

基於增強式學習技術的適性化瀏覽推薦

唐銀伶
國立高雄大學
資訊管理學系
研究生
m1003306@nuk.edu.tw

丁一賢
國立高雄大學
資訊管理學系
助理教授
iting@nuk.edu.tw

摘要

網路上的資訊量大且繁雜，若網站沒有良好的設計，使用者容易在瀏覽時迷失，進而花費許多時間去尋找資料，而網站並沒有最佳化的架構，因此會遇到使用者需求改變的問題，需要使用適性化網站來了解使用者的瀏覽情況，並依照需求調整網頁。然而目前大部分的瀏覽推薦方法並未考慮到使用者在不同時間點的需求來對於網頁進行動態的調整，因此本研究提出的瀏覽推薦方式將以能夠與環境互動的增強式學習為基礎，考量瀏覽次數、瀏覽時間、瀏覽路徑長度、網頁所在階層以及網頁目前排名，再調整以上五種參數之間的權重，目標在於透過所提供的網頁瀏覽推薦方法可以減少使用者到達目標網頁所需要瀏覽的路徑長度。

關鍵詞：網頁使用探勘、適性化網站、增強式學習

Abstract

The explosive growth of the Internet has made information on the web became large and complicated. If website did not have optimized structure, users easily get lost and cannot find the most important information in the right time. The adaptive website can present the information that users need by analyze user behavior. However, visitors may have different needs at different times. Most of recommendation method did not considerate on dynamic or time-dependent needs. This paper presents a recommender system based on reinforcement learning. We considerate five parameters on recommendation, include clicks of the page, time that spent on viewing the page, paths to find the page, hierarchy of the page, and the rank of the page. Using reinforcement learning to adjust the weight of five parameters, we aim to reduce the paths that user needed to find the object page.

Keywords: Web Usage Mining, Adaptive Web Sites, Reinforcement Learning

1. 前言

網路的發展，改變了人們閱讀的習慣，在近幾年變成民眾獲得資訊的主要來源，以新聞媒體為例，2010年網路觀看新聞的人數已經超越了報紙，隨著閱讀方式轉變，廣告管道也轉向網路，在網路廣告的花費也超越報紙，從電視獲取資訊的民眾也不斷減少，特別是年輕族群的觀眾。傳統報章雜誌、書籍等傳播方式能夠呈現的資訊有限，且只能透過圖文靜態呈現，電視傳播透過畫面更生動的提供資訊，但資訊內容仍然由電視台來主導，網路傳播能夠彌補傳統傳播的不足，民眾能有更高的自主權，資訊內容的是由民眾來選擇，能夠更快速有效率獲得所要的資訊。由於網路快速、便利及選擇性高的特性，企業、學校以及政府機關將資訊轉移到網路上，並開始e化作業。

科技使資訊紛紛數位化，造成網路上資訊爆炸，在面對成千上萬的資訊，如何從中找到目標資訊是一門學問，而相對的，網站需要考量如何呈現訊息，才能讓民眾快速找到目標資訊。網路上民眾擁有更高的自主權來選擇想要的資訊，但也表示需要花更多的成本去找尋、去選擇，在資訊e化後，網站不會有專業人員指引，民眾必須靠自己的理解來行動，就像問路及查地圖的差別，因此，在資料量急速增加的環境中，網站需要做出適當的引導，讓民眾能夠依循指引找尋。

為了讓網站能夠引導使用者瀏覽，設計網站時不能忽視網站的架構，但網站並沒有所謂的最佳化架構，若網站架構設計不當或連結錯誤，會造成使用者找不到目標資訊，而版面配置不良則會難以找到連結，以至於需要花費更多時間找尋資料。任何網站架構及呈現方式都可能不符合使用者的需求，網站若能夠依照使用者瀏覽的特性調整，作出合適的引導，即能

夠降低資訊查詢的成本，維持使用者對網站的滿意度。

現在已經有許多網站會針對使用者進行不同的調整，來達到使用者的需求，其中知名的網站為 Amazon，Amazon 對顧客以往的點選、購物等瀏覽記錄進行分析，接著透過其本身的推薦系統進一步推測使用者可能有興趣的資訊，並呈現在網站中，此推薦系統也為 Amazon 增加不少營收，使 Amazon 在 2012 年第二季財務總營收相較於去年同一時期成長了 29%。呈現適合的資訊給使用者能夠提升瀏覽效率，為此網站需要能夠進行調整，調整可以由使用者決定，也可以由網站進行，舉例來說，Google 為了讓資訊更符合需求，讓使用者可以決定如何調整，提供選擇自己所需工具及資訊的功能，並能夠依喜好拖拉區塊來調整呈現，雖然這些做法能夠明確知道使用者的需求，但在一般網站中，大多數的使用者會跳過這個階段，直接找尋所要的資料[11]，因此會使用適性化網站(adaptive web site)從使用者瀏覽模式學習，自動改善網站，並呈現符合使用者需求內容[7]，調整型態分為兩種，第一種為客製化(customization)，針對個別使用者進行網頁的調整，經常使用在入口網站及電子商務，第二種為最適化(optimization)，網站調整本身讓整體使用者瀏覽時能夠有更好的引導。

網際網路分為三類：資訊提供型(informational)、網路交易型(transactional)及網路作業型(operational)[3]，網路的快速及便利使許多作業電子化並轉移到網路上進行，也促使網路交易蓬勃發展，但目前多數公司及公家機關的網站仍然屬於資訊提供型，民眾瀏覽網站只為了找尋特定的資料，而非長時間瀏覽，且較少擁有登入的需求，當網站沒有登入資訊時，區分使用者最常使用的方式為 IP 位置[9]，所記錄的使用者資訊較不精確，無法提供適合的客製化(customization)調整；另外，資訊提供型網站的使用者人數多但瀏覽頻率並不高，大多使用者並沒有足夠的瀏覽記錄來分析，若採用客製化調整，則無法讓使用者瀏覽時快速找到資訊。

資訊提供型網站會蒐集、彙整並管理特定資訊，並提供給使用者，雖然網站的功能簡單，純粹陳列資訊出來，仍然會不符合使用者需求，造成這種情況的原因可能是因為網站最初設計不當，或是使用者的需求改變，最常見到的處理方式是提供動態表單讓使用者能夠快速連結到熱門頁面。利用適性化調整能夠讓

網站自動找出適合推薦的網頁，但網站每次的調整皆為獨立事件，且調整網站對使用者瀏覽狀況所能造成的影響皆為預測。

增強式學習(reinforcement learning)起源於心理學，解釋動物及孩童的學習過程，學習中會與環境不斷的互動，找到正確的行為，如同嬰兒在學習走路時，並沒有人教導如何站立、如何移動，藉由與環境的回饋訊息，例如跌倒，便慢慢的探索出要怎麼使力才能夠站立及走路。由此可知增強式學習屬於非監督式學習，不需要有指導者來監督及教學，透過試誤(trial-and-error)的機制，僅憑著與環境的互動進行學習，學習要素中包含環境(environment)及代理人(agent)兩個項目，環境代表需要改變的問題，代理人則負責學習，代理人下達動作(action)命令後，會改變環境的狀態(state)，不同狀態之下，代理人會依照獎懲值(reward)來調整下一次的動作[14]，學習過程中，每次的行動都會影響到後續的狀態及行動，在經過一連串的學習後，能夠找出符合現況的最佳策略。

本研究以學術網站為實驗目標，網站類型屬於資訊提供型，因此本研究採用最適化型態，以網站整體使用者的需求為目標來調整，由於沒有使用者確切資訊，我們利用網頁探勘(web mining)來分析由網站自動紀錄的日誌檔(log)，找出使用者的瀏覽特性，雖然最適化調整所呈現的網頁無法完全符合所有使用者的需求，但可以讓網站整體的表現更好，使首次瀏覽或不熟悉網站架構的使用者能夠有效率地找到資訊。研究內容中，網站的調整不只以單次資料來進行運算，而是能夠將每次的調整連結起來，從中學習如何調整，因此使用能夠與環境互動的增強式學習，讓網站得知每次調整所造成的影響，進而學習到如何進行調整。總結來說，本研究使用增強式學習讓網站擁有類似人類的學習能力，透過與環境互動瞭解網頁瀏覽狀況的改變，並將每次調整以及其影響加入考慮中，目的在於使網站進行最適化推薦調整，以降低使用者找尋資料的時間。

總結本研究之目的分為以下四項：1.透過伺服器日誌檔，分析整體使用者的瀏覽習慣及特性。2.利用增強式學習使網站透過過去調整經驗學習，並依照使用者需求推薦。3.透過瀏覽推薦減少使用者找到目標資訊所需要瀏覽的路徑長度。4.瞭解網頁各瀏覽特徵之重要性。

本論文共分為六章，依序內容如下：(1)說明研究背景與動機及目的。(2)探討網頁探勘、適性化網站及增強式學習的相關文獻。(3)介紹

研究系統架構，並闡述資料處理過程，定義增強式學習中各個參數，說明調整及呈現方式。
(4)陳述本研究系統之模擬運作情形(5)總結。

2. 文獻探討

本章將會介紹研究中所使用到的技術，包含網頁使用探勘、適性化網站以及增強式學習，並從探討過去研究中，確立本研究之方法。

2.1 網頁使用探勘(Web Usage Mining)

網路快速及便利的特性，讓許多企業朝網路發展，在無法面對面觀察使用者的情況下，很難知道顧客的需求，透過網頁使用探勘，自動從網頁伺服器中追蹤使用者的瀏覽行為（例如：滑鼠點擊），並加以分析使用者特性，藉此提供適當的資訊給顧客，因而拉近雙方的距離。網頁使用探勘最廣泛應用在商業網站中，透過使用者的點擊、購買紀錄等，個人化推薦適合的商品，以提高購買率，而藉由分析顧客瀏覽資料，還能了解促銷活動的成效、目前市場的趨勢等影響公司策略的資訊，甚至找到新的客戶。另外，分析出的使用者瀏覽特性能做為改善網站的依據，讓網站內容更有效率地呈現，使得公司與顧客之間的互動更為順暢。

網頁使用探勘包含三個階段[13]：前置處理(preprocessing)、特徵發覺(pattern discovery)及特徵分析(pattern analysis)。探勘流程最初會從網站伺服器中取出日誌檔，將日誌檔資料進行前置處理，處理過程為資料過濾、資料清理、使用者定義、Session 定義，抓取的資料若不符合分析需求的記錄會將之刪除，每筆資料只留下所需要的部分，其餘將予以清除，以便之後的分析，保留完所需的資訊後，定義區分不同使用者的條件，接著定義 Session 的分界點，一般以 30 分鐘為間隔[2]，以區分不同次的瀏覽紀錄，藉此從分析日誌檔資料中獲得使用者整個瀏覽過程，而非單一筆要求網頁的記錄；接著，處理完的資料經過統計分析，找出瀏覽規則及特徵，最後從分析出的特徵中，找出所需要的資料。

資料來源

日誌檔(log)是網頁使用探勘主要的資料來源，伺服器在使用者瀏覽全球資訊網上瀏覽網頁時，將瀏覽及存取動作紀錄在日誌檔中，伺

服器預設的日誌格式為NCSA的ASCII一般日誌格式(Common Log File Format)，紀錄資訊包含IP、主機名稱、使用者名稱、時間、存取檔案名稱及大小等，格式為[userIP][timestamp][method][url][httpversion][httpresult][size]；另一種常用格式為擴充日誌檔格式(Extended Log File Format)，此種格式增加紀錄使用者帳號、Cookie、瀏覽器版本以及主機的作業系統等[4]，可以自訂要記錄的欄位，選擇重要的欄位而忽視不需要的欄位能夠控制日誌的大小，節省不必要的空間及減少後續的處理。

探勘技術

網頁使用探勘過程中最重要的是分析階段，必須從眾多資料及記錄中找出使用者的瀏覽特性，在特徵發覺(pattern discovery)階段有許多不同的分析方式，常用的技術為統計分析(Statistical Analysis)、關聯法則(Association Rule)、分類分析(Classification)、群集分析(Clustering Analysis)、序列型樣(Sequence Pattern)[13]。

2.2 適性化網站(Adaptive Web Site)

內容資訊豐富的網站，頁面之間的鏈結往往也越趨複雜，若能針對使用者的特性及需求提供不同的資料及呈現方式，可以有效提升瀏覽效率，讓使用者能夠更快速找到所需要的資料，進一步提升使用者對網站的滿意度。適性化網站能夠從使用者存取網頁的特性學習，自動改善網站的組織及呈現[7]。

適性化流程分為兩個階段，第一階段為行為觀察，將紀錄透過網頁使用探勘找出使用者瀏覽時的特性，並將其特性模組化，以便往後的資料能夠快速比對；第二階段為網站調整，通常會事先寫出多個網頁呈現模組，再依照瀏覽特性找出對應的網頁模組並呈現給使用者[8]。在第二階段依網站調整目標不同又分為兩類：內容適性化(content-level adaptation)及鏈結適性化(link-level adaptation)[1]。內容適性化又稱為適性化呈現(adaptive presentation)，主要針對網站頁面呈現做調整，目的是給予不同種類的使用者差異化的內容呈現，其中內容呈現又細分為文字及多媒體適性化，目前多以文字適性化為主；鏈結適性化又稱為適性化瀏覽支援(adaptive navigation support)，主要是針對網頁之間的鏈結來做調整，目的在於提供網站導

覽避免使用者在網站中迷失，能夠快速找到所需資訊。

網站的適性化調整又可以依照調整方式的不同分為兩個方向[7]，第一種是個人化(customization)，網頁立即調整以符合個別使用者的需求，使用者可以依照自己的瀏覽方式選擇呈現的畫面，或是由網站從過去的瀏覽記錄中，自動預測使用者的下一步及目標，提供導覽讓瀏覽能夠更快速；第二種是最適化(optimization)，最適化調整不會為個別使用者做調整，而是從所有使用者的資訊來調整網站，讓整個網站更容易使用，舉例來說：當最常觀看的頁面離首頁相隔很遠，就應該建一個連結直達此頁，當重要訊息公布時，將訊息以較大的版面呈現來提高使用者的注意，將網站以整體使用者的需求來做調整，能夠幫助新進使用者瀏覽。

2.3 傳統網頁探勘適性化網站的優缺點

網站可能遇到的問題有以下四種：第一，不同的使用者有不同的目標，第二，同一個使用者依時間的改變有不同的需求，第三，原始設計不再適合目前的網站，最後，網站的使用方式與原本預計的不同[7]。使用網頁探勘能夠得知目前網站的瀏覽狀況，從眾多資料及記錄中，找出使用者真正的需求，透過適性化調整讓網站進行改善，來解決以上的問題。除了讓網站提供符合需求的資訊之外，適性化網站自動調整的特性也能夠減少人員參與。

傳統的適性化網站雖然能夠得知使用者需求，並依此調整呈現，但網站的調整為一次性，每次的調整為獨立事件，彼此之間無法互相影響，其中調整的條件及方式皆由設計者訂定，如此會造成調整不當的情況，而想要讓網站能夠正確改善，調整過後所造成的影響，依然需要人員來判斷是否合適，並適時的修正調整。為了改善傳統適性化網站的問題，因此調整時需要加入機器學習，讓網站透過每次的學習，來找出最適合的調整。

2.4 增強式學習(Reinforcement Learning)

增強式學習源自於心理學的實驗，實驗內容為訓練鴿子啄擊綠色按鍵，每當啄擊綠色按鍵時會給與飼料，飼料為實驗中的增強物，藉由這個訓練，鴿子可以學習到辨別按鈕，進階實驗讓飼料只有在燈亮時啄擊才會掉落，鴿子

可以學到何時應該啄擊[12]，將此實驗轉換到增強式學習的架構中，鴿子代表代理人，是否亮燈為狀態，飼料為回饋，經過不斷的訓練學習，能夠讓鴿子學會複雜的行為。但在學習過程中，鴿子可能會有與學習目標無關的行為，例如：燈亮時，鴿子邊拍翅膀邊啄擊綠色按鍵，由於達到學習目標，因此鴿子得到飼料，鴿子會誤以為拍翅膀也是需要的行為，以至於後來的學習會不斷重複這個無意義的動作。

增強式學習模擬生物學習的方式，透過失敗吸收教訓，再反覆練習後強化自身能力，是一種非監督式學習(unsupervised learning)，在學習過程中只需要輸入資料，代理人會自動學習調整，由於代理人會依據累積的經驗修改或選擇下一次的行動，因此適用於動態環境。增強式學習在未知環境下，不斷嘗試不同行動，在與環境互動中，利用試誤(trial-and-error)和延遲獎懲(delay reward)的機制來找尋最佳策略的一種學習方法[14]，其架構如圖 1 所示。

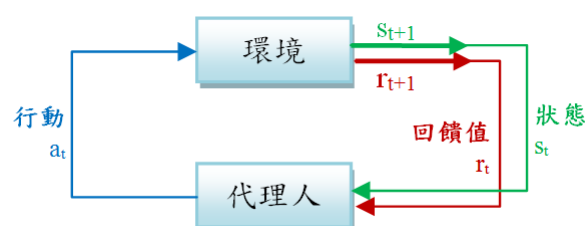


圖 1 增強式學習架構

增強式學習中包含兩個角色，一個是代表需要改變的環境，另一個是負責學習的代理人，在學習過程中代理人在時間點 t 收到環境狀態(s_t)，代理人依據狀態選擇行動(a_t)並送給環境，環境採取行動後會改變目前狀態(S_{t+1})，狀態的改變會依照對目標的影響產生獎懲值(r_{t+1})，並由環境傳回代理人，代理人會根據收到的新狀態(S_{t+1})和獎懲值(r_{t+1})來修改價值函數(value function)，並依照函數變化再次為新的狀態選擇行動，繼續循環以上的步驟。學習過程中，每次行動所產生的變化都會做為後續選擇時的依據，藉由學習經驗的累積，得知不同狀況下所應該採取的最佳行動，進而接近目標。

增強式學習分為主動式(active)與被動式(passive)兩種，被動式增強式學習的代理人用固定策略去學習，主動式學習則需要不停探索環境來找到適合的策略。本研究利用主動式增強式學習中 Q-Learning 演算法，Q-Learning 是由 Watkins 在 1989 年提出的一種無策略演算

法，其演算法建立在被動式演算法 TD 方法基礎之上，藉由試誤及延遲獎懲不斷計算 Q-value (也稱為 state-action-value)，Q-value 為狀態與行動對目標的比較值，值越大代表達到目標的機會越大，代理人依過往經驗來選擇最佳的行動[17]，其公式如下[14]：

$$Q(s, a) \leftarrow Q(s, a) + \alpha(R(s) + \gamma \max_{a'} Q(s', a') - Q(s, a))$$

s 代表目前的環境狀態，a 為代理人的行動，s'是下一時間點的狀態，a'是下一時間點的動作， α 為學習函數， γ 為獎懲的折扣值，R(s) 代表狀態 s 下的實際價值，Q(s, a)代表在狀態 s 下進行行動 a 的期望價值， $\max_{a'} Q(s', a')$ 代表下一時間所有 Q-value 的最大值，Q-Learning 必須計算下一次行動中最大的 Q-value。

2.5 增強式學習應用於網站之相關研究

網站中使用者瀏覽習慣及需求會不斷變動，藉由增強式學習能與環境互動的特性，能夠讓網站隨著環境的變動來調整後續動作，因此過去有許多學者將增強式學習應用在網站的調整中，如表 1，其中透過增強式學習從使用者瀏覽情況來調整的呈現方式，多數研究以推薦為學習結果的呈現方式。

表 1 增強式學習之適性化網站相關研究

學者	目標
Rojanavasu et al. (2005)	在目前商品頁面中，預測使用者會購買的相關商品[10]
Taghipour et al. (2007)	於使用者目前瀏覽路程，預測下一個瀏覽頁面[15]
Mahmood et al. (2007)	在不同種類頁面之下，學習網頁適合推薦的情況[5]
Mahmood et al. (2009)	在不同種類頁面及使用者動作的情況下，學習對應的動作[6]

由過去文獻中顯示，增強式學習之推薦研究皆以商業性網站為實驗對象，各網站擁有會員機制，進而對顧客進行個人化推薦，但在學習過程中，缺少時間性，以 Rojanavasu et al. (2005)的研究為例，代理人無法判斷是否到達冬天，來推薦保暖商品，只能透過使用者點擊了圍巾之後，才進一步推薦手套，無法以環境傳回的狀態來判斷出時間性的推薦。

3. 研究方法

本研究將利用增強式學習來調整網站中的推薦，以利使用者能夠快速找到目標資訊。網站採取最適化調整，以整體使用者為目標，統整網站中所有使用者的瀏覽紀錄，以提供適性化的推薦，讓不熟悉網站架構的使用者減少找尋目標資訊的時間，也能使新的使用者對網站更容易上手，如此可以提升使用者對網站的滿意度；另外，本研究除了最常使用的網頁瀏覽次數及瀏覽時間之外，還增加了三個輔助資訊，分別為到達網頁路徑長度、網頁所在階層以及網頁目前排名，減少路徑為本研究的目標，因此加入學習之中，而後兩項參數則使網站學習能夠將先前的瀏覽及調整狀況加入考量中。本研究系統主要分為三部分：網頁瀏覽特徵萃取、網頁推薦權重學習及推薦表單呈現調整。

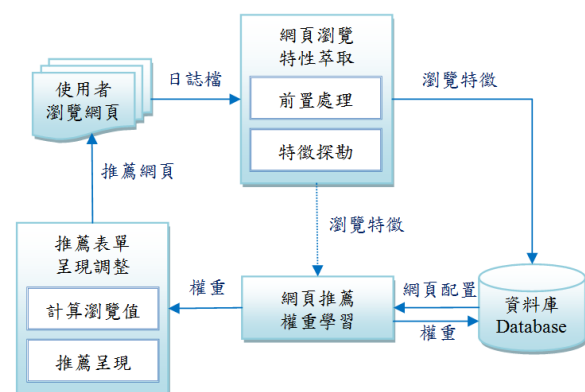


圖 2 研究系統架構圖

圖 2 為本研究之系統架構，使用者在瀏覽網站時，Apache 會自動將使用者所發出的要求記錄成日誌檔，為了分析使用者的瀏覽情況，首先會將此伺服器日誌檔擷取出來，日誌檔經過前置處理會過濾掉不必要的資料，我們將此視為原始資料並記錄到資料庫；接著透過資料探勘找出整體使用者的瀏覽特徵，一樣將此存入資料庫中。從資料庫抓取出使用者特徵後，透過增強式學習，系統會學習到調整推薦需要改變的值，將此加入計算中，各個網頁都會獲得一個瀏覽值(surf value)，網站中的網頁推薦會依學習後的瀏覽值來做調整，最後呈現給使用者，使用者再次瀏覽所產生的日誌檔會再次進行計算，將推薦的網頁不斷調整到符合多數使用者的需求。

3.1 網頁瀏覽特徵萃取

本研究實驗中，前置處理首先會對抓取的資料進行基本的整理，包含資料擷取、資料切割、資料過濾、資料清理及資料格式化，接著定義使用者及 Session，並將處理完的資料存入資料庫。經過上述步驟，可以從日誌檔中了解使用者的特性以及網頁的架構，得知使用者所需要的功能或資訊，及找尋資料時所遇到的困難，網頁即可藉此調整架構，改善網頁的呈現。

(1) 資料切割

由於直接處理過大的日誌檔容易發生多餘的處理，造成分析時間變長，因此截取日誌檔之後，首先將資料切成適合分析的大小以利分析。本研究目的要找出目前網站使用者的瀏覽趨勢，分析時段間隔越小越好，但實驗網站並非大型網站，單一一天的資料量並不足以作為分析，因此以三天為切割單位。

(2) 資料過濾

資料過濾會將研究中不需要的資料過濾掉，本研究實驗所使用的網頁是由 PHP 所撰寫，為了瞭解使用者瀏覽網頁的路徑，我們截取日誌檔中要求 php 檔的記錄，其他如圖檔 (*.jpg、*.png)或執行檔(*.js)等將予以清除，並將功能性的網頁過濾掉，只保留含有網頁本體的記錄(含有 index.php)。由於伺服器中架有多個網站，必須刪除其他資料夾的記錄，僅僅保留此網站頁面的記錄(只有/index.php)。而目前 PHP 網頁多由參數來傳遞，而非網頁之間直接連結，實驗網頁也採用依參數來呈現網頁的方式，因此在過濾完不必要的頁面之後，將傳遞的參數視為網頁，以便接下來的分析。

(3) 資料清理

對於資料過濾後的日誌檔，透過資料清理將記錄中不需要的部分刪除，僅留下後續研究中會使用到的欄位，本研究主要針對瀏覽頁面來做分析，分析前會依瀏覽的日期、時間來進行整理，找出同一次瀏覽網頁的路徑，所以後續分析會使用到的資料欄位為使用者 IP、timestamp 以及 url。

(4) 資料格式化

為了方便將使用者資料匯入資料庫，資料格式化的步驟會將得到的資料轉換為資料庫所需要的格式，我們將資料中 timestamp 中的日期及時間分成兩個欄位，在 url 中由於都是同一個頁面(index.php)，因此用傳遞的參數來代替，並且在每個欄位之間加入"-"來區隔，方便接下來處理及匯入資料庫的動作，格式呈現為[userIP]-[date]-[time]-[url]。

(5) 使用者及 Session 定義

使用者定義一般有以 Cookie、使用者帳號及 IP 等方法來定義，但公告性網站的使用者不需登入即可查詢所需資料，在沒有會員制度的情況下，本研究取得的資料有先天上的限制，所以採用將同個 IP 視為一個使用者的方式，並以 30 分鐘做為 Session 的分界點，利用使用者 IP 及網頁間瀏覽的間隔時間來整理日誌檔，將同一次瀏覽紀錄整理一起。以使用者 IP 做為使用者定義可能會忽略掉的情況，但本研究主要是分析整體使用者的瀏覽行為，此階段目的在整理出單次瀏覽中經過的頁面，不同 IP 同一個使用者及不同使用者同一個 IP 的情形對結果的影響並不大。

(6) 擷取網頁參數

本研究會將網站中各個網頁的重要程度進行衡量，在衡量中會考量使用者的瀏覽時間及網頁的瀏覽次數，藉此資訊瞭解各網頁的重要性，並找出當時網站中使用者的目標網頁。在此階段會計算出各個網頁在期間內的點擊次數、平均點擊次數、停留時間、平均停留時間，網頁停留的時間為此次要求時間與下一次網頁要求時間的間隔，而最後要求的網頁停留時間則採用所有網頁的平均停留時間；此階段同時會計算出網頁的到達路徑及平均到達路徑，網頁的到達路徑為當次瀏覽所發出的第一個網頁要求到目前的頁面之間總共要求過幾個網頁，因此第一個網頁的路徑長度為 0，路徑長度的增加或減少會做為之後增強式學習中網站狀態與動作配對的回饋值。以上網頁的各項參數經過計算後，會獲得網頁的瀏覽值(surf value)，做為最後調整時的依據。

(7) 網頁資料庫

在網頁探勘中，會將資料匯入資料庫，資料表記錄著分析過後的資料(如表 2)，欄位包含瀏覽網頁、網頁點擊次數、停留時間、平均停留時間、所經過的路徑頁數及平均路徑頁數。

表 2 網頁熱門程度衡量表

欄位名稱	說明	範例
Page	目標網頁 URL	tid=238
Count	網頁瀏覽次數	25 (次)
Total_StayTime	網頁瀏覽時間	2250 (秒)
Avg_Stay	平均瀏覽時間	90 (秒)
Total_Path	到達網頁路徑長度	130 (頁)
Avg_Path	平均到達路徑長度	5.2 (頁)

3.2 網頁推薦權重學習

分析網站採用的衡量指標有內容指標及商業指標，而本研究的目標網頁與商業交易無關，因此採用網站內容指標來衡量，衡量網站需要瞭解使用者與網站內容的互動情況，一般會追蹤進站與離站的網頁、瀏覽各別網頁的次數、瀏覽各別網頁的時間長度等資料，每個衡量指標能夠知道網站的不同特性，但彼此之間是無法比較，目前不能肯定哪項指標是最佳的衡量的方式，因此，本研究希望透過自動學習，讓網站能夠學到各資料在排名計算中最適合的權重，增強式學習讓網站不是一次就決定權重配置，而是經過長期學習，參考以往權重調整後對瀏覽狀態的影響，進行之後的調整。

追蹤資料除了點擊次數及瀏覽時間之外，我們增加了到達網頁的路徑長度、網頁目前所在的層級以及網頁目前的排名來當作計算排名時的輔助資料，當兩個網頁之間的瀏覽次數及停留時間相近時，我們推測到達路徑長度比較長的網頁優先推薦能夠使整體平均瀏覽長度減少較多，在底層的網頁也因為相同原因會比高層的網頁優先推薦，而網頁單次的點擊次數及瀏覽時間減少並不一定代表它的重要性降低，因此先前的排名也會加到學習中。

本研究使用增強式學習的模式中，環境代表網頁的瀏覽狀況，環境會將先前由網頁探勘所分析出的資料傳給代理人，讓代理人知道目前網頁的狀態，根據同時由環境傳回的回饋值，衡量前一次調整對網頁表現的影響，用

SARSA 演算法計算其 Q-value 值，接著更新學習資料，最後依照新的狀態採取合適的行動，Q-value 值越大代表選到該行動的機率越大，在學習到各參數的權重後，透過推薦機制會將所有網頁進行排名，最後，系統直接抓取排名在前面的網頁來推薦，並呈現在動態表單中。由於一天的資料量有限，分析出的特徵變化並不明顯，因此本研究網站學習以三天為間隔，處理累積的資料。

SARSA 演算法公式：

$$Q(s, a) \leftarrow Q(s, a) + \alpha(R(s) + \gamma Q(s', a') - Q(s, a))$$

本實驗將 SARSA 演算法減化，設 α 及 γ 為 1：

$$Q(s, a) \leftarrow R(s) + Q(s', a')$$

本研究網站於學習網頁參數權重階段，使用增強式學習的流程如下，並針對學習中的狀態、行動以及回饋值的設定進行詳細解說。

1. 初始推薦
-為了不讓最初的效率太差，會先計算第一次排名來推薦
2. 環境傳回新推薦所造成的狀態
-狀態為各網頁的點擊次數、停留時間、路徑長度、目前層級、目前排名
3. 環境傳回狀態與行動配對的回饋值
-回饋值為過去路徑長度減去目前路徑長度
4. 重新計算增加各參數之權重的 Q 值
5. 代理人根據 Q 值選擇增加權重的參數
-當 Q-value 相同時，隨機選取
6. 推薦機制依照行動所調整的權重來計算排名，並推薦
7. 回到第二步驟

圖 3 增強式學習推薦流程

(1) 參數設定

狀態：目前網站中各網頁的瀏覽狀況。

本研究所追蹤的網頁狀態資料包含瀏覽次數、瀏覽時間及到達網頁的路徑長度，為了讓資料以相同標準比較，會將此三項資料正規化，每個資料所代表的值為此網頁佔全部網頁中的比例，以瀏覽次數來說明，在狀態中的資料為單一網頁瀏覽次數除以總網頁瀏覽次數所得到的值；代表狀態的另外兩項參數為先前此網頁所在的階層及排名，為了方便計算，一樣會進行正規化。

表示方式 State(Count, Time, Path, Level, Rank)

行動：代理人選擇要增加權重的參數。

代理人能夠選擇的動作有每個行動針對不同參數的權重進行調整，為了讓網站最初表現不要太差，會先訂定各個參數的初始權重，瀏覽次數及瀏覽時間最常作為網路評估的參數，因此預設的權重較高，而後三項作為輔助資料的參數其初始權重會較低。選擇行動中，對於各項參數權重的調整皆為固定的值，當選擇增加某一參數時，另外四種參數將會同時減少對應的值，讓所有權重的加總保持在一定的值。每次行動所選擇的參數權重增加 4，其餘參數的權重皆減少 1，讓權重加總維持在 100。表示方式 $Action(W_{Count}, W_{Time}, W_{Path}, W_{Level}, W_{Rank})$

回饋值：到達目標網頁的平均路徑長度變化。

本研究的目標是要減少使用者找尋目標所需要經過的頁面，因此我們以網頁到達路徑的長度變化，作為評估此次狀態與行動配對的標準，計算到達路徑長短的網頁，只侷限在目標網頁，其餘的網頁並非我們所要減少的目標。回饋值的計算是由過去平均網頁到達路徑長度減去此次網頁到達路徑長度，也就代表路徑長度增加，意味著此次調整的權重所帶來的影響不好，反之亦然。

表示方式 $Reward(TotalPath_{past} - TotalPath_{now})$

(2) 參數權重學習演算法

權重學習階段我們會抓取各網頁的瀏覽次數、瀏覽時間、到達網頁的路徑長度、以及網頁所在的階層和排名五項參數作為輸入值，並依照以下之演算法學習。

Reinforcement Learning Recommendation Algorithm

1. $TotalPath_{now} \leftarrow \sum Path$
2. $R \leftarrow TotalPath_{past} - TotalPath_{now}$
3. Update Q
4. $S \leftarrow Count, Time, Path, Level, Rank$
5. Match S in RLmatrix
6. Select Max Q
7. Return A

圖 4 參數權重學習演算法

在系統每次更新推薦中，首先將所有網頁的到達路徑長度加總，獲得總目標網頁路徑長

度，並與前一次之總路徑長度相減，得知路徑長度變化並將此視為獎懲值，以作為前一次調整之評估，透過此獎懲值帶入增強式學習之公式，再次更新 Q 值。接著以各網頁的瀏覽次數、瀏覽時間、到達網頁的路徑長度、以及網頁所在的階層和排名作為目前網站的狀態，並在增強式學習之學習矩陣中比對出相同的狀態，從中選出 Q 值最高的動作來回傳並執行。

系統在執行動作後會改變目前各項參數之權重，並以此權重調整推薦網頁，於三天後分析網頁以獲得五項參數值，並再次執行此演算法。

3.3 推薦表單呈現調整

系統追蹤使用者瀏覽網頁的情況，透過網頁探勘分析出網頁的瀏覽次數、瀏覽時間及到達網頁的路徑長度，並從資料庫中抓取各網頁在網站中所在的階層以及目前的排名等配置狀況，經過增強式學習獲得以上五個影響網頁重要程度的參數之權重，將探勘所分析出的資料與學習所獲得的權重放入推薦機制中，求出各個網頁的瀏覽值(surf value)，接著網站依照瀏覽值將所有網頁進行排名，最後根據排名將推薦網頁呈現在首頁的動態表單。

(1) 瀏覽值(Surf Value)

瀏覽值計算所使用到的參數為網頁瀏覽次數(Count)、網頁瀏覽時間(Time)、到達網頁路徑長度(Path)、網頁目前所在階層(Level)網頁目前排名(Rank)五個項目及其各自的權重，由於各個參數皆經過正規化，因此計算時不再進行其他處理，直接將各參數的數值與權重相乘並加總，即獲得此網頁的瀏覽值，公式如下。由於參數值及權重沒有經過其他的處理，因此權重的變化也代表各個參數的重要性。

瀏覽值公式如下：

SurfValue=

$$Count \times W_{Count} + Stay \times W_{Stay} + Path \times W_{Path} + Level \times W_{Level} + Rank \times W_{Rank}$$

(2) 調整推薦

網站中所有網頁經過計算後會得到各自的瀏覽值，接著網站會依此瀏覽值排名，並選出推薦的網頁呈現在動態表單，推薦分為兩個階層，首先，第一階層推薦的動態表單會呈現

在首頁，系統先將首頁以及其下一層的網頁從排名中過濾，再抓取剩下排名的前三名網頁給予連結，以圖 5 解釋，第一階層推薦網頁為紅色圓圈，所能選擇的範圍於紅色虛線方框中的所有網頁；接著第二層推薦會放在首頁的下一層，為了不使網站原本架構錯亂，在處理過程中，只留下與目前網頁同一類別的網頁，並再次依照排名取出前三名進行推薦，在圖 5 中，我們以顏色區分不同類別，以類別 B 為例，在 B1 頁面所能推薦的網頁為藍色虛線方框內的網頁。

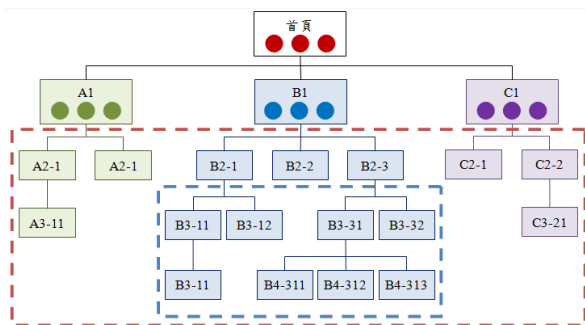


圖 5 推薦階層示意圖

(3) 推薦呈現

本研究實驗目標網站為學校系上網頁，以原有的網站為基礎進行改善，由於使用者不喜歡不斷變動的畫面[16]，因此我們希望不要破壞原本的網站，採取動態表單來呈現推薦結果，並於首頁之中選擇明顯的位置來呈現，如圖 6 所示，以最少的改變來改善整個網站瀏覽效率。



圖 6 網站首頁之推薦配置

4. 實驗

本研究以國立高雄大學資訊管理學系系網頁為實驗目標，以其使用者瀏覽日誌檔作為資料來源，該網站為學術網站，以公布訊息為主要目的，日誌檔格式為 NCSA 的 ASCII 一般日誌檔格式。本研究採用伺服器運作以來的歷年資料，為西元 2008 年 5 月開始至 2012 年 9 月，該紀錄的原始檔案大小為 2.67GigaByte。

本研究以日誌檔紀錄模擬系統運作，以下為網站一月份之模擬情境。

(1) 系統模擬運作(一)：瀏覽次數增加

一月初碩士班學生由於論文計劃書提案報告接近，增加了「計劃書報告網頁」相關資訊頁面的瀏覽次數，期間內，瀏覽次數由 5 次變為 20 次，平均瀏覽時間一樣是 5 分鐘(300 秒)，平均到達路徑長度為 5.5 頁；網站的網頁數量為 120 頁，總瀏覽次數為 600 次，平均每頁為 5 次，平均瀏覽時間為 2.5 分鐘(150 秒)，目標網頁平均到達路徑為 3.5。經過標準化後環境所傳回的狀態為 S (4, 2, 1.57, 0.5, 0.1)。

各參數原本的權重分別為 46, 42, 8, 2, 2，在此狀態下代理人依照 Q 值選擇增加瀏覽次數的權重，因此瀏覽次數的權重增加 4，其餘權重皆減少 1，各參數的權重改變後依序為 50, 41, 7, 1, 1，經過推薦機制計算出的瀏覽值為 294.39。

SurfValue=

$$4 \times 50 + 2 \times 41 + 1.57 \times 7 + 0.5 \times 1 + 0.9 \times 1 = 294.39$$

比較各網頁的瀏覽值，「計劃書報告網頁」在所有網頁排名第二，網站篩掉首頁及第一層網頁後，依然排名第二，系統從排名中直接抓取前三名，因此「計劃書報告網頁」也包含在內，新的前三名直接取代先前推薦的網頁，放到首頁的動態列表中。

調整後經過三天會再次進行學習，「計劃書報告網頁」平均到達路徑縮短為 2.5 頁，瀏覽時間不變，所有目標網頁的平均到達路徑長度由 3.5 變為 3.2，平均瀏覽時間為 2.7，網站傳回回饋值為 0.3 (Reward = 3.5 - 3.2 = 0.3)。

由於路徑減少，計算出的回饋值為正，代表此次的調整是合適的，代理人會重新計算 Q 值，並依照網站所傳來的新狀態選擇下一次的行動。

(2) 系統模擬運作(二)：瀏覽次數減少、停留時間減少

演講行事曆會在學期開始之前公布，在開學階段，由於學生需要查詢整個學期的行程，「演講網頁」的瀏覽次數會增加，由於行事曆非可下載檔案，因此停留時間也會增加，在學期中為了查詢演講資訊或確認變動，「演講網頁」依然保持其重要性，但到了寒假時，整個學期的演講已經結束，演講網頁的瀏覽次數會急速減少，停留時間也會減少。

於1月中時由於進入寒假，網頁瀏覽次數由50次變為2次，平均瀏覽時間為20秒，平均到達路徑為3頁；而網站總瀏覽次數為480次，平均每頁為4次，平均瀏覽時間為2分鐘(120秒)，目標網頁平均到達路徑為3，經過標準化後環境所傳回的狀態為S(0.5, 0.17, 1, 0.83, 0.98)。

各參數原本的權重分別為50, 41, 7, 1, 1，在此狀態下代理人依照Q值選擇增加瀏覽時間的權重，因此瀏覽時間的權重增加4，其餘權重皆減少1，權重改變後依序為49, 45, 6, 0, 0，經過推薦機制計算出的瀏覽值為120.5。

$$\text{SurfValue} = 0.5 \times 49 + 0.17 \times 45 + 1 \times 6 + 0.83 \times 0 + 0.98 \times 0 = 120.5$$

比較各網頁的瀏覽值，「演講網頁」在所有網頁排名20，網站篩掉首頁及第一層網頁後，並不在前三名；接著在第二層推薦，選出與演講網站同類別的頁面，篩選掉動態表單呈現的網頁以及其下一層的網頁之後，仍然不在此類別前三名，因此系統將不會推薦此網頁在任何階層，演講網頁的階層回到第3層。

經過三天後，演講網頁的瀏覽次數以及瀏覽時間沒有變化，平均路徑從3頁增加為5頁，但由於所有目標網頁的平均到達路徑長度由3頁減少為2.5頁，因此判斷此次權重調整為適合的，回饋值為0.5 (Reward = 3-2.5 = 0.5)。

接著代理人會再次用SARSA演算法計算Q-value值，再次依照新的狀態採取合適的行動，並繼續調整網頁推薦。

5. 結論

本研究提出以增強式學習為基礎的網頁瀏覽推薦方法，網站依據使用者的瀏覽特徵以及網頁配置進行推薦，考量多因素的影響，其

中包含網頁瀏覽次數、網頁瀏覽時間、到達網頁所需之路徑長度、網頁目前所在階層、以及網頁目前排名五項參數。透過分析日誌檔萃取出使用者特性後，網站使用增強式學習調整五個參數的權重，並自動評估調整結果，另外，透過增強式學習與環境互動也將時間性考量到本研究之推薦方式中，最後將各項參數值及其相對權重進行運算以推薦網頁，減少網站中整體使用者找尋資訊所花的路徑長度。本研究僅取出五項參數作為調整依據，往後的發展可考慮不同的參數來進行學習推薦。另外，本研究之網站推薦為最適化，未來可針對具有登入機制之網站，使用本研究所提出的方式進行個人化推薦。

參考文獻

- [1] Brusilovsky, P., "Methods and Techniques of Adaptive Hypermedia." *User Modeling and User-Adapted Interaction Journal*, 6, pp. 87-129, 1996
- [2] Catledge, L. and Pitkow, J., "Characterizing Browsing Behaviors on the World Wide Web." *In Proceedings of Computer Networks and ISDN Systems*, 27(6), 1995
- [3] Koh, C.E. and Balthazard, P., "Electronic Commerce and the World Wide Web: a Framework of Business Web Use and a Study of Business Web Practices." *In Proceedings of Decision Sciences Institute*, San Diego, CA, pp. 599-601, 1997
- [4] Kosala, R. and Blockeel, H., "Web Mining Research: A Survey." *ACM SIGKDD Explorations*, 2 (1), pp. 1-15, 2000
- [5] Mahmood, T. and Ricc F., "Learning and Adaptivity in Interactive Recommender Systems." *In Proceedings of International Conference on Electronic Commerce, Minneapolis, USA*, pp. 75-84, 2007
- [6] Mahmood, T. and Ricc F., "Improving Recommender Systems with Adaptive Conversational Strategies." *In Proceedings of the 20th ACM conference on Hypertext and hypermedia*, pp. 73-82, 2009
- [7] Perkowski, M. and Etzioni, O., "Adaptive Web Sites: an AI Challenge." *In Proceedings of International Joint Conferences on Artificial Intelligence*, Nagoya, Japan, 1997

- [8] Perkowitz, M. and Etzioni, O., "Adaptive Web Sites: Automatically Synthesizing Web Pages." *In Proceedings of Association for the Advancement of Artificial Intelligence*, Madison, Wisconsin, pp. 727–732, 1998
- [9] Pierrakos, D., Paliouras, G., Papatheodorou, C., Spyropoulos, C.D., "Web Usage Mining as a Tool for Personalization: A survey." *In Proceedings of User Modeling and User-Adapted Interaction*, 13, pp.311-372, 2003
- [10] Rojanavas, P., Srinil, P., and Pinngern, O., "New Recommendation System Using Reinforcement Learning." *In Proceedings of the Fourth International Conference on eBusiness*, November 19-20, 2005, Bangkok, Thailand, 13(SP3), 2005
- [11] Schwarzkopf, E., "An Adaptive Web Site for The UM2001 Conference." *In Proceedings of the UM2001 Workshop on Machine Learning for User Modeling*, pp. 77–86, 2001
- [12] Skinner, B.F., *Science and human behavior*. New York, p.23, 1953
- [13] Srivastava, J., Cooley, R., Deshpande, M., Tan, P-T., "Web Usage Mining: Discovery and Applications of Usage Patterns from Web Data." *In Proceedings of SIGKDD Explorations*, 1(2), pp. 1-12, 2000
- [14] Sutton, R. S. and Barto, A.G., "Reinforcement Learning: An Introduction." *MIT Press*, Cambridge, MA, 1998
- [15] Taghipour, N., Kardan, A., Ghidary, S.S., "Usage-Based Web Recommendations: a Reinforcement Learning Approach." *In Proceedings of ACM Conference on Recommender Systems*, October 19-20, 2007, New York, USA, pp. 113–120, 2007
- [16] Te'eni, D., Feldman, R., "Performance and Satisfaction in Adaptive Websites: an Experiment on Searches within a Task-Adapted Website." *Journal of the Association for Information Systems*, 2(3), May 2001
- [17] Watkins, C.J.C.H., *Learning from Delayed Rewards*. PhD Thesis, University of Cambridge, England, 1989