

第一章 緒論

隨著電腦技術的進步，使得人們可以處理的資料量快速增加，傳統的統計方法與分析技術已不敷使用。因此，資料挖掘技術的發展，提供我們一個可以處理大量資料並分析其中所隱含有用資訊的方法。同樣的，網路技術的進步與普及，也使得網路使用者與日遽增，網路上蘊藏無限的商機，也使得電子商務與電子資訊化服務的應用也愈來愈多，致使各式各樣交易資料量快速的增加且資料型態變的更為複雜。然而，許多有用的資訊往往隱藏在繁忙的網路行為中，若能夠加以有效的分析、瞭解及運用將可以幫助經營決策者取得較佳的優勢地位。

1.1 研究背景

近年來，隨著網際網路技術迅速的發展，造成網路使用率的普及，加上網路網路快速傳遞訊息的特性，對於一般民眾而言，已經習慣依賴網路上無限的資源來搜尋自己感興趣的資訊外，對於網路是否能針對使用者的需求做出正確的回應，提供適合的資訊以及縮短回應時間等技術上的要求也愈來愈講究。因此，研究如何藉由分析網路使用者瀏覽網頁的行為紀錄，去建立預測模型以預測網路使用者未來的瀏覽路徑，找出使用者未來可能瀏覽的頁面，用來加速提供網頁的速度、提高網站的效率及效能變的更為重要。

在 Web Mining 的研究中[1]，網路使用者的多變性，除了每一個使用者各有自己的獨特性外，就使用者本身而言，在不同時候，不同外在條件，也會使得使用者本身產生多樣的變化，造成了網站上使用者多變的行為模式。另一種則是難以判斷單一使用者，因為網路使用者具有相當的自主性

及隱私權，所以很難強制要求使用者以帳號登入的方式瀏覽網站，如此一來可能會發生不同使用者使用同一台電腦瀏覽網站，卻被視為同一個使用者的問題。目前在 Web Mining 的研究依目的分為兩大類：一類是希望藉由分析網路使用者的瀏覽行為去找出網路使用者的興趣所在，以便網站管理者能夠依據使用者的特性提供最適合使用者的資訊，其主要應用則是在推薦系統以及個人化服務上面[2]。另一類則是希望藉由分析使用者的瀏覽行為去預測使用者未來即將瀏覽的網頁[3]，預先存取相關網頁包含的元件，以改善頻寬不足造成網路壅塞的現象，避免使用者瀏覽網頁延遲時間增加的情形，其主要應用在快取伺服器上[4]。我們希望藉由分析網路使用者歷史的瀏覽行為紀錄，建立預測模型，去預測其他使用者未來可能瀏覽的網頁並預先存取準備好提供給使用者。簡單的說，一個是希望知道使用者有興趣的是什麼，並提供最符合該使用者有用的資訊，另一個則是想要知道使用者即將要產生的動作為何，預先做處理，縮短回應時間。

而本篇論文著重在後者，也就是說，我們希望藉由分析網路使用者歷史的瀏覽行為紀錄，建立預測模型，去預測其他使用者未來可能瀏覽的網頁並預先存取準備好提供給使用者。本篇論文提出一個完整的使用者行為預測系統架構並提出一個新的概念，利用 Petri-net 塑模使用者瀏覽行為的方式，預測使用者未來的瀏覽路徑。因此，研究如何藉由分析使用者的歷史紀錄(Log Database)，去建立預測模型以預測網路使用者未來的瀏覽路徑，找出使用者未來可能的瀏覽的頁面預先處理，用來加速提供網頁的速度、提高網站的效率及效能變的更為重要。在 Web Mining 的研究中，面臨的挑戰有三種：第一種就是，網路使用者的多變性，除了每一個使用者各有自己的獨特性外，就使用者本身而言，在不同時候，不同外在條件，也會使得使用者本身產生多樣的變化，造成了網站上使用者多變的行為模式；第二種是資料來源的不完整與不一致性，通常是因為這方面的研究，

資料來源是記錄在網站主機的 Log Database，而 Log Database 本身的限制會使得資料來源的品質不佳。第三種則是難以判斷單一使用者，因為網路使用者具有相當的自主性，及隱私權，所以，很難強制要求使用者以帳號登入的方式瀏覽網站，如此一來可能會發生不同使用者使用同一台電腦瀏覽網站，卻被視為同一個使用者的問題。

目前已有相當多的論文著重在這個研究的領域。有的是提出不同建立預測系統的方法，如[5][6]；有的則是著重預測效率的提升，而本篇論文則著重在提出一個完整的使用者行為預測系統架構，並提出一個新的預測方法概念來建立預測模型，預測使用者未來的瀏覽路徑(Traveling Path)。雖然已有很多研究提出不同的預測系統，如利用關連式法則[7]、Markov Chain[8]、Longest Repeating Subsequences[9]。但是，大部分的論文都必須有一個額外輔助的資訊，如：網頁的架構、頁面與頁面的相似度，來輔助預測系統。利用這些事先調查好的輔助資訊的確有利於預測系統的運作。然而，網路使用者瀏覽行為的多變性或不確定性卻降低了真正使用上的可行性。另外，也有預測系統是利用分析網路使用者瀏覽行為，將具有相同瀏覽行為或者興趣的瀏覽者分群，也就是說，同一群裡的瀏覽者，具有相同的瀏覽行為或者會瀏覽相同興趣的網頁。因此，當新的瀏覽者的瀏覽模式進來，我們會先判斷此一使用者屬於我們預先分好的何種類型，接下來，則可利用這群的特性來作預測。然而，在此一部分中，大部分的研究都只著重在如何利用使用者瀏覽行為或者使用者瀏覽網頁興趣的相似度有效的做分群，並且利用分群結果建立使用者的行為模型來進行預測。

在本篇論文中我們由預測模式中挑選出具有可供預測參考的參考模式(Reference Pattern)，利用分群來做預測，我們希望就僅知網路使用者過去的瀏覽Log Database 紀錄，去分析建立我們的預測模型，去預測網路使

用者未來的瀏覽路徑。因此本研究採用Petri-net模型化工具將網路組合開發者設計的網路服務組合流程加以模型化，然後運用Petri-net的分析方法模擬網路使用者的動作狀態，也由於採用這樣的預測概念，很有可能能夠提供做為我們預測的參考模式不只一個，而這樣的特性可以讓我們更容易適應多變的使用者行為。

1.2 研究目的

我們希望藉由分析Log Database 提供我們可以將哪些使用者想要瀏覽的檔案預先處理準備，如此一來可讓網路使用者瀏覽網頁，不會因為網路壅塞造成過份延遲的情況。但是，大部分的研究只告訴我們哪些頁面可能被存取的機會較高，卻並沒有提供我們這些存取機會很高的網頁，可能被存取的順序為何。

本篇論文中，我們希望我們的預測系統可以提供這樣的資訊。綜合以上所述，本篇論文的重點主要是提供一個完整的使用者未來行為預測系統架構，對使用者未來的行為進行預測，並以此特性適應於多變的使用者瀏覽行為，本研究以大學課程網為例，瀏覽過此網站的使用者可能包含了一般的使者、各校上傳課程的工作者、網頁管理者等，建立使用者行為資料庫前，我們先以網站歷史的存取紀錄作為建立預測模型的訓練資料，利用發掘的相關技術，經由Data Mining技術、Petri-net技術轉換後，建立使用者行為分析模組和規則表，產生使用者行為資料庫，藉此建立預測模組。

在預測模型階段，藉由分析與處理所找到的資料，建構出使用者行為預測模型(User Behavior Database)。預測模組階段中，憑藉著預測模型，預測未來使用者可能的瀏覽路徑，如果系統預測正確，就能提出修改網頁的建議，如果系統無法判斷，就會回授回去修正機制系統作改善，而本研究

僅探討如何應用Petri-net來分析使用者行為，以及建立預測模組去預測使用者未來的瀏覽路徑。至於探討回授控制的階段，將留待以後再作進一步研究。

根據上述研究動機，我們可看出目前網際網路的發達，相對的也產生了相當大的商機，在電子商務上，顧客關係管理為之重要，要從網際網路上了解顧客習性及行為，必定要能深入的分析使用者行為路徑，故本研究之研究目的為以下幾點：

1. 經由文獻分析介紹資料探勘及使用者行為路徑目前的發展概況。
2. 規劃並建立預測使用者行為路徑之架構圖。
3. 運用轉換控制矩陣建立動態路徑規則表。
4. 找出教育部大學課程網站使用者瀏覽路徑之關聯性。
5. 建立使用者行為之預測模型。

1.3 研究範圍與待答問題

1. 研究範圍

本研究僅針對教育部大學課程網做為研究之平台，對於其他網頁平台不加以討論，且本研究使用者行為預測模型僅能使用於大學課程網站，其餘各大網站不加以探討。可整理如下列幾點：

1. 本研究僅針對教育部大學課程網做為研究平台。
2. 本研究使用者路徑之資料擷取於 2006 年之前。
3. 本研究預測系統方面不探討回授控制的階段

2. 待答問題

1. 如何應用Petri-net技術於Web Mining？
2. 建構的預測模型如何改善網頁位置？
3. 預測之Petri-net模型能否正確找出使用者行為路徑？

1.4 理論技術與研究工具

為了建構網路使用者行者之預測模型，本研究運用了以下幾項的研究工具，和一些理論基礎來做為本論文之研究架構。

1. 資料探勘

從資料庫中發掘未知、潛在、有用，且最終是可被理解的資料型態(Patterns)的過程，也是資料庫中知識發掘(Knowledge Discovery in Databases, KDD)過程的一個步驟。

2. 轉換控制矩陣

特點為只要它的解存在，便可以透過數值方法找到數值解，轉換控制矩陣之所以廣為使用，是因為許多規格及限制條件都可以表示成LMI 型式。

3. Petri-net

基本的Petri-net也陸續延伸出許多更富有特性的高階Petri-net理論，以其延伸的特性來加強對系統塑模的能力，Petri-net可以建立系統的狀態方程式、代數方程式以及其他可以

管理系統行為的數學模組。

4. Artifex軟體

輸入狀態控制矩陣演算法到每一條路徑所對應的動作，建立對應動作的規則庫，所模擬的結果實驗後可得到簡化後的網路流程組合。

1.5 研究架構

1. 緒論

緒論部份描述本研究的研究背景、研究問題和研究目的並且說明本論文的結構分佈，做為本研究之相關背景及知識。

2. 文獻探討

本文章為相關文獻資料的收集與整理。主要在為本研究所提出的論點和解決方法提供相關的文獻說明，並比較現有的解決方法和本研究所提方法的差異。

3. 研究方法設計

本章描述本研究所提出的以Petri-net為基礎的網路的預測模型的測試與效能，如何以Petri-net塑模。

4. 實驗過程與結果

透過使用者行為路徑資料來預測模型的準確率及如何改善網頁

架構。

5. 結論

根據本研究所提出的解決方法和實作結果給予結論，並且說明本研究的研究貢獻、與其他研究的成果比較並且針對未來研究方向提出建議。

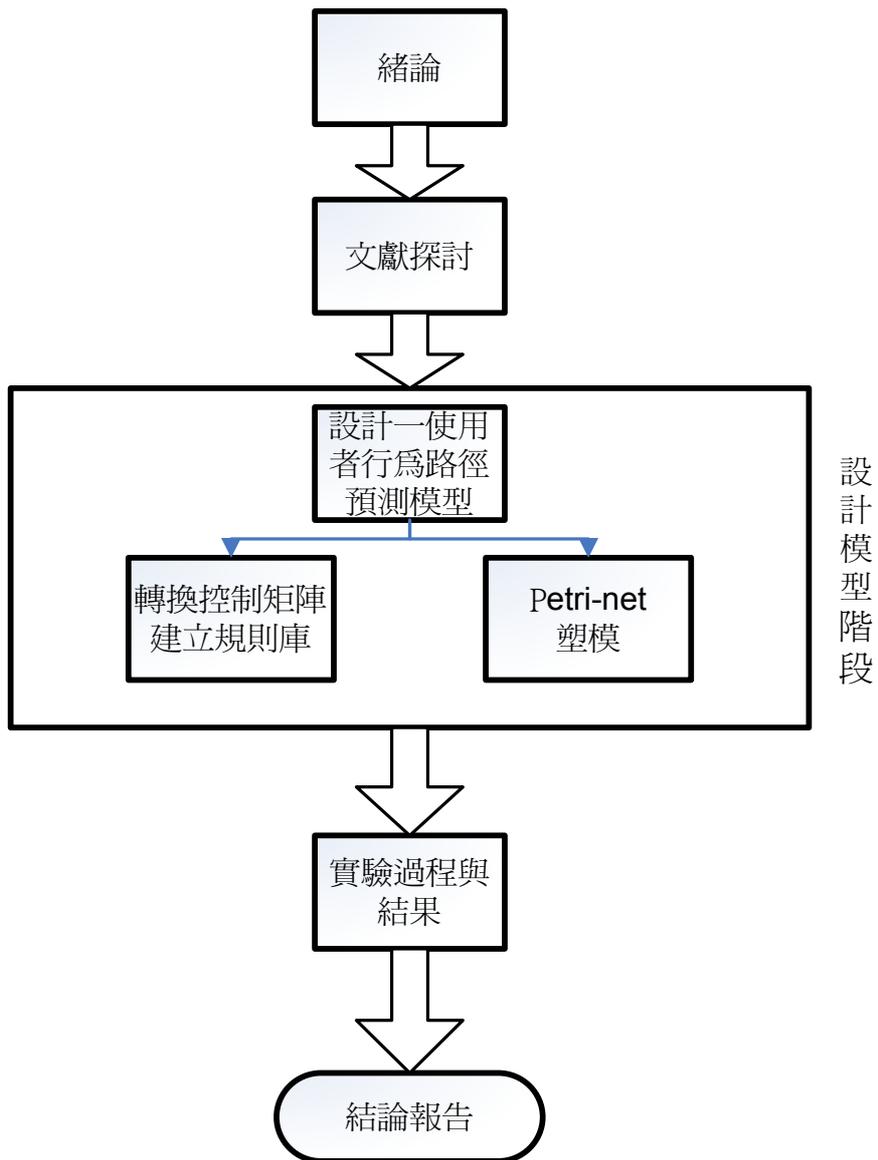


圖1-1 本論文之研究流程圖

(資料來源:本研究自行整理)