

第壹章 緒論

1.1 研究背景

在激烈的商業競爭中，行銷策略的制訂無疑是影響組織能否達成獲利目標很重要的一個環節。而「瞭解客戶行為」又是組織想制訂良好行銷策略時首先應該擁有的資訊。無論是個人或是企業，如果能比競爭者更正確且有效率的從其所擁有的大量資料中找出更多未知的知識或資訊，自然在競爭場上就處於優勢的位置。資料挖掘(Data Mining)，或稱為知識發現(Knowledge Discovery)的研究議題便因為這樣的背景而漸漸的被重視，許多針對不同資料結構的資料發掘方法也被廣泛的討論。(Frawley et al., 1991)

在以往的傳統組織中，往往只能從客戶的消費、服務及維修等日常紀錄中著手收集與客戶有關的行為資訊。當組織想要有系統的收集消費者在商店內的選購行為或選購路徑等進一步的使用者資料時，組織需要面臨付出大量人力與新設備的投資才得以實現。隨著網際網路與資料收集技術的發展，商業網站經營者應用使用者 log 留存的技術，可以很容易的收集到使用者在其網站上的各種行為資料，相較於傳統業者來說，這些資料不僅範圍廣泛，而且十分精確。所以有越來越多的網站經營者開始重視使用者在網站中的行為。在網路泡沫化後，更重視客戶與深入瞭解客戶是他們能夠和傳統業者持續競爭的一個重要利基點。

1.2 研究動機

以往研究全球資訊網中使用者瀏覽行為的文獻，大部份學者皆專注於找出瀏覽頁面路徑部份的規則。(Agrawal and Srikant, 1995、Han, Yang and Kim, 1999、Chen, Park and Yu, 1998; 陳仕昇等，民 88)我們很少看到文獻會談到使用者在同一網頁中所進行的動作或是跨越網頁間動作路徑的問題。但對在全球資訊網上的使用者而言，使用者其實是藉由執行一個個的動作來控制網頁的路徑。網頁的瀏覽路徑也許可以提供使用者行為在意圖上某一程度的資訊，但也許並不周全。表 1-1 顯示了使用者在全球資訊網中可能的網頁-動作路徑。

表 1-1 使用者在全球資訊網中可能的網頁與動作路徑

<u>SessionId</u>	一般的網頁路徑	網頁動作路徑
A00000001	特價品資訊 → 產品資訊頁 → 訂單輸入頁	特價品資訊[特價資訊查詢→查詢產品明細] → 產品資訊頁[檢視產品明細 → 選擇周邊產品 → 檢視可搭配種類] → 訂單輸入頁[輸入訂購資訊 → 完成下單 → 完成交易]
A00000002	產品資訊頁 → 訂單輸入頁	產品資訊頁[檢視產品明細] → 訂單輸入頁[輸入訂購資訊 → 取消交易]
A00000003	特價品資訊 → 產品資訊頁 → 訂單輸入頁	特價品資訊[特價資訊查詢→查詢產品明細] → 產品資訊頁[檢視產品明細] → 訂單輸入頁[輸入訂購資訊 → 取消交易]
A00000004	產品資訊頁 → 訂單輸入頁	產品資訊頁[檢視產品明細 → 檢視可搭配種類] → 訂單輸入頁[輸入訂購資訊 → 完成下單 → 完成交易]
A00000005	特價品資訊 → 訂單輸入頁	特價品資訊[特價資訊查詢] → 訂單輸入頁[輸入訂購資訊 → 取消交易]

以一般的網頁路徑技術來看，我們可以找出「當使用者瀏覽過『產品資訊頁』後，接著瀏覽『訂單輸入頁』的話，有 50%的機率會完成交易」。但是在表 1-1

的資料中，我們可以觀察出「當使用者在『產品資訊頁』中執行過『檢視可搭配種類』後，有 100%的機率在『訂單輸入頁』會『完成交易』」這樣的規則。對於網站的經營者而言，這樣的規則對於更深入了解使用者行為是很有幫助的。於是我們開始思考，是否能夠發展出一種連使用者的行為路徑都考慮進去的資料發掘演算技術。

1.3 研究目的

本研究試圖建立一個商業網站收集及找尋規則的系統，包含建立一個網頁-動作路徑 Log 的留存機制來收集使用者行為與網頁路徑；以及利用本研究所介紹的一個新演算法「網頁-動作路徑演算法」，對收集到的使用者行為記錄找出其瀏覽行為的關聯模式。最後，為了評估系統的效能，我們會對於所建立出的雛形系統做一系列效能的模擬測試。

1.4 研究流程與步驟

本研究的研究方式，首先是確立研究主題，之後是收集研討相關的文獻資料。下一步我們會確定問題的模式。接著我們會發展出解決問題的演算法。從開始一直到這個步驟，我們都在理論架構的建立階段之中。接著我們進行網站系統的建置階段，其中包括需求分析與系統設計、網頁動作 Log 留存機制的建置、網頁-動作路徑演算系統的實作以及應用規則系統的實作。最後一個階段是資料分

析與應用階段，我們會開始收集使用者的瀏覽行為，並且實行演算法找出規則，

最後實作這些找出規則的應用。我們研究流程的示意圖如圖 1-1。

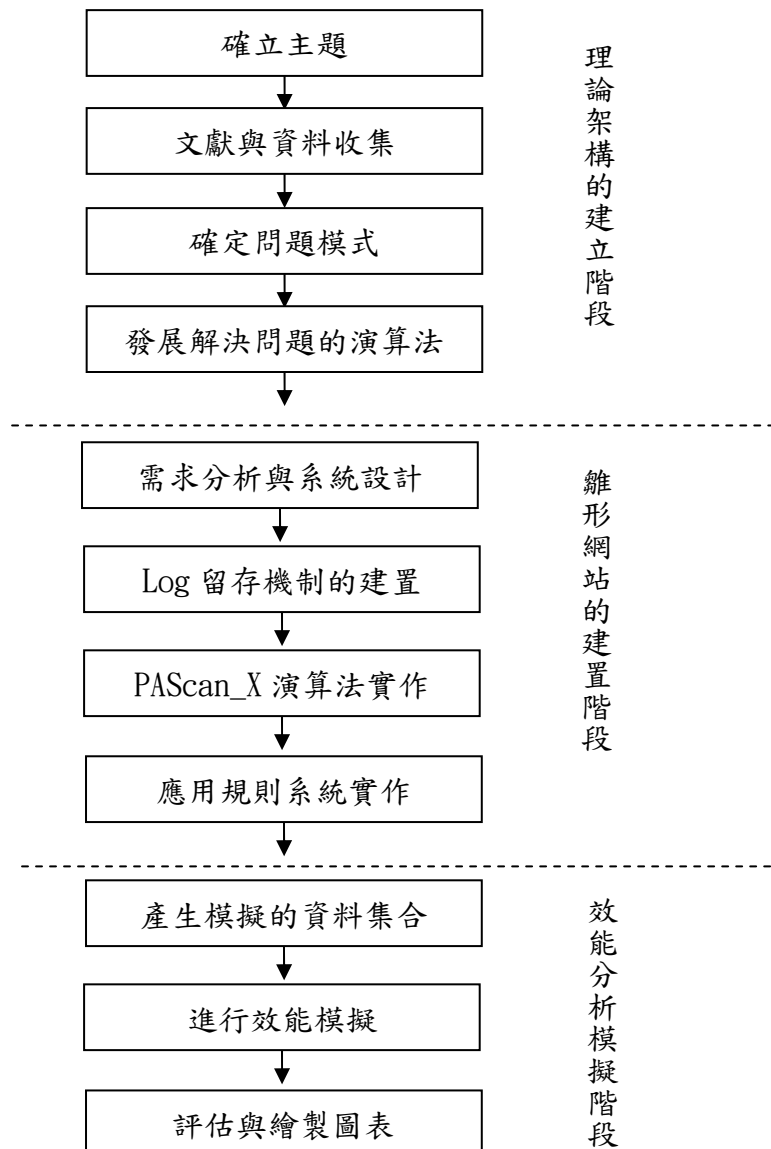


圖 1-1 本研究流程的示意圖

1.5 研究範圍與研究限制

本研究的研究範圍只針對使用者在全球資訊網路上所產生的網頁瀏覽動作

記錄加以研究；對於使用者真正的意圖是否一定會從網頁-動作路徑反應出來，這個我們無從得知，在本研究中，我們假設使用者在網站中瞭解自己做什麼，而且所有的動作行為是出自有意識的行為，而網頁-動作路徑的記錄可以反應出使用者的行為。第二個限制是因本研究希望針對動作做記錄，所以一般網站的系統記錄是無法拿來直接使用的，要取得能夠為本研究所用的記錄資料，網站系統得需要有記錄動作的功能，並且配合利用 Cookie、會員制的方式才能得到完整的網頁-動作路徑資料。第三個會受限的是找出的規則模式，由於本研究將掃描的方式設定為：對於網頁路徑的維度，限制於是必須連續出現，而單頁中的動作，則只考慮其出現的先後順序，不用完整出現。(為什麼這樣設定在 3.2.2 節中有較詳細的說明。)基於這樣設定的話，本研究將無法找出如「先執行過 A 頁，之後有執行過 B 頁」或是「在同一頁出現連續動作」這樣的規則。第四個會受到限制的是重覆計算的部份，目前我們所介紹的重覆計算，針對單一網頁路徑中的多次重覆動作我們還沒辦法處理。目前我們的重覆計算，只針對單一序列在不同頁中先後出現幾次 Pattern 才會處理。

1.6 論文架構

本節將本研究論文的架構做以下的說明：

- 第一章 緒論：

本章對本篇研究進行簡短的說明。描述本研究的背景、動機與目的、研

究流程與步驟，最後介紹研究範圍與限制與論文架構。

- 第二章 文獻探討：

本章以二部份進行文獻整理。第一部份先討論有關資料挖掘相關議題，特別是以 Agrawal 的 Apriori 演算法做資料發掘的方法。第二部份我們會探討以全球資訊網為資料挖掘的現有演算法，及與本研究的差別，以做為本研究資料挖掘方式的對照。

- 第三章 網頁-動作路徑的資料挖掘演算法：

本章分為四部份對網頁-動作路徑演算法做一個整體的介紹。首先會討論網頁-動作路徑的資料結構與其特性。第二部份我們會介紹一個網頁-動作路徑演算法，其中會對整個演算法的各別元件和動作做詳細的介紹。第三部份我們將經由一個小範例，實際操作 PAScan1 及 PAScanN 二類型的網頁-動作路徑如何找出所有的 Large Sequence。最後一部份我們會介紹關於找尋規則的演算過程。

- 第四章 離形系統實作與數據分析：

本章分為二部份來探討，第一部份介紹本研究建立出的離形系統的系統架構與其各別功能的說明。第二部份我們針對三大項六個模擬題目進行效能的測試，最後將測試的資料以圖表的方式呈現並加以討論。

- 第五章 結論與未來研究方向：

將本研究所的結果整理成結論，並對未來可行的研究方向做一些探討。