

# 利用試題反應建立學生學習歷程模型的一些經驗

劉昭麟

政治大學資訊科學系  
chaolin@nccu.edu.tw

## 摘要

高品質的學生模型有助於提高教學輔助系統所提供的協助的品質，是建構智慧型教學系統的重要核心元件。一般而言，我們可以請有經驗的教師直接提供模型。儘管如此，有一些特殊的問題，即使是有經驗的教師也不能確定模型的所有細節。本研究探討利用人工智慧技術，協助教師確認學生模型中無法直接觀測的子結構。我們以貝氏網路建構學生模型，利用模擬的學生答題資料，來猜測學生學習綜合觀念的學習歷程。實驗數據顯示，雖然學生的答題資料只能間接反應學生的能力，更不能忠實地反應學生的學習歷程，但是透過適當的技術，我們仍然有相當好的機會，可以利用這一些間接的證據協助教師找到學生學習歷程的模型。

**關鍵詞：**學生模型、適性化測驗、智慧型教學系統、貝氏網路、機器學習

## 一 緒論

因材施教是教學活動中一個重要的目標，專業教師透過與學生的互動和相關的專業知識，能夠判斷學生需要協助的地方，並且進行必要的導引。在智慧型教學輔助系統中，我們需要能夠把這裡所謂的相關專業知識，儲存在一個軟體系統可以運用的模型中。建構智慧型教學輔助系統時，如果能夠得到高品質的學生學習模型和教師教學模型，則有助於提高整體系統的效用。

一個專業教師的養成需要歷經多年的教育和實務的體驗，因此專業教師在教學活動中角色，並不易完全由電腦軟體來取代。另一方面，學生的學習狀態或者需要輔導的問題與源由，也常常是因人而異和因時而易。因此要建構一個完整的學生模型或者教師模型都不是一件輕而易舉的工作。

在這一篇文章中，我們專注於探討如何利用計算技術來猜測學生學習綜合觀念(composite concepts)的學習歷程[19]，進而討論利用機器學習技術來建構應用問題所需的模型的一些原則問題。經驗顯示學生的外顯行為並不一定忠實反映其內在的學習和成就狀態。最常在文獻[9,11]中被提到的現象是，學生在答題的時候有時會因為不小心而答錯(通稱為 *slip*)、有時會因為運氣好而答對(通稱為 *guess*)，不管是哪一種意外狀況，學生的答題反應都不必然反應他的能力。因此本文採用貝氏網路(Bayesian networks)[6,14]作為表示學習相關模型的計算工具，來捕捉這一些不確定的關係。

綜合觀念是一個需要學生整合兩個以上的基

本觀念或者相對比較基本的綜合觀念而成的觀念。例如進行異分母分數的加法的標準程序，是要能夠通分，然後把分子部分相加。要進行通分則需要先瞭解公倍數的概念，而要瞭解公倍數的概念，當然是要先建構倍數的概念。因此，倍數、公倍數、通分和同分母分數的加法這一些都是進行異分母分數加法的基本概念。然而，學生是否直接整合基本概念來習得綜合觀念卻是一個值得探索的問題。

我們以  $cA$ 、 $cB$ 、 $cC$  和  $cD$  分別來代表前面所說的倍數、公倍數、通分和同分母分數的加法四個基本概念；並且以  $dABCD$  來代表異分母加法這一個綜合觀念。則我們可以想像到至少兩種學習  $dABCD$  的方式，一個像是直接整合  $cA$ 、 $cB$ 、 $cC$  和  $cD$  成為  $dABCD$ ；另一個則是先把  $cA$  和  $cB$  整合成一個叫做  $dAB$  的綜合觀念，然後再把  $dAB$ 、 $cC$  和  $cD$  結合成  $dABCD$ 。

學生是用哪一種方式學到上述的  $dABCD$  呢？在回答這一問題之前，我們可能想要先思考兩個相關的問題。前面這一個問題的解答對於電腦輔助是否有實質的助益呢？專業教師是否能夠提供足夠精細的模型來分辨上述或者其他可以想像得到的學習歷程呢？

一個計算模型如果過於繁雜，以致於即使我們知道一些相關變數的狀態，也沒有相對配套的教學措施，則知道與不知道這一些變數對於教學活動的意義就不是非常顯著。因此，學生模型和教師模型的精確度與複雜度，有其實用面的考量。在綜合觀念的精細模型的需求方面，Carmona 等學者的經驗顯示，在一個多層次的貝氏網路中，如果所用的學生模型可以反應先備知識與高階知識之間的關係，則我們可以更有效率地進行適性化測驗[1]。基於這一個經驗，我們至少可以相信建構先備知識與較高階知識之間的關係的模型在實用上是值得的。

關於專業教師是否能夠就其專業上之經驗與知識，提供足夠精細的模型，顯然是因人而異也因問題複雜度而異。在電腦輔助教學相關的應用之中，許多案例都是由專業人員提供模型的架構，然後利用統計技術來求得模型所需的參數，例如在以試題反應理論[16]為基礎和以貝氏網路[12]為基礎的測驗機制的發展都有這樣的經驗。所以，專業人員可以提供一些有用的模型是有事跡可證而不需加以懷疑的。

儘管專業教師在提供學習相關的模型方面佔有不可或缺的重要角色，專業教師是否能夠提供一個足夠精細的模型，卻是尚且未有定論的一個議題。因此，在利用統計和機器學習技術學習系統參數之外，也有少數的研究者企圖利用機器學習技術建構跟教學活動相關的模型。Vomlel[17]利用學習

貝氏網路的技術，先從學生答題資料中找到大略的貝氏網路的結構，然後參考專家所提供關於如何處理潛在變數(hidden variables)的原則，再將先前所得的模型加以改進。Desmarais[4]等學者則是認為考慮貝氏網路中潛在變數會使得學習貝氏網路的複雜度過高，所以只有考慮試題與試題之間的網路結構，因此這一研究工作所建構的是所謂的『題間知識結構』(item-to-item knowledge structures)。

從貝氏網路的學習技術的角度[5,7,13]來看，以學生的答題反應來猜測學生如何習得綜合觀念時，如果我們已經明確定義了代表觀念的節點，則我們沒有尋找潛在變數的必要。在這一種情形之下，我們雖然不能直接蒐集任何潛在變數的數值，但是明確地知道貝氏網路中有哪些代表觀念的節點<sup>†</sup>。然而，在不能直接蒐集任何潛在變數的數值的情形之下，要猜測這一些潛在變數之間的結構[10,19]，不管從理論的角度[5,13]或者從實務的角度[4,17]來看，都是一個極具挑戰性的問題，也是Desmarais放棄考慮潛在變數轉而只考慮題間知識結構的一個重要原因[4]。

當專業教師可以提供學生學習綜合觀念的模型時，智慧型教學系統的建構者就可以直接將之納入系統之中。然而，當專業教師一時無法提供相關的模型時，我們是否就應該直接訴諸於機器學習技術來建構學生模型？上面所簡述的經驗顯示，直接依賴機器學習技術目前也不能提供簡單的答案。

因此，本文探討一個折衷方案的可行性。我們以專業教師所提供的資訊，配合機器學習技術，來協助專業教師從一些可能的候選模型中挑選出最好的學生模型。這一個方法介於完全依賴專業教師的知識與完全依賴機器學習的技術之間，以機器學習技術協助專業教師鑑別個別模型來尋找最好的模型。我們以模擬的學生答題資料來檢驗這一個想法的可行性，實驗數據顯示在專業教師所提供的教學相關資訊距離學生真實的學習狀態足夠相似時，機器學習技術確實能夠協助專業教師找到比較精細的學生模型；而在專業教師所提供的資訊的導引之下，我們也不會面對學理中學習任意貝氏網路的計算複雜度的問題。

本文第二節進一步說明與本研究相關的背景知識與相關定義。第三節簡報我們模擬學生的答題資料的方法。第四節簡要分析本研究問題的難度的一些相關議題。第五節提出應用機器學習技術來猜測學生學習歷程的方法。第六節詳細討論實驗的程序、結果與相關分析。第七節則進行綜合討論。

## 二 相關背景與定義

在這一節中，我們以比較正式的方式配合貝氏網路來定義我們所要解決的問題。令  $C_i$  代表問題中第  $i$  個觀念， $I_i$  代表我們用來測驗學生是否習得  $C_i$  的試題集合。假定某一測驗中，教師施測的題目牽涉了  $\gamma$  個觀念；我們以  $\Psi = \{C_1, C_2, \dots, C_\gamma\}$  來代表這一些觀念的集合。這一些觀念中有一些可能是**基本觀念**，有一些則是學生直接或者間接地融合基本觀

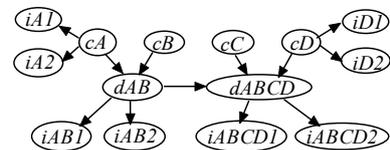
念所成的**綜合觀念**。以緒論裡面所提到的範例來舉例， $cA$ 、 $cB$ 、 $cC$  和  $cD$  是一些基本觀念，而  $dABCD$  則是一個綜合觀念。學生可能是直接整合  $cA$ 、 $cB$ 、 $cC$  和  $cD$  這四個基本觀念為  $dABCD$ ；不過也可能是先把基本觀念中整合成  $dAB$ ，最後才整合  $dAB$ 、 $cC$  和  $cD$  成  $dABCD$ 。依據我們的定義， $dAB$  仍然是綜合觀念。然而，當我們在探討學生學習  $dABCD$  的學習歷程時，如果學生是整合  $dAB$ 、 $cC$  和  $cD$  成  $dABCD$  的話，我們會特別稱呼  $dAB$ 、 $cC$  和  $cD$  是  $dABCD$  的**親代觀念**。一個綜合觀念是由它的親代觀念直接組合而成，親代觀念本身則可以是基本觀念或者是綜合觀念。

利用學生的試題反應來判定學生的學習歷程，就是不直接詢問學生自我所意識到的學習歷程，而只靠學生在考試中的答題表現來探測學生的學習歷程。因此，我們的系統的輸入資料，就是學生對於目前所探索的集合  $\Psi$  中的觀念的試題的答題資料；輸出資料則是系統所猜測的學習歷程模型的類別。

### 二.甲 以貝氏網路表示學生學習歷程

我們可以用貝氏網路中個別的節點(node)來代表學生是否擁有一個別的觀念和學生答覆個別的题目的對錯，由於答題的正確與否和學生是否具備相關的觀念有直接的關連，因此我們可以從代表觀念的節點拉一個連結(link)到代表測驗該觀念的試題的節點。依照前面對於親代觀念的定義，我們還需要從代表親代觀念的節點拉一條連結到代表其所從出的綜合觀念的節點。

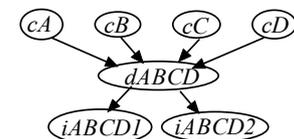
圖一是依據這一些原則所繪製的一個簡化的貝氏網路，這一個網路代表  $dABCD$  的親代觀念是



圖一：一個簡化的學生模型

$dAB$ 、 $cC$  和  $cD$  的信念。為了不讓圖形過度複雜，本圖假設每一觀念只有兩個直接相關的試題，同時也沒有畫出測驗  $cB$  和  $cC$  兩個觀念的試題。在本篇論文中，我們假設所有代表觀念的節點都只有兩個可能的狀態：**具備觀念**和**不具備觀念**。同時我們也假設所有代表試題反應的節點也是只有兩個可能的狀態：**答對**與**答錯**。

圖二是反應學生學習  $dABCD$  的另一種方式。跟圖一的網路結構所表示的信念不一樣，圖



圖二：直接融合基本觀念的模型

二這一個模型是表示學生把所有的**基本觀念**直接融合為  $dABCD$ 。因此，在這一模型所反應的信念之下， $dABCD$  的親代觀念完全是**基本觀念**。

### 二.乙 以 $Q$ 矩陣表示學生能力類型

學生的試題反應跟學生的能力有關，因此在我們的

<sup>†</sup> 因此，這一些變數是不可觀測的(unobservable)變數，而不是完全的潛藏(unknown)變數。

表一：只包含兩個學生類型的  $Q$  矩陣

|      | $cA$ | $cB$ | $cC$ | $dAB$ | $dAC$ | $dBC$ | $dABC$ |
|------|------|------|------|-------|-------|-------|--------|
| $g1$ | 1    | 1    | 1    | 1     | 1     | 1     | 1      |
| $g2$ | 0    | 1    | 1    | 1     | 1     | 0     | 1      |

模型中，也必須考慮不同能力的學生的類別的影響。我們借用 Tatsuoka[15]用以表示試題與答題所需的觀念的關連性的  $Q$  矩陣，來表示不同學生類別的能力[9]。表一是一個只有包含兩種學生能力類型的  $Q$  矩陣，表格的欄標題表示能力的類別，列標題則是學生的類別編號。為了行文方便，我們重複使用觀念的名稱作為能力的名稱。如果一個能力的名稱是一個基本觀念的名稱的話，則相對欄位為一的時候，代表該類別的學生具備該基本觀念，否則就不具備該基本能力。如果一個能力的名稱是一個綜合觀念的名稱的話，則相對欄位為一的時候，該類別的學生具有整合出該綜合觀念的能力。然而，這裡需要特別點出的是，具有整合綜合觀念的能力，不必然代表該類學生一定具有該綜合觀念的能力，這還要看該類別學生是否具有所需的親代觀念的能力。

以表一為例，第一個類型的學生是全能類型的學生，具有所有能力。第二個類型的學生則缺乏  $cA$  的能力，而且無法整合  $dBC$  的親代觀念成為  $dBC$ 。第二類的學生雖然能夠整合  $dAB$  的親代觀念成為  $dAB$ ；但是因為這一類學生缺乏  $cA$ ，所以這一個  $Q$  矩陣間接表示了第二類學生缺乏  $dAB$  的能力。

我們雖然可以利用  $Q$  矩陣的內容直接推論出某一類別的學生是否具備只牽涉到兩個基本觀念的綜合觀念的能力，但是我們並不能從  $Q$  矩陣來推斷某一類別的學生是否具備牽涉到三個或者更多基本觀念的綜合觀念的能力。

以上的討論，容易讓讀者誤以為學生的類別與其能力有決定性的關係，就如本文作者在[9]中所說明的一樣，我們引用一個控制變數，讓  $Q$  矩陣所表示的只是類別中典型學生的表現，個別學生仍然有機會展現不同於典型學生的表現。在本文中，我們以  $\beta$  這一個變數來取代[9]中所討論的  $groupSlip$  和  $groupGuess$  兩個變數。

### 三 模擬器簡介

我們沿用[9]所討論的模擬器，給定一個  $Q$  矩陣、一個貝氏網路結構和一些控制變數，我們就可以利用我們的模擬器產生學生的答題反應。這一個模擬器依據  $Q$  矩陣的限制，以 noisy-and 的模型[14]為貝氏網路中的觀念節點建構條件機率表(conditional probability tables)。依照第二節的說明，我們的貝氏網路的節點分成三大類：代表是否具備觀念的節點、代表答題正確與否的節點和代表學生所屬類型的節點。

就如第二節所提到的，我們利用  $\beta$  引入個別學生的表現跟學生類別中典型學生的表現有可能有所差異的機制。 $\beta$  的數值愈大，則個別學生有愈高的機會偏離典型學生的表現。

建構代表試題反應的條件機率表的時候，我們利用  $\alpha$  這一個變數來引入緒論中曾經提到的  $slip$

和  $guess$  發生的機率。 $\alpha$  的數值愈大，則學生答題的時候發生  $slip$  和  $guess$  的機會就愈大。

我們的模擬器可以模擬多種可能的學生類型的分佈，目前我們是採用均等分佈(uniform distribution)。如果實驗需要的話，我們也可以輕易地採用其他的學生類型分佈。

為所給定的貝氏網路的每一個節點產生了所對應的條件機率表之後，我們就可以利用蒙地卡羅法(Monte-Carlo method)為每一個學生類型產生所需要的學生答題反應。這一個模擬器產生模擬資料的細節請參見[9]的討論。

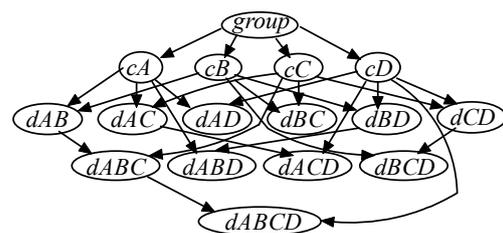
### 四 問題難度的分析

假定我們所要探討問題是一個牽涉  $\lambda$  個基本觀念的綜合觀念的學習歷程。為了簡化問題，我們進一步假設任何綜合觀念的親代觀念都不會牽涉任何相同的基本觀念。舉例來說，在我們現階段的研究中， $dABCD$  的親代觀念不可以是  $dABC$  和  $dCD$ ，因為這兩個觀念同時牽涉到  $cC$ 。

基於這一個假設，因為這一個綜合觀念的親代觀念可以有兩個、三個...甚至  $\lambda$  個。以一個箱子當作一個親代觀念，所以一個綜合觀念的學習歷程的可能個數，就是把  $\lambda$  個基本觀念分派到兩個、三個...甚至  $\lambda$  個箱子，使得每一個箱子都至少有一個基本觀念的總個數。這一個總數跟第二類 Stirling 數[8]有密切的關係。下面的公式(1)顯示，這一個總數會隨著  $\lambda$  的增大而迅速地增大，因此我們所面臨的可能的答案的數量是極大的。

$$\Omega(\lambda) = \sum_{i=2}^{\lambda} \left( \frac{1}{i} \sum_{j=0}^{i-1} (-1)^j \binom{i}{j} (i-j)^{\lambda} \right) \quad (1)$$

圖三是  $\lambda = 4$  時的一個可能的學習歷程，為了讓圖形維持基本的可讀性，這一個圖形結構沒有畫任何代表試題的節點。利用公式(1)計算  $dABCD$  所有的可能學習歷程時，所得的結果是 14。但是這一個數字僅僅只有考慮  $dABCD$  的學習歷程的數目，並沒有考慮其他牽涉到三個基本觀念的綜合觀念的學習歷程的可能個數。如果我們應用公式(1)來計算的話，一個牽涉到三個基本觀念的綜合觀念的學習歷程的可能個數是 6 個。因為圖三有四個牽涉到三個基本觀念的綜合觀念，所以依照這一些推算，如果我們沒有先探討那些只有牽涉三個基本觀念的綜合觀念的學習歷程，就直接研究牽涉四個基本觀念的綜合觀念的學習歷程的話，我們會直接面對總共  $18144(=\Omega^4(3) \times \Omega(4) = 6^4 \times 14)$  種可能的結構。以公式(1)的符號來定義的話，當我們直接處理牽涉到  $\lambda$  個基本觀念的綜合觀念時，我們將面對



圖三：一個  $\lambda = 4$  的可能結構 ABC~D

公式(2)所代表的這一個極大數量的候選結構。

$$\prod_{k=3}^{\lambda} (\Omega(k))^{\binom{\lambda}{k}} \quad (2)$$

除了學習綜合觀念的可能的方法的總數很大之外，我們所面臨的可能學生類型的數量也是極大。利用 $\lambda$ 個基本觀念可以組合出來的所有的基本觀念和綜合觀念的總數有 $(2^{\lambda}-1)$ 種。以 $\phi$ 代表這一個總數。在假設學生的學習狀態只能是會某種觀念或者不會的情形之下，我們最多可以有 $2^{\phi}$ 種的學生類型。再以 $\kappa$ 代表 $2^{\phi}$ 。除非我們所使用的 $Q$ 矩陣包含這所有的 $\kappa$ 種的學生類型；否則，如果我們要選擇一部份的學生類型來進行實驗，則我們就面臨了如何從這 $\kappa$ 種的學生類型選出一個子集合來進行實驗的問題，這時候，我們就會有 $(2^{\kappa}-1)$ 種可選用的 $Q$ 矩陣。總而言之，在研究一個牽涉到 $\lambda$ 個基本觀念的綜合觀念時，下面的公式(3)提供我們所有可能的 $Q$ 矩陣的總數。即使 $\lambda$ 只不過是4，我們也面對高達 $(2^{32768}-1)$ 種不同的 $Q$ 矩陣。

$$2^{2^{(2^{\lambda}-1)}} - 1 \quad (3)$$

在實務上，並非所有公式(1)的貝氏網路結構和公式(3)的 $Q$ 矩陣都能夠得到專業教師的認可，因此上面的分析其實僅僅是資訊科學中所謂最壞狀況的分析而已。然而，這一分析確實也讓我們窺知問題的規模和可能的難度。

## 五 分類器的設計

基於第四節關於公式(2)的分析，我們應該先確認所牽涉的基本觀念的數目比較少的綜合觀念的學習歷程，然後逐步探討更加高階的綜合觀念的學習歷程。這樣我們所面對的可能的結構的總數，就是公式(1)的數目，雖然這一個總數仍然是隨著 $\lambda$ 的增大而快速增大，但是還是比直接面對公式(2)所代表的總數要好得多。

為了簡化文字，我們將以特殊的符號來表示一個個可能的貝氏網路結構。以圖三為例，在研究 $dABCD$ 的學習歷程時，假定我們已經確認了所有只有牽涉到三個基本觀念的綜合觀念的學習歷程，所以我們只要針對學習 $dABCD$ 的14種方式即可，圖三就是其中的一種。同時，因為這一個結構中， $dABCD$ 的親代觀念是 $dABC$ 和 $cD$ ；所以我們用 $ABC \sim D$ 這一個簡單的符號來代表這一個結構。應用同樣的規則，我們以 $AB \sim C \sim D$ 和 $A \sim B \sim C \sim D$ 分別代表圖一和圖二的結構。

要從公式(1)所表達的數量的可能網路結構中，找出最足以代表學生學習歷程的結構，除了直接由專業教師提供或者是全部由機器學習技術代勞之外，這一篇論文所要探討的是一個折衷方案：由專業教師提出幾個可能的候選結構，然後由機器學習技術來評比這一些候選結構。

### 五.甲 引入與估計互見資訊

再以圖三的结构作為討論的例子，如果我們知道結構中的節點的狀態(state)，則剛剛所陳述的問題的重點，其實只要透過檢驗各個節點的條件獨立

(conditionally independent)關係是否成立，就有一個解決的方案。然而，圖三裡面的節點代表的都是潛在變數，我們實際上不能直接得到這一些節點的狀態，我們所能夠蒐集的資訊是學生的答題表現。由於圖三的结构並沒有畫出這一些代表答題反應的節點，所以我們可以再參考圖一的结构。因為考慮了前面說明過的 $slip$ 和 $guess$ ，代表觀念能力的節點和代表答題反應的節點之間是一個機率分佈；所以雖然我們可以用學生的答題反應來推測學生的能力，但是由於這樣的機率關係，我們並不能用試題節點間的條件獨立關係成立與否，來確認觀念節點間的條件獨立關係是否成立[10]。

觀察圖一、圖二和圖三的结构，我們可以發現，在直覺上學生學習的歷程應該會影響答題對錯的關連性。比較直接相關的觀念的試題的對錯、比起相對關連性較弱的觀念的試題的對錯，應該有比較高的關連性。基於這一個觀點，如果正確的结构是圖三的話，則學生回答關於 $dABCD$ 的試題的對錯，應該跟學生回答關於 $dABC$ 和 $cD$ 的試題的對錯有比較密切的關係，而跟學生回答關於 $dAB$ 、 $cC$ 和 $cD$ 的試題的對錯的關係比較弱。

我們利用資訊理論裡面的互見訊息(mutual information)[3]，即公式(4)的 $MI(X;Y)$ ，作為量測 $X$ 和 $Y$ 這兩群機率變數之間的關連性，公式裡面的 $\hat{X}$ 和 $\hat{Y}$ 分別代表 $X$ 和 $Y$ 兩群機率變數各自所有可能的數值。

$$MI(X;Y) =$$

$$\sum_{x \in \hat{X}} \sum_{y \in \hat{Y}} \Pr(X=x, Y=y) \ln \frac{\Pr(X=x, Y=y)}{\Pr(X=x) \Pr(Y=y)} \quad (4)$$

舉例來說，如果 $MI(cA, cB, cC, cD; dABCD)$ 大於 $MI(dABC, cD; dABCD)$ 的話，則表示 $cA$ 、 $cB$ 、 $cC$ 和 $cD$ 四者跟 $dABCD$ 有比較直接的關係，因此相較之下，會認定學生的答題表現比較支持圖二所代表的學習歷程。

然而，就如我們剛剛已經討論過的，我們並沒有學生們會或者不會相關觀念的資料，我們只有學生的答題反應。所以，雖然學生的答題反應並不必然反應他們的能力，但是我們也只能利用他們的答題反應去猜測他們的能力。

假定某一次的測驗中，每一個觀念都有 $n$ 道直接相關的試題，則學生答對關於某一觀念的試題的比例就有 $0$ 、 $1/n$ 、 $2/n$ 、 $\dots$ 、和 $1$ 總共 $(n+1)$ 種可能。我們可以用全體學生對於某一觀念的試題的答對比例的分佈來暫代全體學生在這一觀念的能力的分佈。我們當然也可利用學生對於某一些觀念組合的答題反應，來估計學生對於那一些觀念組合的能力的聯合機率分佈(joint probability distribution)，公式(5)是一個簡單的範例。

$$\Pr(cA = 1/n, dBC = 2/n) = \frac{\text{答對1題關於}cA\text{和2題關於}dAB\text{的試題學生數}}{\text{受測學生總數}} \quad (5)$$

跟所有以估計的機率作為推理核心資訊的系統一樣，我們也需要處理零機率的問題，目前我們是把所有機率變數的基本狀態的出現次數先加上一個很小的量(目前是用 $0.001$ )，作為我們平緩

(smoothing)機率分佈的方法。

有了估計個別觀念和觀念組合的能力的機率分佈情形的方法，我們就可以套用互見資訊的定義，來計算可以用來評比不同貝氏網路結構所需的分數。以 $\lambda$ 是4的問題來說，我們需要為所有14種可能的學習歷程，計算一個分數，作為評比的基礎。

## 五.乙 機器學習技術

然而，在進行一些小規模的實驗之後，我們發現，雖然互見資訊的效果不錯[19]，不過有一些時候代表不同學習歷程的分數有一些時候並沒有很大的差距，這時候互見資訊的大小的可靠度就變得不高。因此，除了原始的互見資訊的資料之外，我們還建議找出所有基本的互見資訊的最大值、次大值和平均值，把所有基本的互見資訊除以最大值，當作決策時的特徵項目(features)。此外，把最大值除以最小值和最大值除以平均值也是兩個可能有用的特徵項目。

以研究 $dABCD$ 的學習歷程為例，我們需要計算14個基本的互見資訊，把這14個互見資訊除以最大值之後，我們會得到另外14個新的特徵，最後再把最大值除以次大值，並且把最大值除以平均值。所以我們一共會產生30個可以用以評比不同結構的特徵項目。

我們可以蒐集受測學生的答題反應，然後利用上述這一些有用的特徵來評比不同結構，從中選出一個最適當的。然而，經驗[10,19]顯示，這樣的作法雖然有其優點，但是當學生的外在表現與其內在能力的關連性相對不高的時候，直接利用這一些特徵項目來猜測學生的學習歷程的效果並不盡理想。以第二節所說明的詞彙來說明，當我們用比較大的 $\alpha$ 和 $\beta$ 來產生學生答題反應的模擬資料時，並且用上述的特徵項目來猜測產生這一些模擬資料的學習歷程時，所得的效果並不好。

本文所討論的方法，是假設我們有資料來源讓我們可以預先找出幾個可能的學習歷程，配合一個選定的 $Q$ 矩陣，來產生學生的答題反應。由於在這一階段是我們選擇特定的學習歷程來產生這一些答題反應，所以我們可以把已知的學習歷程類別當作是訓練資料的類別標籤(class labels)，配合所產生的答題反應當作是監督式學習(supervised learning)的訓練資料，來產生一些分類器，最後再以這一些分類器來猜測真實的學生的學習歷程的類別。在這一理念之下，我們可以採用支持向量機(support vector machines)、類神經網路(artificial neural networks)或者其他可用的機器學習技術來建構我們的分類器。在下一節中，我們採用支持向量機作為猜測學生學習歷程的分類器。

## 六 實驗程序、結果與分析

由於我們並沒有蒐集真實的學生的答題資料，因此我們全部都是採用模擬的學生資料進行實驗。在一組實驗中，我們需要兩組貝氏網路，其中有一組只有一個貝氏網路，是用來代表學生真正的學習歷程，我們用這一個貝氏網路來產生測試分類器的學生答題資料。另一組的貝氏網路包含多個貝氏網

路，分別代表我們所猜測的學習歷程，我們利用這一組貝氏網路分別產生許多組的學生答題資料，配合所選定的 $Q$ 矩陣，以第三節所描述的模擬器產生模擬的學生資料，然後應用第五節所描述的原則訓練我們的分類器。我們以訓練所得的分類器來猜測產生測試資料時所用的貝氏網路的結構，同時記錄和分析正確的比例。

在目前的實驗中，我們假設所有學生都是採用同一個學習歷程來學綜合觀念的。這樣一來，我們的實驗只要找出最有可能的學習歷程即可，可以簡化實驗的設計和分析。當然，這樣的假設不是絕對必要的，如果我們認為不同學生可能用不同的方式學習某一綜合觀念的話，則我們可以讓我們的系統找出前幾名最有可能的結構即可。

## 六.甲 實驗所用的貝氏網路結構

我們進行四組實驗，分別假設學生的學習類型是 $A\sim BCD$ 、 $AB\sim CD$ 、 $A\sim BC\sim D$ 和 $A\sim B\sim C\sim D$ 。除了 $dABCD$ 有這四個可能的結構之外，我們假設所有其他綜合觀念的親代觀念都是基本觀念。從親代觀念的個數與組合上，這四個可能的答案分別代表四大類不同的情況。 $dABCD$ 的親代觀念有兩個時，有兩類情況，也就是兩個親代觀念分別牽涉一個和三個基本觀念，或者是兩個親代觀念都牽涉兩個基本觀念。 $dABCD$ 的親代觀念有三個時，只有一類情況，只能有一個親代觀念牽涉到兩個基本觀念，另外兩個都必須是基本觀念。 $dABCD$ 的親代觀念有四個時，則每一個親代觀念都必須是基本觀念。

我們暫時假設依據專業資訊所猜測的貝氏網路結構群，總是能夠包含正確的結構。為了能夠同時比較正確結構對於實驗結果的影響，我們就假設依據專業資訊所猜測的貝氏網路結構群是 $\{A\sim BCD, AB\sim CD, A\sim BC\sim D, A\sim B\sim C\sim D\}$ 。所以不管實驗時真正的答案是上一段落所列的哪一個，全部都屬於這一個集合。

## 六.乙 實驗所用的 $Q$ 矩陣

我們在第三節所討論的模擬器，除了需要貝氏網路的結構之外，還需要再提供關於學生類型的 $Q$ 矩陣。在我們完全依賴模擬資料所進行的實驗中，觀念上需要兩個 $Q$ 矩陣：第一個 $Q$ 矩陣代表依照某種專業資訊所猜測的學生能力類型，第二個 $Q$ 矩陣則是代表真實學生的類型。如果專業資訊是完全可靠的話，則這兩個矩陣就會相近甚至相同。在目前的實驗中，我們暫時假設這兩個矩陣是完全相同的。當然，在真實的研究中，即使是基於專業資訊所臆測的 $Q$ 矩陣也不一定能夠精確掌握學生的類型。這一部份的不確定性，部分可以由第三節解釋過的控制變因 $\beta$ 來模擬。但是，比較嚴格的考驗是真的以兩個不同的 $Q$ 矩陣來進行實驗，這一部份作者已經嘗試過，不過由於論文篇幅的關係，不能在這一短篇短文中報告結果。

下一頁上的表二是本文所採用的 $Q$ 矩陣。它的格式跟表一的格式類似，最左邊的欄位是學生類型的編號，其他欄位代表基本能力的有無和是否具備產生綜合觀念的綜合能力。這一個 $Q$ 矩陣跟[10]一文所使用的 $Q$ 矩陣完全一樣，不過本文與[10]

表二：實驗所採用的  $Q$  矩陣

|       | $cA$ | $cB$ | $cC$ | $cD$ | $dAB$ | $dAC$ | $dAD$ | $dBC$ | $dBD$ | $dCD$ | $dABC$ | $dABD$ | $dACD$ | $dBCD$ | $dABCD$ |
|-------|------|------|------|------|-------|-------|-------|-------|-------|-------|--------|--------|--------|--------|---------|
| $g1$  | 1    | 1    | 1    | 1    | 1     | 1     | 1     | 1     | 1     | 1     | 1      | 1      | 1      | 1      | 1       |
| $g2$  | 1    | 1    | 1    | 1    | 1     | 0     | 0     | 0     | 0     | 1     | 1      | 1      | 1      | 0      | 1       |
| $g3$  | 1    | 1    | 1    | 1    | 0     | 1     | 0     | 1     | 0     | 1     | 1      | 1      | 0      | 1      | 1       |
| $g4$  | 1    | 1    | 1    | 1    | 1     | 0     | 0     | 0     | 0     | 1     | 1      | 1      | 0      | 0      | 1       |
| $g5$  | 1    | 1    | 1    | 1    | 1     | 0     | 1     | 0     | 1     | 0     | 1      | 0      | 1      | 1      | 1       |
| $g6$  | 1    | 1    | 1    | 1    | 1     | 0     | 0     | 0     | 0     | 1     | 1      | 0      | 1      | 0      | 1       |
| $g7$  | 1    | 1    | 1    | 1    | 1     | 1     | 1     | 0     | 0     | 0     | 1      | 0      | 0      | 1      | 1       |
| $g8$  | 1    | 1    | 1    | 1    | 1     | 0     | 0     | 0     | 0     | 1     | 1      | 0      | 0      | 0      | 1       |
| $g9$  | 1    | 1    | 1    | 1    | 0     | 0     | 0     | 1     | 1     | 1     | 0      | 1      | 1      | 1      | 1       |
| $g10$ | 1    | 1    | 1    | 1    | 1     | 0     | 0     | 0     | 0     | 1     | 0      | 1      | 1      | 0      | 1       |
| $g11$ | 1    | 1    | 1    | 1    | 1     | 1     | 0     | 0     | 1     | 1     | 0      | 1      | 0      | 1      | 1       |
| $g12$ | 1    | 1    | 1    | 1    | 1     | 0     | 0     | 0     | 0     | 1     | 0      | 1      | 0      | 0      | 1       |
| $g13$ | 1    | 1    | 1    | 1    | 0     | 0     | 1     | 1     | 0     | 0     | 0      | 0      | 1      | 1      | 1       |
| $g14$ | 1    | 1    | 1    | 1    | 1     | 0     | 0     | 0     | 0     | 1     | 0      | 0      | 1      | 0      | 1       |
| $g15$ | 1    | 1    | 1    | 1    | 0     | 0     | 0     | 0     | 1     | 1     | 0      | 0      | 0      | 1      | 1       |
| $g16$ | 1    | 1    | 1    | 1    | 1     | 0     | 0     | 0     | 0     | 1     | 0      | 0      | 0      | 0      | 1       |

所用的貝氏網路的結構並不相同，所以實驗結果可以相互比較。

表二中 16 類的學生類型都具備基本能力，而且有能力整合學習  $dABCD$  的親代觀念的能力，所以都可以遂行  $A\sim B\sim C\sim D$  這一種學習模式。其中， $i$  是奇數的  $g_i$  類別，都是可以運用  $A\sim BCD$  的學習模式。類別  $g1$ 、 $g2$ 、 $g4$ 、 $g6$ 、 $g8$ 、 $g10$ 、 $g11$ 、 $g12$ 、 $g14$  和  $g16$  可以運用  $AB\sim CD$  的模式。類別  $g1$ 、 $g3$ 、 $g9$  和  $g13$  可以運用  $A\sim BC\sim D$  的模式。

從觀念的類別來分析表二的設定的話，上面已經提到所有學生類別都具備基本能力，而且有能力整合學習  $dABCD$  的親代觀念的能力。如果我們的目標是要找尋學生學習  $dABCD$  的方式的話，模擬一些至少有機會能夠學會  $dABCD$  的學生的表現應該是合理的。 $dABC$ 、 $dABD$ 、 $dACD$  和  $dBCD$  這四個牽涉到三個基本觀念的綜合觀念，總共可以有 16 種不同的能力組合，這也是表二包含 16 類學生的主要原因。由於組合數量的原因，表二並沒有能夠包含所有牽涉到兩個基本觀念的綜合觀念的所有能力組合，因為那將需要 64 個類別，我們任選了 16 種方式來設定這一些能力的組合。

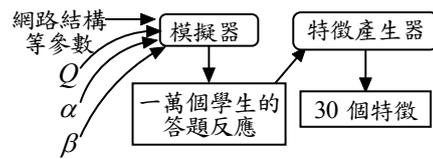
我們在第四節曾經討論過  $Q$  矩陣的數量的問題， $Q$  矩陣的選擇必然會影響實驗的結果，除了上面所說，個別學生類型可以支持所要試驗的學習歷程之外，我們還有其他方式來分析  $Q$  矩陣的影響，作者將於近日發表關於這一些比較深入的議題的討論。

除了透過控制變因  $\beta$  的輔助，來產生一些學生類別的不確定性之外，我們還利用控制變因  $\alpha$  來模擬學生答題時發生 *slip* 和 *guess* 的機率。在我們的實驗中， $\alpha$  和  $\beta$  這兩個控制變因會分別被設定為 0.05、0.10、0.15、0.20、0.25 和 0.30 這六個不同的值。在文獻中，尚未看到有討論 *slip* 和 *guess* 發生機率超過 0.3 的必要。在我們的模擬器中  $\alpha$  和  $\beta$  的大小是發生「異常」表現的機率的上限。發生異常的確實機率是透過一個亂數產生器所設定，而這個亂數產生器以均等的機會提供  $[0,1]$  這一個區間之內任何的實數。所謂「異常」表現是指發生

*slip*、*guess* 或者整體表現不像所屬類別的典型學生這一些情形。

### 六.丙 基礎流程

圖四提供實驗基礎流程的簡圖。特徵產生器利用第五節所說明的方法，依據一萬個學生答題的平均狀況來估計代表觀念的節點的狀態，然後計算所需要的互見資訊和特徵。在我們的實驗中，我們假定學生針對每一個觀念都回答了三道試題，以第二節的詞彙來說的話，對測試範圍之內所有的觀念，我們都假設  $|I_i| = 3$ 。我們會針對每一個可能的結構，也就是  $A\sim BCD$ 、 $AB\sim CD$ 、 $A\sim BC\sim D$  和  $A\sim B\sim C\sim D$ ，執行圖四的流程；每一個結構都會執行 600 次。針對任一特定  $\alpha$  和  $\beta$  的組合，我們有 2400 筆案例，我們從每一結構的 600 個案例裡面任選 500 筆案例當作訓練資料，以剩下的 100 筆作為測試資料。



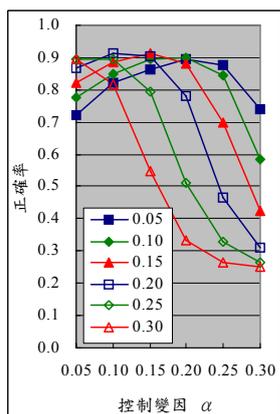
圖四：實驗的基礎流程

這裡所說的訓練資料，就是第五.乙節所討論到的利用專業資訊所直接提供或者所產生的學生答題表現。測試資料則是指該小節所說的真實的學生答題反應。

我們使用 LIBSVM[2] 作為實踐支持向量機的工具，我們使用 c-SVC 型態的支持向量機，採用 radial basis 核心函數(kernel)。以下所有以支持向量機所產生的實驗結果，都是先以訓練資料自我測試，找出在 0.1、0.2、...、1.9 之間最好的  $C$  和  $\gamma$  兩個參數的組合。在比較了總共 361 種組合之後，我們找出一個最好的組合之後，再用這一個組合來猜測產生真正的測試資料的貝氏網路結構。

## 六.丁 基本分析

我們先用一個簡單的程序分析一下所要使用的原始資料。我們先用第五.甲節所說明的方法，以所產生的資料中的互見資訊的大小，來猜測產生各筆資料所使用的貝氏網路結構。經過一個簡單的程序，我們可以計算在不同的 $\alpha$ 和 $\beta$ 的組合下，總共 2400 筆中，直接以互見資訊來猜測原始結構的正確率(accuracy)。我們直接比較代表 14 種不同結構的互見資訊的大小，以最大的互見資訊所代表的結構當作我們所猜測的結構，然後檢視所猜測的結構是否就是原本用以產生該筆資料的結構。



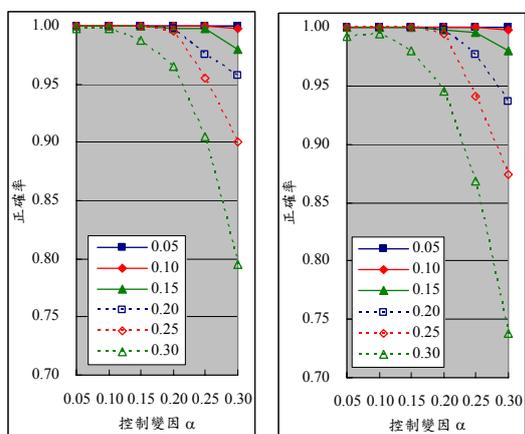
圖五：實驗資料的基本分析

圖五的縱軸是正確率的大小，橫軸是控制變因 $\alpha$ 的值，圖例則是控制變因 $\beta$ 的值。我們可以輕易地由圖五的曲線發現，如果只有依賴第五.甲節所陳述的方式，以估測的互見資訊來猜測正確的結構並不容易達到很好的效果。在所測試的範圍之內，隨著控制變因的組合的變化，最好的效果接近九成，而最壞的時候則低到接近只有兩成五。

這裡的兩成五並非因為我們系統是從四個已知類別中隨意猜測而得的。實際上我們的系統是比較 14 種結構的相對可能性，所以如果是隨意亂猜的話，最壞的正確率可以低到一成以下。這裡的兩成五是因為 A~B~C~D 是一個競爭力非常強的候選結構，在 $\alpha$ 和 $\beta$ 兩個控制變因都很大的時候，我們的系統常常會猜測學習歷程是 A~B~C~D。所以在真實答案真的是 A~B~C~D 的時候，我們的系統都還是會猜對這 600 筆測試案例的類別，因此表現出接近兩成五的結果。

## 六.戊 實驗結果與分析

圖六的左圖是我們執行第六.丙小節所描述的程序



圖六：找出正確結構的比例隨控制變因而異

所得的實驗結果。圖六裡面兩個圖的縱軸、橫軸和圖例的意義跟圖五都是一樣的。

比較圖五和圖六的左圖的曲線，我們發現，當我們可以利用專業知識限縮可能的結構的範圍的時候，機器學習技術確實可以大幅提高找到潛藏的學習模型的機會。儘管在 $\alpha$ 和 $\beta$ 兩個控制變因都很大的時候，我們的系統表現仍然是相對的不好，不過比起上一小節所做的基本分析來說，利用機器學習技術已經大幅提高了猜到正確學習歷程的比例。

除了大幅提高整體的正確率之外，透過限縮可能的答案的範圍，圖六的曲線的趨勢也比較規則化，正確率大都隨著 $\alpha$ 和 $\beta$ 兩個控制變因的變大而降低。

以上的實驗中，我們選定了 A~BCD、AB~CD、A~BC~D 和 A~B~C~D 作為可能的答案的集合。如果我們加入 A~B~CD，並且重複同樣的實驗程序，我們得到圖六的右圖這一個結果。比較圖六的兩個圖形，我們觀察到增加可能的答案的範圍對於我們的系統的影響。

除了以正確率來量測實驗結果之外，我們也可以用精確率(precision)與召回率(recall)所組合而成的 F 分數(F measure) [18]來量測實驗結果。由於我們現在的可能答案有四類，所以可以為四個類別分別計算精確率和召回率，並且分別計算四個精確率和四個召回率的平均值，最後賦予平均精確率和平均召回率相同的權重來計算 F 分數。結果顯示，對於我們所執行過的實驗來說，使用正確率和 F 分數來量測實驗結果並沒有顯著的不同。以繪製圖六左圖的數據作為分母，以同一實驗結果所計算出的 F 分數作為分子來比較兩者的差異，我們所得的 36 組比例中最小的是 1.0000，最大的則是 1.0015。F 分數雖然都大於正確率，但是兩者差異極小。

## 七 討論

本研究的主要目的是要檢驗我們是否能夠利用基於專業經驗所提供的資訊，來協助我們從極大量數的可能答案中找到潛在變數之間的關連性。這一個觀念可以應用到許多領域；在本文中，我們以找尋無法直接觀測的學習歷程為例，實驗結果顯示本研究所欲驗證的研究法是可行的。

在這樣的短文中，其實有許多問題是無法仔細交代清楚的。首先，一個以模擬方式所得的結論是否意味著未來以真實的數據從事研究時，我們會得到相同的結論？跟大部分工程技術的研究一樣，模擬系統所採用的模型必然都簡化了現實世界中的某一些狀況，因此模擬系統所得的結論確實是需要接受進一步檢驗的。不過，模擬系統的結果和經驗應該可以提供設計未來的實驗的參考。

前面幾個小節所報告的實驗設計和結果，還有許多其他面向需要進一步檢驗。比較本文和[10]所報告的結果，我們可以看到不同的答案集合對於實驗結果的影響。 $Q$  矩陣的設定直接影響了所產生的貝氏網路所採用的機率分佈，因此必然影響實驗的結果。因此如何選定  $Q$  矩陣當然是一個需要討論的議題。

此外，基於專業資訊所猜測的可能答案和真實

的學生的能力狀態，兩者是否真的如第六.甲節和第六.乙節所假設的一樣：兩者是完美的吻合？如果我們所猜測的可能答案的集合沒有包含學生真正的學習模式，則監督式學習法當然是派不上用場了。不過，我們相信專業資訊應該可以提供合理的猜測，不讓這樣極端的情形發生。相對而言，要猜到正確的  $Q$  矩陣就比較不容易。如果訓練支持向量機時所用的  $Q$  矩陣不能夠合理反應真實的學生能力，對於研究結果又會有怎樣的影響？本文作者將另文討論以上這一些問題。

以本文所應用的範圍來說，以機器學習法找尋學生學習歷程確實有一個很難解的難題。我們所賴以進行分類的依據是來自於一個學生群體所表現的行為的統計數據，所以如果學生群體的人數不多，則所得的統計數據是否可靠，將是一個很大的挑戰。一個極端的難題是，本文所討論的方法恐怕很難用來找出那一些具有獨特學習方式的學生的學習歷程。

### 致謝

本研究承蒙國科會 NSC-94-2213-E-004-008 和 NSC-95-2213-E-004-013 研究計畫的持續資助與教育部頂尖大學計畫之部分補助，謹此致謝。

### 參考文獻

1. C. Carmona, E. Millán, J. L. Pérez-de-la-Cruz, M. Trella, and R. Conejo, Introducing prerequisite relations in a multi-layered Bayesian student model, *Proc. of the 10th Int'l Conf. on User Modeling*, 347–356, 2005
2. C.-C. Chang and C.-J. Lin, LIBSVM: A library for support vector machines, 2001, <http://www.csie.ntu.edu.tw/~cjlin/libsvm>
3. T. M. Cover and J. A. Thomas, *Elements of Information Theory*, John Wiley & Sons, 1990
4. M. C. Desmarais, P. Meshkinfam, and M. Gagnon, Learned student models with item to item knowledge structures, *User Modeling and User-Adapted Interaction* **16**(5), 403–434, 2006
5. D. Heckerman, A tutorial on learning with Bayesian networks, in [7], 301–354, 1999
6. F. V. Jensen and T. D. Nielsen, *Bayesian Networks and Decision Graphs*, Springer, 2007
7. M. I. Jordan, ed., *Learning in Graphical Models*, The MIT Press, 1999
8. D. E. Knuth, *The Art of Computer Programming: Fundamental Algorithms*, p. 73, Addison-Wesley, 1973
9. C.-L. Liu, Using mutual information for adaptive item comparison and student assessment, *J. of Educational Technology & Society* **8**(4), 100–119, 2005
10. C.-L. Liu, Learning students' learning patterns with support vector machines, *Lecture Notes in Computer Science* 3488, 601–611, 2006
11. E. Millán and J. L. Pérez-de-la-Cruz, A Bayesian diagnostic algorithm for student modeling and its evaluation, *User Modeling and User-Adapted Interaction* **12**(2-3), 281–330, 2002
12. R. J. Mislevy, R. G. Almond, D. Yan, and L. S. Steinberg, Bayes nets in educational assessment: Where

- do the numbers come from?, *Proc. of the 15th Conf. on Uncertainty in Artificial Intelligence*, 437–446, 1999
13. R. E. Neapolitan, *Learning Bayesian Networks*, Prentice Hall, 2003
  14. J. Pearl, *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*, Morgan Kaufmann, 1988
  15. K. K. Tatsuoka, Rule space: An approach for dealing with misconceptions based on item response theory, *J. of Educational Measurement* **20**, 345–354, 1983
  16. W. J. van der Linden and R. K. Hambleton, eds., *Handbook of Modern Item Response Theory*, Springer, 1997
  17. J. Vomlel, Bayesian networks in educational testing, *Int'l J. of Uncertainty, Fuzziness and Knowledge-Based Systems* **12**(Supplement 1), 83–100, 2004
  18. I. H. Witten and E. Frank, *Data Mining: Practical Machine Learning Tools and Techniques*, Elsevier, 2005
  19. 王鈺婷及劉昭麟，從答題表現探討綜合觀念學習歷程的模擬分析，*第九屆人工智慧與應用研討會論文集*，CD-ROM，2004