

第三章 應用Petri-net進行大學課程網使用者行為之塑模

在本章中，將介紹我們所提出用來預測多變的使用者行為的預測系統，並仔細介紹架構中各個組成的元件，以及各組成要素的建構步驟與方法。首先，我們在3.1 節中提出預測系統的完整架構，在3.2 節，應用Petri-net進行網路使用者行為之塑模，將利用資料探勘的方法建立預測模型，接著在3.3 節中，會詳細說明預測階段中，以大學課程網簡易查詢網頁為例進行塑模。並利用3.2節所建立的預測模型，依據使用者瀏覽行為趨勢的特性來預測使用者多變的瀏覽行為。在3.4 節中我們建立轉換控制矩陣表，藉著這個預測模式集，以提高整體的預測正確率。最後，我們在3.5 節中提出的預測模組架構以評估預測模型的準確度。由於使用者行為本身具有多變性與獨特性；因此，依據與單一群別或類別，過去使用者瀏覽行為的相似度，去預測當下使用者未來可能的瀏覽行為並不十分恰當。所以本篇論文的重點主要是提出一個利用Petri-net的概念去預測使用者未來瀏覽路徑的預測系統架構，藉此特性適應於多變的使用者行為，並產生具有順序性的預測結果。

3.1 研究架構

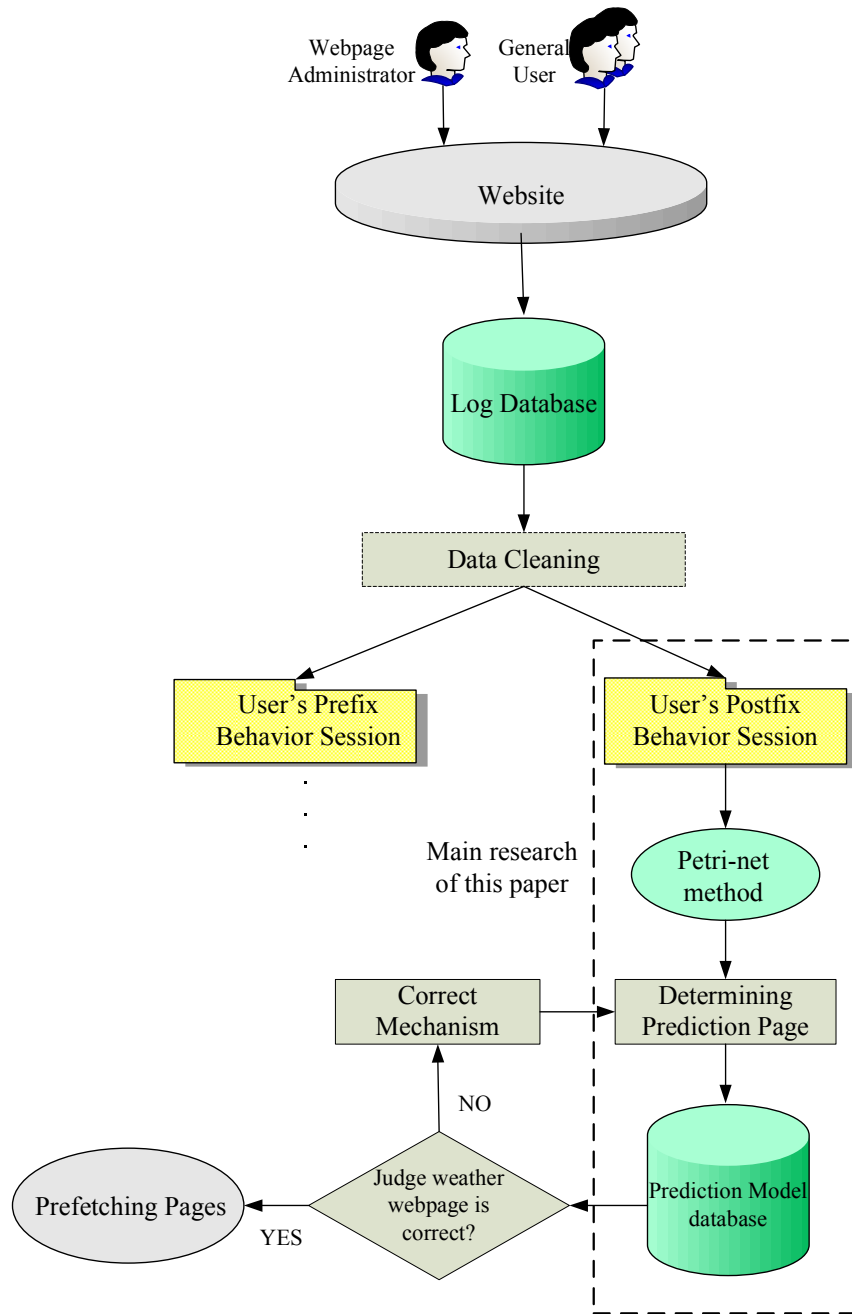


圖3-1 研究架構圖

(資料來源:本研究自行整理)

圖 3-1 為本研究架構流程圖，以大學課程網網站為例，瀏覽過此網站的使用者可能包含了一般的使用者、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，如：資料挖掘、分群或分類等技術找到可做為預測參考的樣式。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，建構出一種使用者行為預測資料庫(User Behavior Database)。

圖 3-1 中的預測模組階段中，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，即能判斷使用者下一步可能瀏覽的路徑，如果系統無法判斷，就會藉由回授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。

3.2 應用Petri-net進行網路使用者行為之塑模

The screenshot shows a university website interface. At the top, there is a navigation bar with links: 回首頁, 管理專區, 軟體下載, 線上報名, 線上教學, 常見問題, 流量統計. Below this is a main content area with three columns. The left column has '課程查詢' (Course Search) with sub-links for '簡易查詢', '進階查詢', and '科系資料查詢'. The middle column is a '公佈欄' (Notice Board) with several announcements regarding system maintenance and course uploads. The right column features a '課程搜尋' (Course Search) box and a '統計資料' (Statistics) table.

學年度	課程數量
94	276796
93	268765
92	266852
91	243445
90	227999
89	150552
88	10001
87	966
86	604
總計	1445980

Below the statistics table, it says: Last update: 2006/6/20 上午 09:45:29. At the bottom of the page, it states: 共有 254416 來訪.

圖3-2 大學課程網網頁畫面
(資料來源:本研究自行整理)

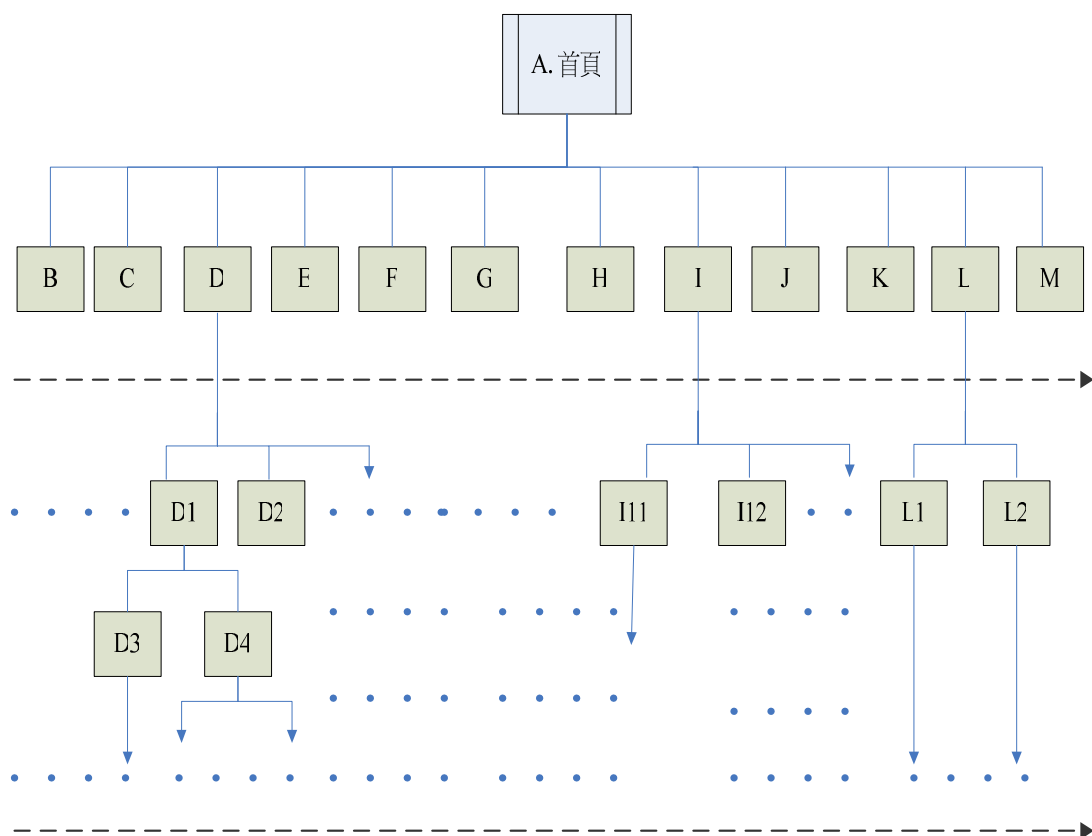


圖 3-3 線上瀏覽者知識發現行為路徑之網頁架構圖

(資料來源:本研究自行整理)

圖 3-2 為大學課程網的實際網頁畫面，我們先將網頁的每一個按鈕和連結轉換為大學課程網站的路徑之網頁架構圖(圖 3-3)，以此來模擬大學課程網使用者的行為路徑推測，以便觀察或模擬使用者行為路徑之狀態位置。圖 3-3 中各項主網頁、次網頁、按鈕和資訊，以選項符號來代表，整理如表 3-1、表 3-2 所示。圖 3-3 中的符號定義，其中符號 A 代表此網站的主網頁(首頁)，符號 B、C、D... 為次網頁，再下一層的網頁為 B1、B2、C1、C2...，以此類推。

表 3-1 大學課程網主網頁與次網頁的按鈕或連結項目定義表

B	按下「回首頁」按鈕	B1	按下「公佈欄」按鈕
C	按下「管理專區」按鈕	C1	按下「作業登入」按鈕
D	按下「軟體下載」按鈕	D1	按下「立即下載」按鈕
E	按下「線上報名」按鈕	E1	按下「大學校院課程上網研討會」按鈕
F	按下「線上教學」按鈕	E2	按下「技職校院課程上網研討會」按鈕
G	按下「常見問題」按鈕	F1	按下「回首頁」按鈕
H	按下「流量統計」按鈕	F2	按下「播放器下載」按鈕
I	按下「簡易查詢」按鈕	F3	按下「單元一」按鈕
J	按下「進階查詢」按鈕	F4	按下「單元二」按鈕
K	按下「科系資料查詢」按鈕	F5	按下「單元三」按鈕
L	按下「主題活動查詢」按鈕	F6	按下「單元四」按鈕
M	按下「主題活動成效」按鈕	F7	按下「單元五」按鈕
N	按下「各校資料查詢」按鈕		

(資料來源:本研究自行整理)

表 3-2 大學課程資料倉儲網主網頁與次網頁的欄位選項定義表

C11	輸入「帳號」欄位	J17	選擇「必修選修」欄位
C12	輸入「密碼」欄位	J18	選擇「科目類別」欄位
I11	選擇「關鍵字」欄位	J19	選擇「全外語」欄位
I12	選擇「縣市區域」欄位	J20	選擇「檢視欄位」欄位
I13	選擇「學年度」欄位	K11	輸入「系所名」欄位
I14	選擇「學期」欄位	K12	選擇「開課年度」欄位

J11	選擇「開課學校」欄位	L11	輸入「關鍵字」欄位
J12	選擇「學分數」欄位	L12	選擇「開課學校」欄位
J13	選擇「開課年級」欄位	L13	選擇「學年度」欄位
J14	選擇「學制」欄位	L14	選擇「學分數」欄位
J15	選擇「開課學期」欄位	M11	按下「教育主題」任一個按鈕
J16	選擇「部校定」欄位	N11	按下「查詢縣市」任一個按鈕

(資料來源:本研究自行整理)

3.3 以大學課程網簡易查詢網頁為例進行塑模

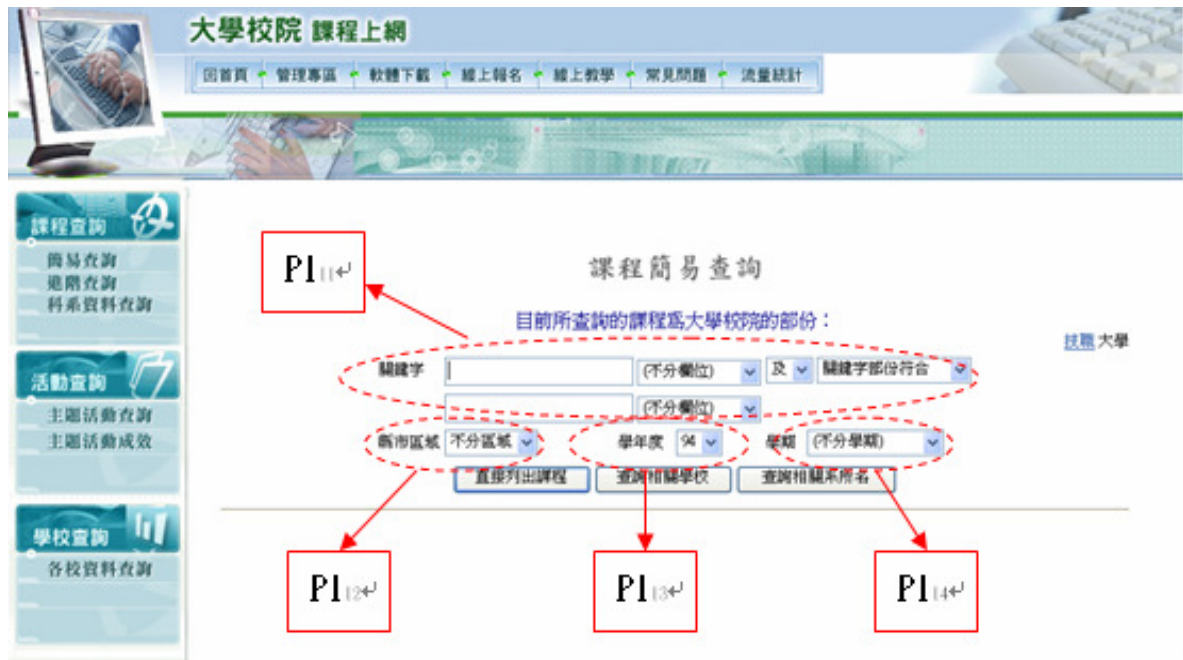


圖 3-4 大學課程網簡易查詢網頁畫面

(資料來源:本研究自行整理)

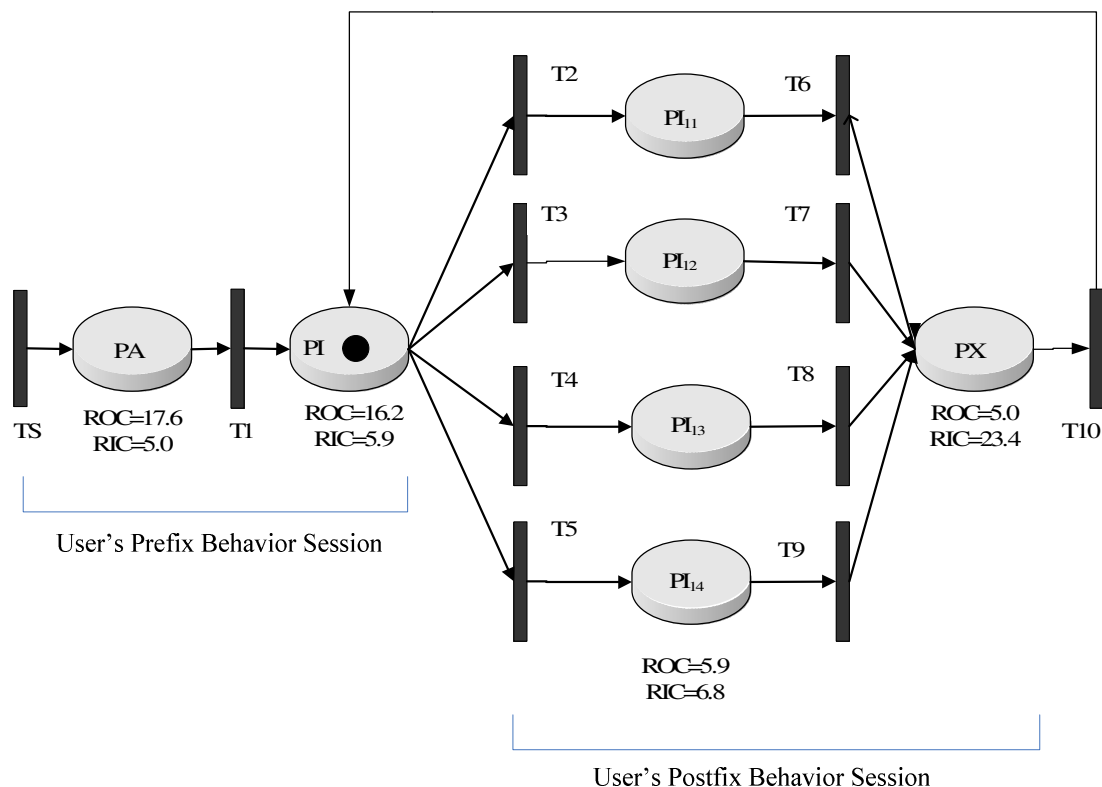


圖 3-5 應用 Petri-net 分析使用者進行簡易查詢動態行為圖

(資料來源:本研究自行整理)

表 3-3 使用者動態行為表格符號定義表

PA	「首頁」網頁狀態	T2	觸發進入「關鍵字」網頁
PI	「簡易查詢」網頁狀態	T3	觸發進入「縣市區域」網頁
PI₁₁	「關鍵字」網頁狀態	T4	觸發進入「學年」網頁
PI₁₂	「縣市區域」網頁狀態	T5	觸發進入「學期」網頁
PI₁₃	「學年」網頁狀態	T6	觸發進入「查詢結果」網頁
PI₁₄	「學期」網頁狀態	T7	觸發進入「查詢結果」網頁
PX	「查詢結果」網頁狀態	T8	觸發進入「查詢結果」網頁
TS	觸發進入「首頁」網頁	T9	觸發進入「查詢結果」網頁
T1	觸發進入「簡易查詢」網頁	T10	觸發進入「簡易查詢」網頁

(資料來源:本研究自行整理)

網路使用者在一個網站上瀏覽的路徑千變萬化，本研究以使用者路徑之中一個瀏覽路徑來分析，我們以使用者進入大學課程網「簡易查詢」的網頁作為我們的研究範圍(圖 3-4 所示)，我們根據圖 3-4，網路使用者做簡易查詢的動作，必須經過關鍵字(PI_{11})、縣市區域(PI_{12})、學年(PI_{13})、學期(PI_{14})等其中一個動作來查詢。應用 Petri-net 建構使用者在該網頁中的動態行為(圖 3-5 所示)，而表 3-3 為此「簡易查詢」的網頁使用者動態行為表格符號定義。我們定義 Petri-net 對使用者在網頁的行為如下[21]:

$P = \{P_1, P_2, P_3 \dots P_n\}$, ($n \geq 0$) 為一有限位置或狀態集合

$T = \{T_1, T_2, T_3 \dots T_n\}$, ($n \geq 0$) 為一有限轉變或處理集合

$P \cap T = \{\}$, $F \subseteq (P \times T) \cup (T \times P)$

$I(T) = \{P \mid (P, T)\}$, $O(T) = \{P \mid (T, P)\}$

$I(P) = \{P \mid (T, P) \in F\}$, $O(P) = \{P \mid (P, T) \in F\}$

以大學課程網站為例，可描述如下:

$P = \{\text{網頁畫面所成的狀態}\}$, $T = \{\text{大學課程網站上的按鈕或觸發集合}\}$

我們定義大學課程網資料庫的基本流向控制矩陣為 UCDW(University Course Data Warehouse)，則可獲得:

$Up = \{P_A, P_I, PI_{11}, PI_{12}, PI_{13}, PI_{14}, P_X\}$

$T = \{T_1, T_2, T_3, T_4, T_5, T_6, T_7, T_8, T_9, T_{10}\}$

3.4 建立轉換控制矩陣表

表 3-4 轉換控制矩陣表(CW_M)

CW_M	PA	PI	PI ₁₁	PI ₁₂	PI ₁₃	PI ₁₄	PX	COD	ROC
PA	0	1	2	2	2	2	3	12	11.4
PI	7	0	1	1	1	1	2	13	10.5
PI ₁₁	7	2	0	3	3	3	1	19	7.2
PI ₁₂	7	2	3	0	3	3	1	19	7.2
PI ₁₃	7	2	3	3	0	3	1	19	7.2
PI ₁₄	7	2	3	3	3	0	1	19	7.2
PX	7	1	7	7	7	7	0	36	3.8
CID	42	10	19	19	19	19	9	CD=137	
RIC	3.3	13.7	7.2	7.2	7.2	7.2	15.2		

(資料來源:本研究自行整理)

表3-4為 CW_M (Converted Weight Matrix) [22][23]，是分析圖3-5所得到的結果。其中 COD (Converted Output Degree)為向外內分支度， ROC (Relative Output Degree)為相對向外分支度， CID (Converted Input Degree)為向內分支度， RIC (Relative Input Degree)為相對向內分支度。

表 3-4 的第一欄為下錨節點 N_A ， $N_A = \{node \mid node \in N_{anchor}\}$ ，第一列為目的節點 N_D ， $N_D = \{node \mid node \in N_{destination}\}$ [22]，

$$CW_{Mij} = \begin{cases} M_{wij} & \text{if } w_{ij} \neq \infty \\ K & \text{otherwise} \end{cases} \dots\dots\dots(1)$$

$$COD_i = \sum_j CW_{Mij}, ROC = \frac{CD}{COD_i} \dots\dots\dots(2)$$

$$CID_j = \sum_i CW_{Mij}, RIC = \frac{CD}{CID_j} \dots\dots\dots(3)$$

$$CD = ROC_i \times COD_i = RIC_j \times CID_j \dots \dots \dots (4)$$

$$CD = \sum_j CW_{Mij} = \sum_j CW_{Mij} = \sum_j COD = \sum_i CID \dots \dots \dots (5)$$

上述各式中 i、j 分別代表橫軸與縱軸，(1)式為 COD_i 、 ROC 方程式，(2)式為 CID_j 、 RIC 方程式，(3)、(4)式 CD 方程式，在此所選取的路徑為最短的路徑距離，若是無法定義的流向，我們在此給權重 $K=7$ ，則可得到表 3-4 的數據，在此我們看到 PI_{11} ， PI_{12} ， PI_{13} ， PI_{14} 的 RIC 都為 7.2， ROC 都為 7.2，由此可知此網頁架構的權重都相同，代表網頁位置的出線機率相同。

表 3-5 轉換矩陣 CW_M 的 ROC 與 RIC 變化關係表

		ROC	
		變大	變小
RIC	變大	發散出去多 接收進來多	發散出去少 接收進來多 (如 PX 節點)
	變小	發散出去多 接收進來少 (如 PA 節點)	發散出去少 接收進來少

(資料來源:本研究自行整理)

表 3-5 為轉換矩陣 CW_M 的 ROC 與 RIC 變化所得的關係分析表，當 ROC 值愈大時，從此點所發散出去的節點關連度愈多；反之，當 ROC 值愈小時，從此點所發散出去的節點關連度愈少。當 RIC 值愈大時，接收到此點的節點關連度愈多；反之，當 RIC 值愈小時，接收到此點的節點關連度愈少。運用此法則建立規則庫和經驗判斷法則來建立預測模型，由表 3-5 可

形成規則如下:

Rule 1: 當 *ROC* 值越高且 *RIC* 值越低時, 表示此點至對外的節點關連度越大; 亦即該點至其餘各點的路徑越短。通常為起始節點, 例如: PA。

Rule 2: *ROC* 值越低且 *RIC* 值越高時, 表示此點至對外的節點關連度越小; 亦即該點至其餘各點的路徑越長。通常為終結節點, 例如: PX。

3.5 預測模型架構

根據第一節第二節以大學課程網簡易查詢網頁進行塑模, 再以第三節的轉換控制矩陣表計算出各個節點之間相互關聯度, 運用此法則建立規則庫來建立預測模型。建立預測模型階段時, 當使用者進入「簡易查詢」網頁時, 利用資料探勘的方式來處理使用者的歷史資料, 萃取出訓練模式 (Training Pattern) 的資料, 表 3-6 為選取 2006 年 12 月份大學課程網使用者進入簡易查詢的瀏覽路徑資料。

表 3-6 使用者路徑資料

	T2	T3	T4	T5
User1	PI ₁₁	PI ₁₂		
User2				PI ₁₄
User3	PI ₁₁		PI ₁₃	
User4	PI ₁₁			
User5	PI ₁₁		PI ₁₃	PI ₁₄
User6			PI ₁₃	
User7	PI ₁₁			
User8	PI ₁₁	PI ₁₂	PI ₁₃	PI ₁₄

User9			PI ₁₃	
User10	PI ₁₁			
User11				PI ₁₄
User12	PI ₁₁		PI ₁₃	
:	:	:	:	:
:	:	:	:	:
Total	132	48	91	69


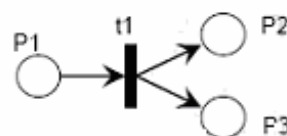
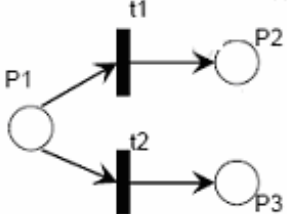
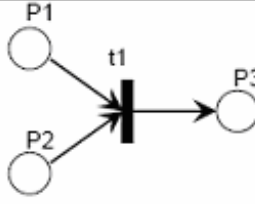
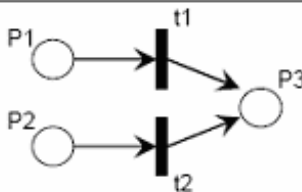
(資料來源:本研究自行整理)

使用者進入大學課程網進行簡易查詢動作時，經由 PA 再到 PI 狀態，接著進入 PI₁₁、PI₁₂、PI₁₃、PI₁₄ 其中一條路徑，可能行經的路徑為使用者經由 T2 觸發動作進入 PI₁₁ (「關鍵字」網頁狀態)，T4 觸發動作進入 PI₁₃ (「學年」網頁狀態)，T5 觸發動作進入 PI₁₄ (「學期」網頁狀態)，T3 觸發動作進入 PI₁₂ (「縣市區域」網頁狀態)，根據表 3-6 所計算出的結果，200 筆使用者路徑資料庫得知使用者作簡易查詢動作時，最常查詢的動作觸發頻率為觸發進入「關鍵字」網頁(T2= 132)，依序頻率為觸發進入「學年」網頁(T4= 91)、觸發進入「學期」網頁(T5= 69)、觸發進入「縣市區域」網頁(T3= 48)，而網頁所代表的狀態為依序 PI₁₁、PI₁₃、PI₁₄、PI₁₂，以此排序來建立模組。

本研究採用Petri-net技術中的Incidence Matrix and State Equation方法[24]，分析方法是將Petri-net模型以輸入矩陣 D^- 和輸出 D^+ 矩陣的方式表達，每個矩陣皆擁有m列和n行，分別代表Petri-net模型中的transition和place。輸入矩陣 D^- 代表輸入到transition的place，輸出矩陣 D^+ 代表由transition輸出的place。而代表整個Petri-net系統模型的矩陣D可由以下的式子求得： $D = D^+ - D^-$ 。表3-7為輸入矩陣 D^- 、輸出矩陣

D^+ 和系統模型矩陣 D 的表達範例。

表3-7系統模型矩陣的表達範例。

(a)		$D^- = [1 \ 0]$ $D^+ = [0 \ 1]$ $D = [-1 \ 1]$
(b)		$D^- = [1 \ 0 \ 0]$ $D^+ = [0 \ 1 \ 1]$ $D = [-1 \ 1 \ 1]$
(c)		$D^- = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix}$ $D^+ = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$ $D = \begin{bmatrix} -1 & 1 & 0 \\ -1 & 0 & 1 \end{bmatrix}$
(d)		$D^- = [1 \ 1 \ 0]$ $D^+ = [0 \ 0 \ 1]$ $D = [-1 \ -1 \ 1]$
(e)		$D^- = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$ $D^+ = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 1 \end{bmatrix}$ $D = \begin{bmatrix} -1 & 0 & 1 \\ 0 & -1 & 1 \end{bmatrix}$

(資料來源:參考戚玉樑碩士論文)

假設 M_i 為系統起始的狀態， S 為標記 M_i 開始執行的觸發序列

$\{T_1, T_2, T_3, T_4, T_5, T_6, T_7, T_8, T_9, T_{10}\}$ ， M 為觸發狀態的集合。

#符號代表為取絕對值後的正整數。

$D^- [j, i] = \# (P_j, I(t_j))$ ， D^- 為 Transition 的輸入(I)

$D^+ [j, i] = \# (P_j, O(t_j))$ ， D^+ 為 Transition 的輸出(O)

$UCDW = (P, T, D^-, D^+, M)$ ，最後我們所要的狀態為 $D = D^+ - D^-$

UCDW (University Course Data Warehouse) 為我們定義的大學課程網資料倉儲。