上，因廣告主、網站經營者和廣告商的不同需求，演化出兩種不同的服務機制：單站廣告機制以及網路廣告連播制。所謂單站廣告機制就是網站經營者獨自處理有關網路廣告的業務，廣告主必須與其接洽，購買這個網站上的廣告時段及頁面，而該網站必須依廣告主購買時段、頁面以不同的技術播放網路廣告。另外，網路廣告連播制就是網站經營者以加盟的方式加入聯播網，將所有廣告的業務委外由聯播網承受，如此一來網站可以致力於網站本業的經營。在廣告連播制服務的機制下，聯播網扮演著非常重要的角色，它不但降低了網站經營者建置於維護廣告主機的成本，同時也減少了廣告主搜尋適當網站所產生的交易成本。

然而在各個聯播網中，目前大都是提供廣告主及廣告商有關網路使用者點選各類廣

告的人口背景統計資料，以協助完成網路廣告的計劃、購買和檢討等複雜的工作，進而提供廣告主最適合的網路廣告建議。在這樣的運作模式下所遞送的廣告並非都能滿足所有廣告主和使用者，實際上，使用者的需求是能夠收到有興趣的資訊（廣告），而廣告主則希望能透過廣告資訊服務者（廣告網站或聯播網）來接觸到潛在的消費者。因此聯播網業者若能夠依照使用者的偏好來提供相關廣告資訊，不但能消除使用者對網路廣告不滿的刻板印象，也能夠有效提供廣告主合適的銷售對象，進而促成交易。

## 二、文獻探討

### （一）網路廣告播放技術

網路廣告播放的技術影響著廣告的效果，其種類大約可以分成四種[5][16]：

#### 1. 無目標式（Untargeted）

是較早期系統所使用的方式，在一定的時段以亂數或固定的方式播放廣告，此方式的優點是技術簡單，但是不易對準使用者的胃口，而使得廣告效益很低。

#### 2. 編輯式（Editorial）

針對網站內容或網頁主題的目標使用者，排定相關廣告，以增加廣告效益。例如 Yahoo 的網站[14]約有 1 百萬種不同的網頁主題，廣告主可以根據其潛在客戶可能有興趣的主題中安插廣告。

#### 3. 目標式（Targeted）

是所有技術中目前應用最廣泛的一種，廣告主可以指定客戶的某些參數，如客戶所使用的瀏覽器類型、國家、上網的時間或 ISP 等。而廣告的播放則是透過廣告伺服器來分析使用者是否符合這些參數，一旦使用者符合了某廣告主所設定的參數，廣告伺服器則送出此廣告。

#### 4. 個人化（Personalized）

是最新一代的網路廣告播放技術，以類神經網路 [8][18]、相似度指標（Similarity Indexing）[20]等方式，利用使用者個別的瀏覽行為與互動記錄來學習使用者的偏好，並且選擇與使用者偏好相關的廣告。

### （二）相關系統

近年來由於網際網路的濫觴，導致了資訊過載的情況日益嚴重，使用者暴露在如此資訊氾濫的環境中要如何才能有效的取得所需的資料也成為了一項重要的課題。個人化的技

術也因應而生，廣泛地應用在各個不同的領域中，協助使用者過濾不必要的資訊。在網路廣告的相關產業中也逐漸重視這個問題，期望這樣的技術能夠幫助他們解決點選率（Click Through Rate, CTR）低迷不振的窘境。因此有越來越多的研究便逐漸朝向個人化、客製化的網路廣告來思考。

ADWIZ 系統[14]便是以非干擾式的方法來取得使用者的短期興趣，並且來推薦客製化的廣告，所謂的短期興趣指的是，使用者所輸入欲查詢資訊的關鍵字，或是由瀏覽網頁的 URL 所轉換而成的關鍵字。以相似度指標的方式來儲存並過濾某廣告的關鍵字的出現機率是否與使用者的短期興趣相符，並且推薦出與短期興趣相符程度較高的廣告。

李麗華 & 陳致傑[5]提出了一個以代理人為主之智慧型網路廣告行銷架構（Agent-base Internet Smart Ad architecture, AISA），其利用已知的屬性及統計資訊當作學習範例，並且利用自適共振網路（Adaptive Resonance Theory Network, ART）從中學習內聚規則，系統則依照使用者與廣告主需求之特性，找出其適合聚類，作為分類之依據。

李維平等人[4]提出了以資料探勘的技術來產生個人化廣告。此研究主要是以資料探勘的兩種技術：分類（Classification）與順序（Sequence），方法的整合應用，提出一套完整的瀏覽模式。在此模式中依據歷史客戶的特性將資料加以分類，建立一套能夠預測未來新進使用者瀏覽行為的瀏覽樹，以提供適切的廣告。

丁一賢等人[1]提出了一個廣告模式，這個模式同時使用了網頁使用探勘與網頁內容探勘，來過濾使用者的行為並且擷取其偏好。以便日後作為客製化廣告推薦之用。

### （三）瀏覽行為的監視

對於使用者行為的監視在 USENET news 環境中的研究已經相當的成熟了，在其中常見被監視的項目包含了有三大類 [15]：檢視（Examination）、記憶（Retention）、參考（Reference）。這樣的一個監視活動主要的就是在獲得使用者對文章隱性回饋，也就是使用者對文章的偏好程度。在 [12] 研究當中更證明了閱讀的時間（檢視類）是用來判斷隱性回饋相當有用的根據，同時若一併使用記憶類（如列印）相關的操作行為來作分析的依據，更能有效的來判斷使用者興趣。

然而近年來，已經有相關的研究應用了上述的理論在網際網路上。在 [7] 的研究中就以

是否有將網頁標籤 (Bookmark) 的行為作為判斷使用者的偏好。而在[1]中則單是以閱讀時間來判斷偏好。

另外, WAIR 系統 [17] 是以多階層類神經網路來學習其不同行為的權重, 其中包含了閱讀的時間、是否有將網頁標籤、是否有使用捲軸 (Scroll Bar) 以及是否有經由此網頁連結到別的網頁中。將學習後的權重作為日後分析使用者興趣的依據。而在 Powerize Server [11] 系統中, 除了上述幾種行為外更是加入了如是否有標示文字、是否有重複閱讀、是否有儲存以及是否有剪貼的行為發生等項目來判斷使用者偏好。

### 三、個人化廣告資訊推薦系統

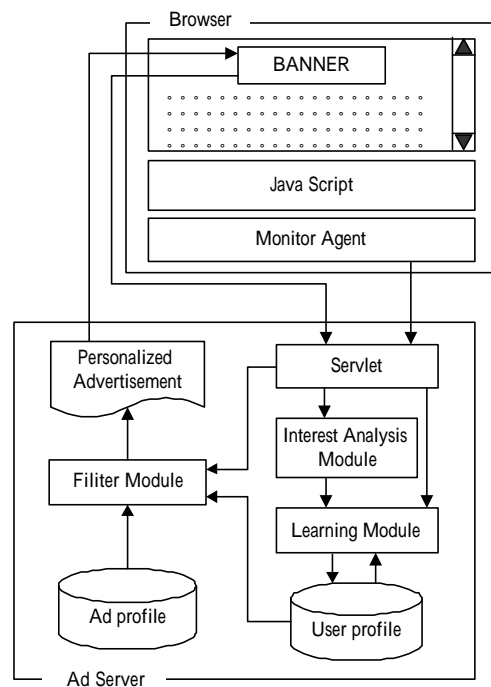
個人化廣告資訊推薦架構如圖一所示, 整個架構是由兩個部分所組成的, Browser (Client 端) 和 Ad Server (Server 端)。在網路廣告聯播制度下, 當使用者向廣告聯盟的成員網站要求 (Request) 一個網頁時, 該網站會回應 (Response) 被要求的網頁, 然而在這個網頁中會有一個橫幅廣告 (Banner) 是來自於 Ad Server。當 Browser 再度向 Ad Server 要求 Banner 時, Ad Server 會取回以往在使用者的電腦中所儲存的 cookie (若不存在則表示為新的使用者, 此時必須在使用者的電腦中新設定一個能辨別使用者身份的 ID), 來辨別使用者的身份, 以便傳送個人化的廣告資訊 (即 Banner) 給使用者。除了在網頁中嵌入 Banner 外, 還必須在網頁中加入用來監視使用者操作習慣的 JavaScript 與一個以 Java 所撰寫而成的 Monitor Agent。Monitor Agent 是透過 JavaScript 來取得使用者的操作習慣, 並且當使用者切換網頁時將收集到的資訊傳回 Ad Server, 以作為分析使用者興趣之用。如表一, 為使用者被監督的項目。

Ad Server 的主要功能在於學習使用者的偏好, 並且將依照偏好來過濾使用者有興趣的廣告資訊, 將廣告傳回 Client 端。當 Ad Server 接收到 Monitor Agent 所送回來的資訊後, 就會準備開始進行學習的動作。首先, 興趣分析模組會取得滑鼠移動事件、被標示文字的個數、Scroll Bar 改變的幅度、網頁中所有

文字的數目和瀏覽時間來分析使用者興趣。在取得使用者興趣後, 將此回饋和網頁關鍵字輸入輪廓學習模組, 進行長期興趣的學習。最後將結果存回使用者輪廓資料庫中, 以便日後作為過濾廣告資訊的依據。

廣告主在刊登廣告時被要求輸入目標族群感興趣的關鍵字, 系統將關鍵字存入廣告輪廓 (Ad profile) 資料庫中。當使用者進入網站後系統會透過過濾模組進行廣告輪廓和使用者的輪廓的比對工作, 找出和使用者輪廓相似度最高的廣告, 也就是使用者有興趣的廣告, 將這個廣告傳回。

本系統包含了使用者興趣分析模組 (Interest Analysis Module)、使用者輪廓學習模組 (Learning Module) 和廣告資訊過濾模組 (Filter Module), 在以下各節分別說明。



圖一. 個人化廣告資訊推薦架構

表一. 使用者瀏覽行為

項目	說明	用途
滑鼠移動事件	在瀏覽器中滑鼠移動事件發生的次數	判別使用者對網頁的操作複雜度 (興趣分析模組)
被標示文字的個數	在閱讀的過程中, 網頁內的文字被標示個數的總和	判別使用者對網頁的操作複雜度 (興趣分析模組)

Scroll Bar 改變的幅度	在瀏覽網頁時，使用者每次移動 Scroll Bar 幅度的總和	判別使用者對網頁的操作複雜度（興趣分析模組）
網頁中所有文字的數目	此文字的數目並不包含網頁中的 HTML 標籤	判別使用者對網頁的興趣程度（興趣分析模組）
瀏覽時間	使用者在此網頁停留的時間	判別使用者對網頁的興趣程度（興趣分析模組）
網頁關鍵字	網站經營者對此網頁的敘述，用數個關鍵字來表達	作為學習使用者長期興趣之用（輪廓學習模組）
網頁的內容	網頁內所有的文字。若是網站經營者不提供關鍵字時，則 Monitor Agent 會將網頁內容取回，並且自動擷取出代表網頁的關鍵字	作為學習使用者長期興趣之用（輪廓學習模組）

### (一) 使用者興趣分析模組

在資訊擷取與過濾的過程中，使用者通常無法用很明確的字眼表示他的興趣，但是當使用者在看完一篇文章後卻能輕易的表達出對此文章是否喜好，以及喜好的程度，我們稱之為相關性回饋（Relevance Feedback）。然而這樣的回饋方式大體可以分成兩類，顯性回饋（explicit feedback）和隱性回饋（implicit feedback）。顯性回饋指的是使用者在閱讀完文章後，明確地回應系統此文件和他所偏好資訊的相關程度。此一方式的優點是，簡單並且可以精確的取得使用者回饋，但是他的缺點是，必須要求使用者輸入對此文件的分數，這樣的動作會干擾到使用者。另一個方法是隱性回饋，藉由觀察使用者的行為來獲得對文件的偏好，它的優點是不會干擾到使用者。

興趣分析模組主要的功能就是以隱性回饋的方式來分析使用者對於這篇網頁的偏好程度，以避免干擾使用者為前提來獲得使用者偏好。在此我們利用二階段的模糊邏輯推論（採用 Mamdani 推論引擎）來分析使用者在瀏覽器中操作的情況以及網頁的特性，並且推論出使用者的興趣。第一階段的推論用來決定使用者操作的複雜程度，第二階段則是利用前一階段的輸出加上其他的變數來推斷使用者的興趣。在下一章中會更詳細說明分析的原理。

### (二) 使用者輪廓學習模組

此模組主要的功能是學習使用者的輪廓（User profile）。透過 Monitor Agent，我們可以取得可以用來表示此網頁的關鍵字，並且也可以由興趣分析模組得到使用者對於此網頁的偏好程度，進一步將以代表此網頁的關鍵字與偏好程度進行長期興趣的學習，並將訓練結果存回使用這輪廓資料庫。

而使用者輪廓的結構是採用向量空間模型（Vector Space Model, VSM）[19]，也就是

說代表使用者興趣的興趣描述檔是以關鍵字形成的向量表示。同時每個關鍵字皆有對應的權重，在此關鍵字的權重代表的也就是使用者對關鍵字的偏好程度。

對於使用者輪廓學習的方法，本研究使用 WH (Widrow-Hoff) 演算法[13]來學習使用者輪廓，它是一個 on-line incremental 的演算法，透過此演算法我們可以即時修改使用者的偏好。

### (三) 廣告資訊過濾模組

過濾資訊模組的功能是當使用者的瀏覽器向 Ad Server 要求廣告時，將個人化廣告資訊傳遞給使用者。個人化廣告的取得是分別由使用者輪廓資料庫與廣告輪廓資料庫取得使用者與廣告的輪廓檔，再將兩者作相似度運算，取出相似度最高的廣告推薦給使用者。

相似度衡量的公式皆源自於下列幾種 [10]: 隸屬度和與差 (difference and the sum of grades of membership)、幾何距離模式 (geometric distance model) 和集合理論 (set-theoretic approach) 中的聯集與交集運算。而在 [6] 的研究中更證明 [9] 於 2000 年所提出的相似度衡量公式，在眾多的公式中其衡量結果較佳。因此本系統即採用 [9] 中所提出的相似度衡量方法。

## 四、使用者興趣分析

當使用者在一個網頁上瀏覽時，他對這篇文章的感覺除了存在他的心中外，可以令我們直接感受到就是對瀏覽器的操作。例如，使用者對某網頁感興趣時，他的操作行為會是不斷的翻頁（利用 Scroll Bar）來看完這篇文章，並且利用滑鼠來指向目光所正在閱讀的位置，更進一步的行為就是將重要的地方標示起來。對於這樣閱讀的操作習慣，我們可以監視

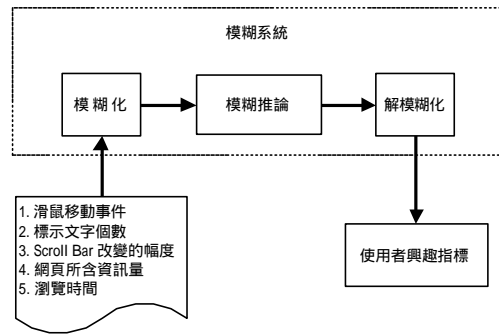
他的行為來加以猜測使用者對網頁的偏好程度。

就上述的觀點，本研究是以模糊推論的方式求得使用者興趣。藉由將滑鼠移動事件、被標示文字的個數、Scroll Bar 改變的幅度、網頁所含的資訊量和瀏覽時間等變數模糊化，並以模糊推論求得使用者興趣，最後再將模糊語句解模糊化成為使用者興趣指標，整個架構如圖二。

在模糊系統的設計上，我們採用 Mamdani 推論引擎與二階段模糊邏輯推論來分析使用者在瀏覽器中操作的情況以及網頁的特性，並且推論出其興趣，如圖三為此一模糊系統的結構。第一階段是在判斷使用者對網頁的操作複雜度。所謂操作複雜度是使用者在瀏覽網頁的同時，由直接對瀏覽器操作的行為所構成的，這些行為可能包括鍵盤或是滑鼠的使用。例如在一個網頁上複雜程度很高的操作行為會是，使用者經常移動滑鼠、不斷的翻動網頁，甚至於會有標示某些文字的動作。

第二階段的推論則是判別使用者對網頁的興趣程度。單以操作的複雜度並不能完全代表偏好程度，我們需要其它的資訊來輔助才能更精確的來判斷，其中包含了網頁中所含的資訊含量，以及瀏覽時間。舉例來說，在一個資訊含量很高的網頁中，瀏覽的時間卻很小，無

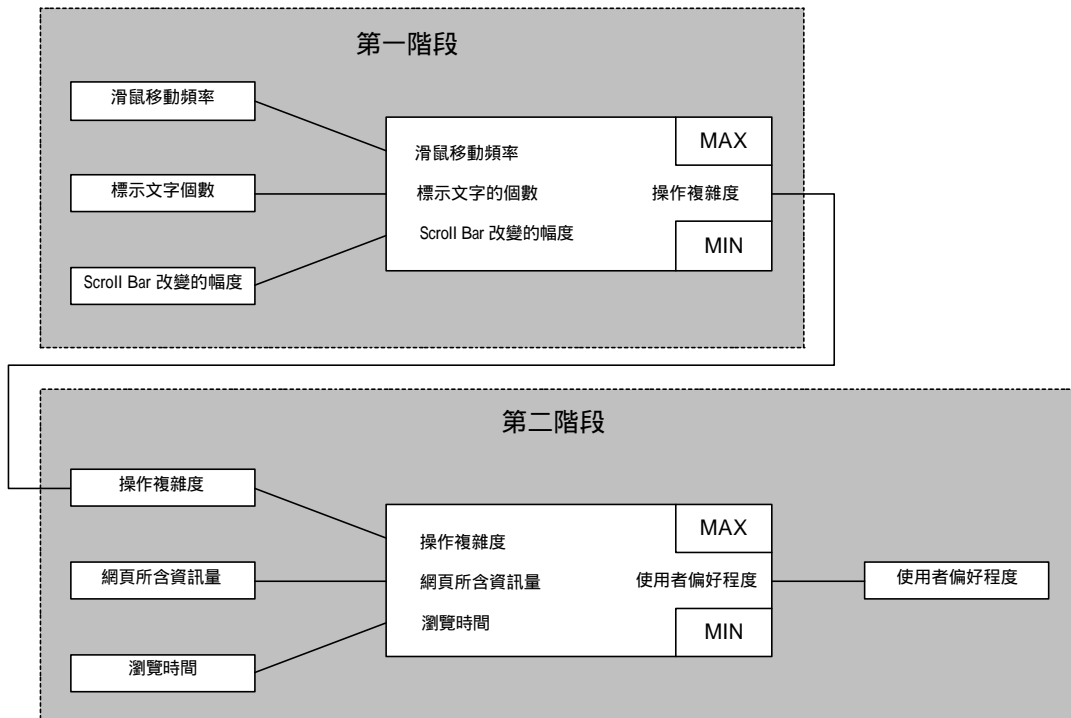
論你的操作複雜度是多少，我們都可以判定偏好程度是很低的。以下我們分成四小節來詳細說明模糊系統。



圖二. 使用者興趣推論模式

### (一) 定義輸出入語意變數

第一階段的推論是使用“滑鼠移動事件”、“標示文字個數”與“Scroll Bar 改變的幅度”用來決定“操作的複雜程度”。第二階段則是使用“操作的複雜程度”加上“網頁所含資訊量”，和“瀏覽時間”來推斷“使用者的偏好程度”。在不論是第一階段或是第二階段在語意值的部分，在此都是以三個尺度的語意值來表示，如表二及表三分別代表第一及第二階段的輸出入語意變數和語意值。



圖三. 模糊系統結構圖

表二. 第一階段輸出入語意變數和語意值

輸入		
語意變數	說明	語意值
Mouse_Move	滑鼠移動事件發生的次數	“少”，“普通”，“多”
Text_Highlight	被標示文字之多寡程度	“少”，“普通”，“多”
Scrolling_Range	Scroll Bar 改變的幅度	“小”，“普通”，“大”
輸出		
語意變數	說明	語意值
Op_Complexity	操作的複雜程度	“低”，“普通”，“高”

表三. 第二階段輸出入語意變數和語意值

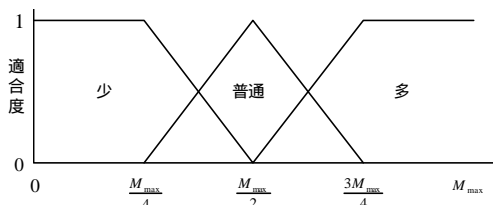
輸入		
語意變數	說明	語意值
Op_Complexity	操作的複雜程度	“低”，“普通”，“高”
Information_Capacity	此網頁所含的資訊量	“少”，“普通”，“多”
Browse_Time	使用者瀏覽網頁的時間	“短”，“普通”，“長”
輸出		
語意變數	說明	語意值
User_Interest	使用者對此網頁的偏好	“低”，“普通”，“高”

(二) 訂定歸屬函數

在模糊邏輯相關的文獻中，歸屬函數有許多不同的形式，但在實際的應用中，大都還是以三角形、Z型與S型歸屬函數為主。在此本研究即是採用這三種歸屬函數，相關輸出變數的歸屬函數定義如下。

1. Mouse\_Move 的歸屬函數

滑鼠移動事件所發生的次數多寡之判斷，原則上會因為檔案越大移動的次數越多，所以歸屬函數的範圍是動態可以更改的，換句話說檔案越大意味著範圍會也會越大。因此 Mouse\_Move 的範圍會介於 0  $M_{max}$  之間，其中  $M_{max}$  為此次使用者所瀏覽網頁的檔案大小所相對應到的最大滑鼠移動次數，而最大滑鼠移動次數的取得是根據系統的歷史資料而來的。在此 Mouse\_Move 的歸屬函數如圖四所示。

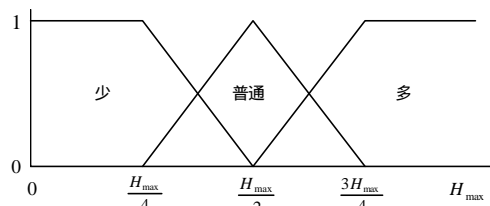


圖四. Mouse\_Move 的歸屬函數

2. Text\_Highlight 的歸屬函數

被標示文字之多寡程度的判斷是根據每

次閱讀文章中文字的個數，也就是說此歸屬函數的範圍是可以動態更改的，它是隨著每次閱讀文章中文字個數的不同其範圍也會隨著變化。例如，對於一個大小為 100 個字的文章 A，使用者的操作行為中標示文字的個數為 50，此時在 Text\_Highlight 中語意值“普通”的適合度為 1。但是在同樣的操作情況下（標示文字的個數為 50），在一個大小為 50 個字的文章 B 中，Text\_Highlight 的語意值“普通”之適合度卻為 0。因此就使用者瀏覽的網頁而言，Text\_Highlight 的範圍會介於 0  $H_{max}$  之間，其中  $H_{max}$  為這次閱讀文章中的文字總數。在此 Text\_Highlight 的歸屬函數表示如圖五。



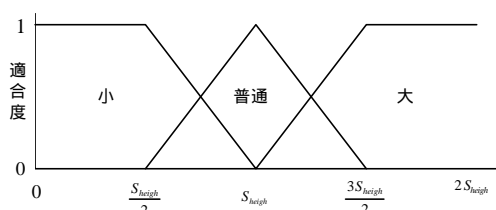
圖五. Text\_Highlight 的歸屬函數

3. Scrolling\_Range 的歸屬函數

使用者在瀏覽時進行翻页動作的改變幅度判斷是根據捲軸的高度，所謂捲軸的高度指的是捲軸的底端和頂端之間的差。我們的視埠（viewport）會在捲軸的頂端與底端之間遊走，每當使用者在進行翻页的動作時視埠就會改變，藉由紀錄視埠在捲軸的底端和頂端之間

移動的位置差，可以將使用者在這次瀏覽網頁中改變視埠的差值加總起來，並且與捲軸的高度來作比較，以便求得翻頁的改變幅度。

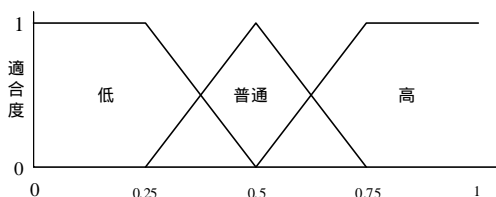
假設在一般的情況下，使用者會看完網頁，換句話說視埠會移動約和捲軸的高度相同，我們令這種情況的語意值為“普通”。所以 Scrolling\_Range 的範圍會介於  $0 \sim 2S_{high}$  之間，其中  $S_{high}$  為捲軸的高度。在此 Scrolling\_Range 的歸屬函數如圖六所示。



圖六. Scrolling\_Range 的歸屬函數

#### 4. Op\_Complexity 的歸屬函數

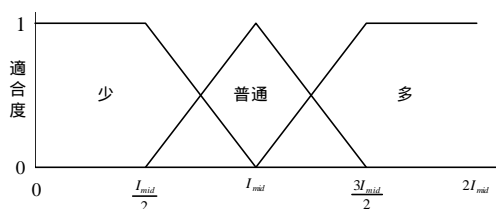
對於第一階段的輸出，也就是使用者操作的複雜度，其歸屬函數的範圍會介於  $0 \sim 1$  之間，而數值越大所代表的意義是使用者操作的複雜度越高。所以 Op\_Complexity 的歸屬函數表示如圖七。



圖七. Op\_Complexity 的歸屬函數

#### 5. Information\_Capacity 的歸屬函數

網頁所含的資訊量是以網頁中所含的文字數目來衡量，文字數目越多，表示資訊含量越大。然而對於 Information\_Capacity 歸屬函數的設定我們是依照系統所收集的歷史資料而來的，假設 Information\_Capacity 的範圍會介於  $0 \sim 2I_{mid}$  之間，其中  $I_{mid}$  為網頁平均字數，此網頁平均字數是由系統日誌檔中“網頁字數”欄位而求得的。所以 Information\_Capacity 的歸屬函數表示如圖八。

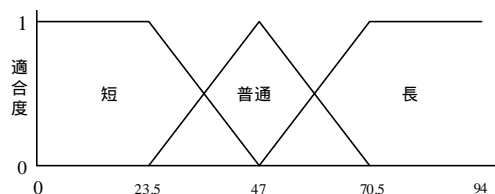


圖八. Information\_Capacity 的歸屬函數

#### 6. Browse\_Time 的歸屬函數

根據 Nielsen/NetRatings 對全球 21 個國家進行全球網際網路指標 (Global Internet Index) 之調查[2]，在 2001 年三月公布的全球趨勢研究報告指出，使用者在瀏覽網頁時平均的停留時間為 47 秒。瀏覽網頁時間的意義是使用者在閱讀網頁時所花費的時間，而在系統中，會在網頁被載入時開始計時，直到使用者離開此網頁為止，它是除了 Op\_Complexity 外衡量使用者興趣的一個重要指標。

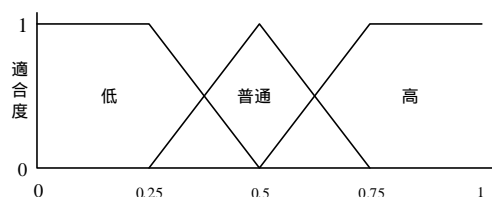
在 Browse\_Time 的歸屬函數制訂中，本研究採用的是上述報告的結果，也就是一般而言使用者在網頁的停留時間大約是 47 秒，換句話說若網頁的停留時間為是 47 秒時，則 Browse\_Time 的語意值“普通”的適合度為 1。所以 Browse\_Time 的範圍會介於  $0 \sim 94$  之間。在此 Browse\_Time 的歸屬函數表示如圖九。



圖九. Browse\_Time 的歸屬函數

#### 7. User\_Interest 的歸屬函數

與對於第一階段的輸出相同，使用者對此網頁的偏好程度，其歸屬函數的範圍會介於  $0 \sim 1$  之間，而數值越大所代表的意義是使用者對此網頁的偏好程度越高。所以 User\_Interest 的歸屬函數表示如圖十。



圖十. User\_Interest 的歸屬函數

### (三) 模糊規則制訂

興趣分析模組在本系統中擔任了分析使用者興趣的任務，對學習模組而言它的輸出更是一項學習使用者偏好的重要資訊。然而模糊規則的制訂更是整個興趣分析模組的關鍵，它關係到了推論結果的正確性，所以在制訂規則時我們必須格外小心。在進行各階段的規則制訂前，我們必須瞭解各階段輸入的語意變數中真正的意涵，各變數間相互的關係與限制，如此才能夠有效的利用規則來描述現實的情

況，以及簡化規則庫中的模糊規則進而提升推論的速度。

在第一階段中的三個輸入變數，“Mouse\_Move”、“Text\_Highlight”與“Scrolling\_Range”所代表的都是使用者在瀏覽網頁中常見的行為，而其中“Mouse\_Move”與“Text\_Highlight”的關係最為密切，原因是當操作網頁時這兩項動作都移動到滑鼠。這會使我們聯想到，當“Text\_Highlight”很高時“Mouse\_Move”必定不可能是很低，其實不然，當我們在標示文字時動作是先按住滑鼠左鍵在移動滑鼠，這一個動作與單純移動滑鼠的“Mouse\_Move”並不同，所以兩個輸入是獨立的。另外有關於“Scrolling\_Range”的部分，在本系統中用來偵測的方式是利用監視可視文章最頂端的相對位置值是否有改變，所以不論使用者利用滑鼠或是鍵盤來翻頁，翻頁的幅度都會被計算到。

為了使得規則庫中的模糊規則更能夠代表實際的情況，以及刪除不合理的規則，第一階段中規則的建立符合下列幾項原則：

- 基本上代表“Mouse\_Move”、“Text\_Highlight”和“Scrolling\_Range”三個輸入變數的語意值越大，“Op\_Complexity”也越高。
- 在“Mouse\_Move”、“Text\_Highlight”和“Scrolling\_Range”三個輸入變數的語意值中，只要有兩個變數的語意值為低（小或少），無論第三個變數的語意值為何者，“Op\_Complexity”的語意值都為“低”。例如在“Mouse\_Move”為“少”和“Text\_Highlight”為“少”的情況下，無論“Scrolling\_Range”的語意值為任一個，“Op\_Complexity”的語意值都為“低”。
- 不使用“Mouse\_Move”為“少”和

“Text\_Highlight”為“多”的情況。雖然“Mouse\_Move”和“Text\_Highlight”兩個輸入是獨立的，系統也允許這種情況，但是除非使用者刻意造成，否則這並不符合實際操作的可能行為。

另外在第二階段中所包含的輸入變數有“Op\_Complexity”、“Information\_Capacity”和“Browse\_Time”。同樣地，在第二階段中規則的建立也循下列幾項原則：

- 基本上“Op\_Complexity”與“Browse\_Time”的語意值越大，“User\_Interest”也越高。
- “Information\_Capacity”對於此階段的輸出並沒有直接顯著的影響，而其主要的目的在於檢視瀏覽時間和操作的複雜度是否合理。例如在“Browse\_Time”為“短”和“Information\_Capacity”為“多”或“普通”的情況下，“User\_Interest”為“低”。原因是當網頁具有一定的資訊量時，使用者必須花費一些時間來閱讀，但是若瀏覽網頁的時間很短，這就表示使用者對此網頁並不感興趣。或者是“Information\_Capacity”為“少”和“Browse\_Time”為“長”的情況下，“User\_Interest”的語意值會直接參考“Op\_Complexity”的語意值，原因可能是使用者並沒有在閱讀網頁，但是也沒有將網頁關掉，所以一旦有這樣的情況發生，我們就直接參考操作的複雜度作為第二階段的輸出。

根據了上述原則，本研究訂出的模糊規則共有 36 條，其中第一階段有 21 條，第二階段有 15 條，所有的模糊規則如表四和表五所示。

表四. 第一階段模糊規則

	前件部			後件部
	Mouse_Move	Text_Highlight	Scrolling_Range	Op_Complexity
模糊規則 1	多	多	大	高
模糊規則 2	多	多	普通	高
模糊規則 3	多	多	小	普通
模糊規則 4	多	普通	大	高
模糊規則 5	多	普通	普通	高
模糊規則 6	多	普通	小	普通
模糊規則 7	多	少	大	普通
模糊規則 8	多	少	普通	普通
模糊規則 9	普通	多	大	高



模糊規則 10	普通	多	普通	普通
模糊規則 11	普通	多	小	普通
模糊規則 12	普通	普通	大	普通
模糊規則 13	普通	普通	普通	普通
模糊規則 14	普通	普通	小	低
模糊規則 15	普通	少	大	普通
模糊規則 16	普通	少	普通	普通
模糊規則 17	少	普通	大	普通
模糊規則 18	少	普通	普通	普通
模糊規則 19	少	普通	小	低
模糊規則 20		少	小	低
模糊規則 21	少	少		低

表五. 第二階段模糊規則

	前件部			後件部
	Op_Complexity	Information_Capacity	Browse_Time	User_Interest
模糊規則 1	高		長	高
模糊規則 2	高		普通	高
模糊規則 3		多	短	低
模糊規則 4		普通	短	低
模糊規則 5	高	少	短	普通
模糊規則 6	普通	多	長	高
模糊規則 7	普通		普通	普通
模糊規則 8	普通	普通	長	高
模糊規則 9	普通	少	長	普通
模糊規則 10	普通	少	短	普通
模糊規則 11	低	多	長	普通
模糊規則 12	低		普通	普通
模糊規則 13	低	普通	長	普通
模糊規則 14	低	少	長	低
模糊規則 15	低		短	低

#### (四) 解模糊規化

模糊推論的結果仍是一群語意值，所以必須將推論結果量化以得到明確的輸出值。常見的解模糊化方式有三種包含了面積中心法 (Center of Area, CoA)、平均中心法 (Center Average Defuzzifier)、最大值法 (Maximum Defuzzifier) 較為常見。在此本研究採用的是平均中心法。

## 五、結論

在網路廣告聯播機制中聯播網所扮演的角色即為網路仲介商，其存在的價值在於降低廣告主與使用者的搜尋成本，促進雙方的交易。然而以往的運作模式雖然完善但是仍然缺乏對使用者興趣的瞭解，以致於無法針對使用者推薦適切的廣告資訊，而導致廣告效益不彰

的窘境。因此本研究提出了一個以模糊推論為基礎的技術來監督並分析使用者瀏覽行為及其偏好。同時配合個人化廣告資訊推薦架構，此架構是以網路廣告聯播制為主，藉由上述的使用者興趣分析技術以及學習與過濾資訊的方法來推薦個人化的廣告資訊，目的在於準確地建立起廣告主和潛在消費者間溝通管道，協助使用者過濾不必要的廣告資訊，進一步來提升廣告的效益。

## 六、參考文獻

- [1] 丁一賢, 游耿能, 蕭如淵, 林素穗, “網頁探勘的應用：以客製化網路廣告為例”, 第二屆電子化企業經營管理理論暨實務研討會, pp426-435, 4月, 2001.
- [2] 江政達, “今年3月全球家庭上網人口

- 達三億七千九百萬人”，  
[http://www.find.org.tw/news\\_disp.asp?news\\_id=1434](http://www.find.org.tw/news_disp.asp?news_id=1434).
- [3] 汪欣潔, “MIC 最新調查：台灣網路廣告市場今年為 8.7 億元 預估 2004 年將突破 60 億元”，  
<http://news.pchome.com.tw/times/science/20001222/index-20001222161621068191.html>.
- [4] 李維平、李政權、黃仁傑、黃寶嘉, “使用資料探勘技術產生個人化廣告之研究”, 第二屆網站經營學術暨實務研討會, pp101-110, 12 月, 2000.
- [5] 李麗華 陳致傑, “一個新的代理人智慧型網路廣告行銷架構”, 第二屆網站經營學術暨實務研討會, pp159-167, 12 月, 2000.
- [6] 陳振東, 戴偉勝, “網際網路個人化資訊相似度衡量方法之研究”.
- [7] 楊振偉, “利用書籤功能達到網際網路資訊分享與過濾的技術探討”, 國立中央大學資訊管理研究所, 碩士論文, 1998.
- [8] Aptex Software Inc., “SelectCast Affinity Server”, <http://www.aptex.com/>
- [9] Chen, C. T., “Extensions of the TOPSIS for group decision-making under fuzzy environment”, *Fuzzy Sets and Systems*, Vol 114-1, pp1-9, 2000.
- [10] Chen, S. M., M. S. Yeh and P. Y. Hsiao, “A comparison of similarity measures of fuzzy value”, *Fuzzy Sets and Systems*, Vol 72-1, pp79-89, 1995.
- [11] K. Jinmook., O. Douglas , and K. Romanik, “Using implicit feedback for user modeling in Internet and Intranet searching” Technical Report, College of Library and Information Services, University of Maryland at College Park.
- [12] K. Jinmook, O. Douglas , and K. Romanik., “User modeling for information filtering base on implicit feedback”, <http://citeseer.nj.nec.com/319653.html>.
- [13] L. David D., S. Robert E., C. James P., and P. Ron, “Training Algorithms for Linear Text Classifiers”, *Proceedings of the 19th annual international ACM SIGIR conference on Research and development in information retrieval*, pp298-306, 1996.
- [14] L. Marc, N. Atsuyoshi, A. Naoki, K. Tomonari, and K. Yoshiyuki ,”Un-intrusive Customization Techniques for Web Advertising”, *The Eighth International World Wide Web Conference*, 1999.
- [15] O. Douglas and K. Jinmook, “Implicit feedback for recommender system”, In *Proceedings of the AAAI Workshop on Recommender Systems*, 1998.
- [16] P. Baudisch, D. Leopold, “User-configurable advertising profile applied to Web page banners”, *Proc. of the first Berlin Economics Workshop*, 1997.
- [17] S. Young-Woo and Z. Byoung-Tak, ” Learning user’s preferences by analyzing Web-browsing behaviors”, *Proc. of Int’l Conf. on Autonomous Agents*, pp381-387, 2000.
- [18] T. Ah-Hwee and T. Christine, “Learning user profiles for personalized information dissemination”, In *Proceedings, 1998 IEEE International Joint Conference on Neural Networks*, pp183-188, 1998.
- [19] Y. Budi and Dik L. Lee, “Search and Ranking Algorithms for Locating Resources on the World Wide Web”, *proc. 12<sup>th</sup> Int’l conf. Data Engineering*, New Orleans, pp164-171, 1996.
- [20] Yu P. S., “Data mining and personalization technologies”, *Proceedings of the 6th International Conference on Database Systems for Advanced Applications*, pp6-13, 1999.