

以序列探勘技術建構個人化網路學習路徑

黃河銓 曾俊銘

國立高雄應用科技大學 資訊管理研究所

hchuang@cc.kuas.edu.tw

1094345109@cc.kuas.edu.tw

摘要

在現代的社會中，網際網路已成為生活中不可或缺的一部分，網路學習亦成為一種廣受歡迎的學習方式。而要如何利用網路學習系統中之先前學習者的學習經驗以輔助新的學習者便成了一個很重要的研究議題。本研究提出了一種結合學習風格與序列探勘之方法，並同時考量到教材學習時間之因素來為學習者建構適性化之學習路徑。利用學習風格可以讓學習者學習符合自身風格的適性化教材，而序列探勘的應用則讓學習者能夠依適當順序學習教材。在學習時間有限的情況下，考量學習教材所需要的時間可讓學習者擁有足夠的時間學習教材。因此本研究整合了上述優點，應用序列探勘分析舊學習者之學習路徑，將其分析結果加以整合，並加入學習風格相似度及學習教材所需要之時間做為路徑選擇之依據，建構出適性化之學習路徑。

關鍵詞：適性化學習系統、學習風格、序列探勘

1. 前言

隨著網路的普及，知識的學習開始以網際網路做為傳遞知識的媒介。數位學習系統讓學習者擺脫傳統的學習方式，讓學習者可自由的選擇學習的時間、地點與學習的主題。傳統的數位學習系統其網頁的內容與網站之鏈結無法隨著學習者的特質與能力變化，只能提供相同的內容給所有的學習者。然而不同的學習者其能力與特性亦會不同，這導致傳統的教學無法讓每位學習者達成有效的學習成果。適性化學習系統(Adaptive Learning System)的發

展讓數位學習系統得以考慮個別的學習者的目標、偏好以及知識，為每位學習者提供最需要的個別化資訊[3]。藉由對於不同的學習者提供符合學習者特性的教材，適性化學習系統可讓學習者更有效、更快速的獲得各種知識與技能。一個適性化學習系統必需要具備下面幾個能力：(1)監控使用者的活動；(2)基於特定領域的模型解釋這些活動；(3)利用活動之解釋推論使用者之需求與偏好並表示成關聯的模型；(4)依照使用者的知識與現有的題材讓學習過程更加容易[7]。在近代的學習系統中，大多運用超媒體技術以輔助教材的建置，而有關適性化超媒體之技術主要可分為兩大類，分別是適性化呈現(Adaptive Presentation)與適性化導航(Adaptive Navigation Support)。適性化呈現又可分為文字適性化與多媒體適性化技術，適性化導航則分成鏈結的隱藏、排序、註解、直接指導、導覽圖適性化等技術[3]。適性化呈現主要是針對學習者想學的教材或內容作適當的調整，而適性化導覽則是根據學習者的特質提供個別化的超媒體連結與內容。現在已有許多系統利用這些技術提供適性化之學習，Wolf 所提出的 iWeaver 系統便利用分析學習者的學習風格來提供適性化之學習[12]。該系統針對各種學習風格的學習者提供合適版本的教材與各式學習工具，卻忽略了課程中各項知識間的相關性與順序性。Morales et al.則利用序列探勘進行鏈結的推薦，藉此找出最佳之學習路徑，但卻忽略了學習者的學習風格[6]。因此，本研究結合了序列探勘與學習風格之優點，考量到知識間的相關性、順序性及教材的適合性與學習者對知識的理解程度，分析出適合學習者之適性化學習路徑，以提供學習者適性化之學習環境。

2. 學習風格與序列探勘

為了要讓適性化學習系統能夠落實個別化學習輔助的特性，我們利用學習理論中所提出的各項影響要素輔助學習者達成有效的學習。Carroll [4] 提出了學校的學習模型，他將學校的學習表示成一個時間的函數，公式如下：

$$\text{School learning} = f(\text{Time spent} / \text{Time needed}) \quad (1)$$

Time spent 代表真正花費在學習上的時間，由機會與毅力所組成。Time needed 則是代表學習所需要的時間，由才能、理解教學的能力及教學品質所構成。Slavin 以 Carroll's model 為基礎，提出了 QAIT 模型，將重點放在模型中學校與教師可直接修改掌控之元素[9]。QAIT 模型由下列之元件所構成：

- 教學的品質 (Quality of Instruction): 對資訊或技能的表達程度，能否讓學習者簡單的學習。
- 等級合適的教學 (Appropriate Levels of Instruction): 確定學生在學習某個課程前，已具備學習新課程所需要的知識與技能。
- 動機 (Incentive): 激發學生對於致力於教學目標的達成與教材的學習意願。
- 時間 (Time): 學生是否有獲得足夠的時間學習其教材。

在學習過程中，滿足 QAIT 四個因素，可達成有效的教學，對於學習者的學習成效亦有很大的幫助。

此外，學習風格(Learning Style)包含了個人的行為與偏好模式[2]，若是能夠滿足這些偏好則可大幅提升學習者的學習成效。Felder and Silverman [5] 運用四種不同的維度為學習者的學習風格分類，共可組成 16 種不同的學習風格。這四個維度分別是：

■ 主動(active) - 反思維度(reflective)

用於衡量學習者處理資訊的方式，主動型的學習者獲得知識的方法是藉由試驗並觀察其如何運作，並偏好團隊合作。反思型的學習者則是偏好在行動之前先進行徹底的思考，並且喜歡獨自工作。

■ 感覺(sensing) - 直覺維度(intuitive)

用來量測學習者對何種類型之資訊感知較佳，

感覺是透過感官的途徑來察覺，使用直接的感覺來收集資料。而直覺則是透過間接的感覺，在本身無特別意識的情況下發覺與觀察。感覺型的學習者喜歡學習事實，解決問題時偏好使用固定的方法，不喜歡複雜與意料外的事。直覺型的學習者通常較喜歡探索可能性與關聯(理論方面)，有著更多創新，但不喜歡重複性質的工作。

■ 視覺(visual) - 口語(verbal)維度

這個維度則是測驗出了學習者對視覺化或口語化感知訊息的喜好程度。視覺型的學習者擅長記憶圖型化的資訊，可以從圖表、影片、流程圖等獲取更多的知識；口語型的學習者則擅常記憶文字與聲音資訊，透過讀與寫解釋說明可以學得更有效率。

■ 連續(sequential) - 總體維度(global)

用來衡量學習者促進理解之最佳方式為何。連續型的學習者解決問題時是隨著線性的步驟推理，而下一步通常可由前一步推理出來。總體型的學習者擅長跳躍性的思考，讓此類學習者自行找出問題解決方法效果會比強迫接受他人方法來的好。

學習者的學習風格可利用 Felder and Soloman 所設計的學習風格量表測得[10]。該量表包含 44 個問題評估學習者在上述四個維度上的偏好。當學習者對某一維度有強烈偏好時，學習者在不支援該偏好的環境中學習可能會遇上困難。每一個題目都有兩個答案可供選擇，分別對應到某個維度上的某項特性。每個維度上皆有兩種互斥的特性，受測者對每一個維度只會測出一個特性，共可測得四個風格特徵值。舉例來說，對於同一學習者，問卷只會測出主動與反思兩種特性其中之一。學習風格可由所測得的四項風格特徵值劃分，其範圍在 1~11 分之間，代表該項特性之強弱程度。分數越大代表該名學習者對該特性之表現越強烈。

序列樣式(sequence pattern)在資料探勘領域中是一個很重要的研究主題，在舊有的關聯法則中，只能找出各個項目之間的關係，並沒有考慮到項目間的先後順序。序列探勘(sequence mining)解決了這個問題，它找出了各個項目的關聯與序列性。一個序列是由多筆依照時間排序的交易所組成，序列的長度為交易資料的筆數。而一個由交易 A、B、C(依

時間排序)所構成序列可表示為 $A \rightarrow B \rightarrow C$ 。每一筆序列的支持度(support)代表在所有客戶中,有多少百分比的客戶符合該序列。Agrawal and Srikant [1] 一文中介紹了三種不同的演算法以解決發現序列樣式的問題。AprioriAll、ApriorSome 與 DynamicSome 便是藉由分析所有序列找出經常出現的序列樣式。Ramakrishnan and Rakesh [11]介紹了 GSP(Generalized Sequential Patterns)演算法,該演算法以比 AprioriAll 更快的速度找出符合使用者自訂之支持度的序列樣式。Parthasarathy et al. [8]發表了 ISM (Incremental Sequence Mining)演算法,改善了舊有的演算法在資料庫更新後需重新分析的缺陷,透過演算法中所建立之 ISL(Increment Sequence Lattice)以支援漸進式序列探勘。

3. 適性化學習路徑

本研究主要是利用序列探勘與學習風格以產生適性化學習路徑。而本研究之適性化學習路徑主要是依照教學目標針對課程中的每一項知識尋找合適的教材並安排適當之學習順序。我們為每一個別的知識建立了多個不同學習風格(版本)的教材,以讓系統為學習者安排適性化之教材。在系統中將每一份教材視為一個節點,而學習者學習這些教材的順序即為學習路徑。在一開始系統會先為學習者做兩項測驗,分別是學習風格測驗以及學習者對課程中各項知識的了解程度測驗。接著會利用序列探勘分析舊學習者的學習路徑以取得序列樣式,並將序列樣式整合成圖形結構,我們稱之為學習路徑樣式(learning path patterns)。此外,對於學習者完全了解的節點系統將會自動地從學習路徑中移除。下一步則是計算學習路徑樣式中每一條路徑的權重值,最後再依照權重值選擇出適當的學習路徑。

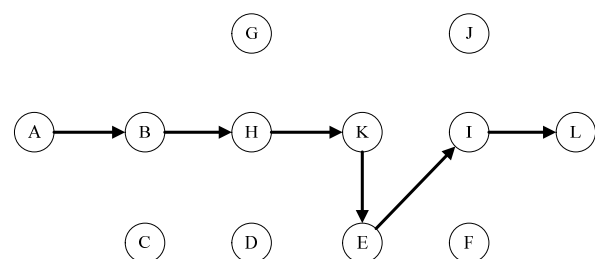
在序列探勘的部分,我們採用 ISM 演算法,演算法會於初次使用時建構 ISL 儲存資訊。ISL 分為兩個部份,FS(Frequent Set)儲存了所有頻繁序列的資訊,NB(Negative Border)儲存了所有最有可能成為頻繁序列的序列。在 NB 中所儲存的序列有著

下列幾項特點:(1)序列所分割出之子序列皆為頻繁序列;(2)序列本身並非頻繁序列。利用 NB 其所儲存之資訊可讓系統在新增資訊時快速的計算出新的頻繁序列。ISM 演算法最主要分為兩個階段,第一階段主要是當有新增資訊時,更新 NB 與 FS 中每一個序列的支持值,並找出 NB 中能成為頻繁序列的序列,將其加入 NB-to-FS 列表中。第二階段則是從前一階段所找出的 NB-to-FS 中之頻繁序列資訊存入 ISL 中,同時利用新的頻繁序列與其他相同字尾之頻繁序列做交集運算,並將產生之新序列加入 NB 或 FS 中。

在取得序列樣式之後,將序列樣式轉換成學習路徑樣式。圖 1 是序列探勘所找出的序列樣式,在序列樣式轉換前,系統會先檢查系統中有哪些教材,以教材做為結構中之節點。節點與節點間的路徑是根據序列樣式來建立路徑。如圖 2 所示,使用編號 1 之頻繁序列時,系統將建立 $A \rightarrow B$ 、 $B \rightarrow H$ 、 $H \rightarrow K$ 、 $K \rightarrow E$ 、 $E \rightarrow I$ 以及 $I \rightarrow L$ 之間的連結。而對於學習者完全了解之節點,我們將取消所有通往該節點之路徑,例如若學習者完全了解節點 H,我們將移除所有通向節點 H 之路徑(如 $B \rightarrow H$)。

編號	頻繁序列
1	$A \rightarrow B \rightarrow H \rightarrow K \rightarrow E \rightarrow I \rightarrow L$
2	$A \rightarrow C \rightarrow H \rightarrow K \rightarrow E \rightarrow F \rightarrow L$
3	$A \rightarrow C \rightarrow D \rightarrow H \rightarrow E \rightarrow F \rightarrow L$
4	$A \rightarrow B \rightarrow G \rightarrow J \rightarrow K \rightarrow I \rightarrow L$
5	$A \rightarrow C \rightarrow D \rightarrow E \rightarrow I \rightarrow J \rightarrow L$

圖 1 序列探勘找出的頻繁序列



使用的頻繁序列： $A \rightarrow B \rightarrow H \rightarrow K \rightarrow E \rightarrow I \rightarrow L$

圖 2 路徑建構示意圖

成功取得學習路徑樣式之後，我們將開始計算每一條路徑的權重值。權重值的組成可分為兩個部份，分別是學習風格相似度以及學習教材所需要的時間。首先我們先來看學習風格相似度的部分，要計算相似度必須要先計算學習風格特徵平均值以及取得新學習者之風格特徵值。以路徑 $A \rightarrow B$ 為例，圖 3 為通過路徑 $A \rightarrow B$ 之學習者其資訊之彙整。通過人數是指舊學習者中，有多少人是依照 $A \rightarrow B$ 路徑學習。特徵值總合則是將這些通過此路徑之舊學習者，其某項學習風格特性之特徵值總合。學習風格特徵平均值就是以特徵值總合除以使用其風格的總人數。

$A \rightarrow B$			
學習風格特性	頻率次數	特徵值總合	平均值
主動(act)	25	175	7
反思(ref)	7	35	5
感覺(sen)	11	33	3
直覺(int)	14	60	5
視覺(vis)	30	180	9
口語(ver)	10	30	3
連續(seq)	24	168	7
總體(glo)	17	85	5

圖 3 $A \rightarrow B$ 路徑之資訊

以圖 3 中的主動特性為例，其平均值為 $175/25=7$ 。在為每一條路徑計算完 8 項學習風格平均值後，即可為每條路徑計算學習風格相似程度。如(2)式所示，計算學習風格相似度需要風格平均值以及新學習者的學習風格特徵值，計算出之數值越小代表路徑之風格平均值與學習者之學習風格越近似：

$$Sim_{path(i,j)} = \sqrt{\sum_d (LS_{d,average} - LS_{d,learner})^2} \quad (2)$$

其中 $Sim_{path(i,j)}$ 代表 $i \rightarrow j$ 路徑上之學習風格相似度。 d 代表新學習者所測得之學習風格特性。 $LS_{d,average}$ 代表路徑上 d 項特性之平均值。 $LS_{d,learner}$ 代表新學習者其學習風格之 d 特性的特徵值。舉例來說，假

設有一新學習者，其學習風格為反思 5 分、感覺 7 分、視覺 11 分、連續 5 分，則該名學習者在 $A \rightarrow B$ 路徑上之學習風格相似度為：

$$\begin{aligned} Sim_{path(i,j)} &= \sqrt{\sum_{d \in (Ref, Sen, Vis, Seq)} (LS_{d,average} - LS_{d,learner})^2} \\ &= \sqrt{(5-5)^2 + (3-7)^2 + (9-11)^2 + (7-5)^2} \\ &= 4.899 \end{aligned}$$

經過計算，學習風格相似度為 4.899，相似度越接近 0 代表與通過路徑之學習者其學習風格越近似。權重值的第二個部份是學習教材所需要的時間。為了要讓學習者擁有足夠的時間學習，對學習所需之時間做出控管是有其必要性存在。這部份是透過測驗學習者對各個節點之了解程度以及預設的教材學習時間所得到。其評量方法如公式(3)所示：

$$T_{L,j} = (1 - U_j) * T_{R,j} \quad (3)$$

$T_{L,j}$ 是新學習者學習節點 j 真正所需要花的時間，而 $T_{R,j}$ 則是節點 j 預設所需學習的時間， U_j 則是學習者對節點 j 的了解程度，範圍在 0 到 1 之間，0 代表完全不了解該節點，而 1 代表完全了解該節點。利用對節點的了解程度以及預設節點所需要的學習時間來估計出學習該節點真正所需要的時間是多少。最後的權重值計算如公式(4)所示：

$$W_{path(i,j)} = W_{LS} * Sim_{path(i,j)} + W_{time} * T_{L,j} \quad (4)$$

$W_{path(i,j)}$ 為選擇路徑時所用到的路徑權重。 W_{LS} 是風格權重， W_{time} 則是時間權重，學習者可自行調整這兩項權重的比例，藉此調整學習風格與需要的學習時間對路徑選擇的影響程度。

最後的學習路徑將藉由計算權重後的學習路徑樣式所決定。路徑之抉擇將以路徑上之權重為主，以權重值較小之路徑為優先的選擇目標。如圖 4 所示，圖 4 為計算權重後的學習路徑樣式，系統推薦的適性化學習路徑為 $A \rightarrow C \rightarrow H \rightarrow E \rightarrow I \rightarrow L$ 。學習路徑之選擇，利用了結合學習風格與學習時間兩項因素之權重值。權重小的路徑代表依照該路徑學習之舊學習者與新學習者的學習風格較相似或新學習者對該節點之了解程度較大，需要的學習時間較短。而權重較大之路徑則代表著按照該路徑學習

之舊學習者與新學習者之學習風格差異較大，或是對於該節點之了解程度較少，需要更多的時間學習。在學習路徑產生後，系統會將所推薦之學習路徑存入學習者基本資料中，以作為提供學習者適性化之教材。

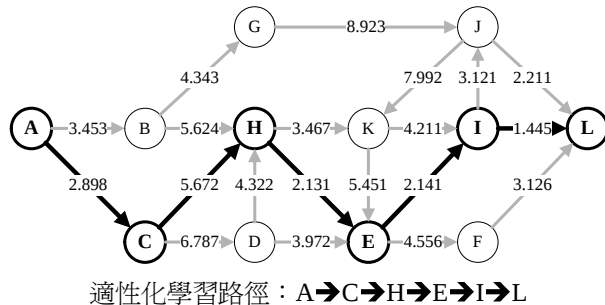


圖 4 選擇適性化學習路徑

4. 系統架構與實作

本節將介紹本研究的系統架構與實作方法。圖 5 為本系統之架構圖，本系統主要分成 4 個模組，分別是學習者資料管理模組、學習路徑產生模組、序列探勘模組以及教材呈現模組。下面將針對各個模組解說其功能：

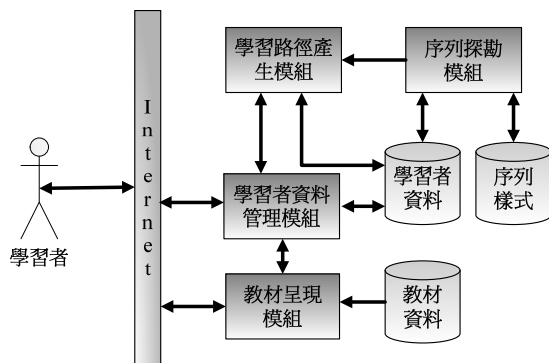


圖 5 系統架構

學習資料管理模組：本模組負責管理學習者的基本資料與各種學習相關資訊，此模組會針對系統所需的各種有關學習者的資訊加以測驗紀錄，並同時將學習者在學習中的點選紀錄存入學習者資料庫中。模組會針對新學習者測驗學習風格及課程中各個知識的了解程度，並將測驗結果存入學習者資料庫中。若學習者是舊的學習者，系統將讀取學習者資料，並將學習路徑傳送至教材呈現模組以支援該模組配置適當的教材。

序列探勘模組：此模組實作 ISM 序列探勘演算法，主要目的在於挖掘出經常出現之學習路徑樣式，以挖掘出之頻繁樣式作為未來路徑選擇之基礎，並於學習路徑產生模組要求資料時，回傳已挖掘出的序列樣式。模組會從學習者資料庫中取得新增的學習者學習路徑，並從序列樣式資料庫中讀取 ISL，針對更新的部分加以運算。

學習路徑產生模組：本模組主要的功能是結合序列探勘模組所得到的序列樣式與學習者的學習風格產生出適性化的學習路徑。模組會先讀取序列探勘模組所產生的序列樣式、教材所需之學習時間及舊學習者的學習風格特徵值，以先前所介紹之路徑建構方法產生路徑。最後所產生之路徑會存入學習者資料庫中，以供教材適性呈現模組使用。

教材適性呈現模組：本模組將從學習者資料中讀取學習者基本資料、學習路徑圖與系統推薦之學習路徑，並藉由適性化導覽技術提供學習者適當之教材，推薦適當之鏈結，同時將學習者之教材點選紀錄傳送至學習資料管理模組加以儲存，以供未來新學習者參考。同時藉由提供多種版本的教材並搭配學習風格與序列探勘，讓學習者可依照適當的順序研讀適合自己的教材，藉此達成有效的學習。

系統執行於 Windows 系統環境並使用 Tomcat 5.5 作為網頁伺服器。資料庫使用 MySQL，並採用 JSP 語言架構網站以呈現適當的教材。資料庫之存取使用 JDO，其實作版本採用 JPOX。學習教材採用「程式設計」課程為學習。系統主要是整合舊有學習者之學習路徑，並針對學習風格與學習時間找出適性化學習路徑。每個學習節點是由一到多個網頁所構成並且都會有一個預設學習時間。圖 6 為適性化學習路徑及其節點資訊。

5. 結論

本研究提出了一個適性化學習路徑的建置模式，並實作於數位學習系統上。研究採用了 Felder & Silverman 的學習風格理論，學習風格會影響學習者偏好某些教材特性，亦影響到學習路徑之抉擇。且因學習能力不同，精熟教材所需之時間亦有差異，

故不能僅考量學習風格，亦須考慮學習能力及所需之學習時間等因素。故採用序列探勘技術分析舊有學習者的學習路徑，以找出適當的學習教材與閱讀順序。同時考量到舊學習者的學習風格、新學習者之學習風格及學習教材所需要的時間，綜合上述特性選擇適性化之學習路徑。



圖 6 適性化學習路徑

在後續研究方面，本研究將針對學習路徑的推薦機制，加入序列比對之功能；將新學習者現在的學習路徑與頻繁序列互相比對，根據學習者現在的學習狀況來即時修正推薦的學習路徑。此外，針對學習知識能力的評量機制方面，將藉由加入試題反應理論(Item Response Theory, IRT)以便於精確地測驗出學習者的能力，利用所測得之能力值找出適當難度的學習路徑，讓系統所推薦的學習路徑可以更加符合學習者的需求與能力。

參考文獻

- [1] R. Agrawal and R. Srikant, "Mining sequential patterns", 11th International Conference on Data Engineering (ICDE'95), pp. 3-14, 1995.
- [2] C. I. Bennett, "Comprehensive Multicultural Education: Theory and Practice", Allyn and Bacon, 1999
- [3] P. Brusilovsky, "Adaptive Hypermedia", User Modeling and User-Adapted Interaction, Vol. 11, No. 1, pp.87-110, 2001.
- [4] J. B. Carroll, "A model of school learning", Teachers College Record, Vol. 64, No. 8, pp.723-733, 1963.
- [5] R. M. Felder and L. K. Silverman, "Learning and Teaching Styles in Engineering Education", Engineering Education, Vol. 78, No. 7, pp.674-681, 1988.
- [6] C. R. Morales, A. R. P. Pérez, S. V. Soto, H. Martínez and A. Z. Gómez, "Using sequential pattern mining for links recommendation in Adaptive Hypermedia Educational Systems", Fourth International Conference On Multimedia and Information and Communication Technologies In Education, 2006 of Conference.
- [7] A. Paramythis and S. Loidl-Reisinger, "Adaptive Learning Environments and e-Learning Standards", Electronic Journal of eLearning, Vol. 2, No. 1, pp.181-194, 2004.
- [8] S. Parthasarathy, M. J. Zaki, M. Ogihara and S. Dwarkadas, "Incremental and interactive sequence mining", Proceedings of the eighth international conference on Information and knowledge management, pp. 251-258, 1999.
- [9] R. E. Slavin, "A Model of Effective Instruction", Educational Forum, Vol. 59, No. pp.166-176, 1995.
- [10] B. A. Soloman and R. M. Felder, "Index of Learning Styles Questionnaire", <http://www2.ncsu.edu/unity/lockers/users/f/felder/public/ILSdir/ilsweb.html>,
- [11] R. Srikant and R. Agrawal, "Mining Sequential Patterns: Generalizations and Performance Improvements", Proceedings of the 5th International Conference on Extending Database Technology: Advances in Database Technology, pp. 3-17, 1996.
- [12] C. Wolf, "iWeaver: Towards' Learning Style'-based e-Learning in Computer Science Education", Proceedings of the fifth Australasian computing education conference on Computing education, pp. 273-279, 2003.