

南 華 大 學

資訊管理學系碩士班

碩士論文

應用資料採礦技術分析國小課業表現的研究

A study of Applying Data Mining Technology on Analyzing
courses performance of Elementary School Students



研 究 生：林意順

指 導 教 授：邱宏彬

中華民國 99 年 6 月 25 日

南 華 大 學

資訊管理學系資訊管理研究所

碩 士 學 位 論 文

應用資料採礦技術分析國小課業表現的研究

研究生： 林意順

經考試合格特此證明

口試委員：謝品霖
李翔詣
邱宏彬

指導教授：邱宏彬

系主任(所長)：鍾國貴

口試日期：中華民國 99 年 06 月 25 日

南華大學資訊管理學系碩士論文著作財產權同意書

立書人：_____林意順_____之碩士畢業論文

中文題目：應用資料採礦技術分析國小課業表現的研究

英文題目：A study of Applying Data Mining Technology on
Analyzing courses performance of Elementary
School Students

指導教授： 邱宏彬 博士

學生與指導老師就本篇論文內容及資料其著作財產權歸屬如下：

- 共同享有著作權
- 共同享有著作權，學生願「拋棄」著作財產權
- 學生獨自享有著作財產權

學生：_____林意順_____ (請親自簽名)

指導老師：_____邱宏彬_____ (請親自簽名)

中 華 民 國 99 年 5 月 20 日

南華大學碩士班研究生
論文指導教授推薦函

資訊管理系碩士班林意順君所提之論文
應用資料採礦技術分析國小課業表現的研究
係由本人指導撰述，同意提付審查。

指導教授


99年5月20日

誌 謝

本文能夠順利完成，首先感謝我的指導教授邱宏彬老師。有他的鼓勵與指導，我才能克服許多論文寫作上的困難和盲點，得以初窺資料採礦技術應用之奧妙，老師的教誨和體諒包容，學生永遠感激在心。同時，也謝謝口試委員李翔詣教授及謝昆霖教授對本文提供寶貴的經驗與建議，讓論文能更趨於完備，特此致謝。

感謝親愛的老婆怡真，這段時間大多有妳幫忙照顧孩子，我才能有較多的時間專心完成論文；也感謝我親愛的家人，體諒我這兩年無法常回家參與家庭活動的心情。

感謝專班的所有同學們，沒有你們這段時間的鼓勵與扶持，就沒有今天得以成長的我。誠如鐘國貴所長言：「唸研究所獲得的最寶貴資產就是人脈，就是認識資管所大家庭裡的這一群良師益友。」

最後，回首兩年的求學歷程，辛苦卻也很充實。在人生旅途中，此段經驗將是我記憶裡最耀眼的那顆晨星。

意順

謹誌於南華大學資訊管理學系碩士班

2010/06

應用資料採礦技術分析國小課業表現的研究

學生：林意順

指導教授：邱宏彬 博士

南華大學資訊管理學系碩士班

摘 要

資料採礦 (Data Mining) 技術自 80 年代問世至今，其應用已經無所不在。資料採礦利用統計原理與演算法等相關軟體技術，可從大量資料中挖掘出隱藏在背後的因果關係和有意義的規則，幫助企業決策者做出有力決策並創造最大利益。基於此，本研究試圖從國小中年級學生的原始學籍成績資料，經過預處理 (Pre-Processing)、轉化成可用的數據，再利用關聯法則 (Association Rule) 挖掘法找出隱藏在成績背後的關聯性。最後，將研究結果與建議，提供學校做為日後補救教學與課程規劃的參考。

關鍵詞：課業表現、資料採礦、關聯法則

A study of Applying Data Mining Technology on Analyzing courses
performance of Elementary School Students

Student : Lin, Yi-Shun

Advisors : Dr. Chiu, Hung-Pin

Department of Information Management
The Graduated Program
Nan-Hua University

ABSTRACT

Data Mining technology since the 80's come out so far, its application has been everywhere. Use of statistical theory and algorithms, Data Mining can identify causal relationships and meaningful rules hidden in large amounts of data behind, to help decision makers to make effective decisions and create the best interest. Based on this, this study attempted to use elementary students in the original score, after Pre-Processing, into usable data, and then use the Association Rule mining method to find hidden behind the relevance of results. Finally, the study findings and recommendations to the school as remedial teaching and curriculum planning future reference.

Keyword : courses performance, Data Mining, Association Rule

目錄

書名頁	i
論文口試合格證明	ii
著作財產權同意書	iii
論文指導教授推薦書	iv
誌謝	v
中文摘要	vi
英文摘要	vii
目錄	viii
表目錄	x
圖目錄	xi
第一章 緒論	1
第一節 研究背景與動機	1
第二節 研究目的	2
第三節 研究範圍與限制	3
第四節 論文架構	3
第二章 文獻探討	5
第一節 課業表現的定義及研究	5
第二節 資料採礦的定義、類型及應用	8
第三節 關聯法則	10
第四節 相關研究之回顧與探討	12

第三章 研究方法	14
第一節 研究對象及資料來源	14
第二節 研究流程	14
第三節 資料處理	16
第四節 資料採礦工具 SQL SERVER 2005 簡介	20
第四章 資料分析	21
第一節 資料特性分析	21
第二節 關聯法則模型的建置過程	26
第三節 規則尋找的結果及說明	34
第五章 結論與建議	46
第一節 研究結論	46
第二節 研究建議	47
參考文獻	49
一、中文部份	49
二、西文部份	51

表目錄

表 1 整合過的 Excel 資料表	17
表 2 各科目成績的分級標準	18
表 3 六級距分類表	18
表 4 資料表 S2	19
表 5 訓練組資料表 S1	19
表 6 三級距分類表	22
表 7 國語科成績人數分佈表	22
表 8 英語科成績人數分佈表	23
表 9 數學科成績人數分佈表	23
表 10 社會科成績人數分佈表	24
表 11 自然與科技科成績人數分佈表	24
表 12 各科成績男女人數分佈之增益值表	25
表 13 各科成績間的關聯規則總表	37
表 14 各科成績間的關聯規則總表—整理一	41
表 15 各科成績間的關聯規則總表—整理二	43

圖目錄

圖 1 條件機率示意圖	11
圖 2 研究流程	15
圖 3 SQL Server 2005 資料平台的配置	20
圖 4 學生性別比例分佈圖	21
圖 5 新增資料庫	26
圖 6 匯入 Excel 資料表	27
圖 7 設定適當的資料類型	27
圖 8 新增採礦專案	28
圖 9 連接 Score 資料庫	28
圖 10 連接 dbo.S1\$ 資料表	29
圖 11 選取關聯規則採礦技術	30
圖 12 勾選「輸入」和「可預測」的選項	30
圖 13 採礦工具畫面	31
圖 14 採礦模型標籤頁	31
圖 15 採礦模型檢視器－「項目集」標籤頁	32
圖 16 採礦模型檢視器－「規則」標籤頁	33
圖 17 採礦模型檢視器－「相依性網路」標籤頁	33
圖 18 「國語→數學」採礦模型設定	35
圖 19 「國語→數學」項目集	35
圖 20 「國語→數學」有效規則	36
圖 21 「國語→數學」相依性網路	36

第一章 緒論

學生的課業表現 (courses performance) ，或稱學習成就 (learning achievement) 、學習表現 (learning performance) 、學業成績 (academic achievement)……等等，均是代表同樣含義。一直以來都是教師、學校、家長、學生自己，乃至國家教育相關機構與學者們都相當關心的議題。一個國家的競爭力強弱與否，在物質層面，除了可從經濟發展的程度得窺一二外；在精神層面來說，國民的教育程度水平無疑是另一個更重要的參考指標。有鑑於此，世界各國無不投注大量金錢與心力積極地在研究學生的課業表現上，希望能把研究的成果作為教育改革政策與課程修訂的重要依據。

過去，這些成績資料只能存放在紙本上提供學校與教師當做歷史記錄來參考，無法從中得知任何有用的訊息。如今，資料採礦技術的蓬勃發展，讓我們得以從存放大量成績資料的資料庫中找出隱藏的、有用的知識，有機會為學校教育和學生的學習做出更有益的貢獻。

第一節 研究背景與動機

本人從事國小教育工作今年邁入第十一年，其間輾轉服務過幾所不同類型的學校，深切感受到不論學校規模、家長社經地位或教育資源分配的條件有多麼懸殊不均，學生的成績表現這項頭號參考指標，在大家的心目中可說是沒有過一絲一毫的動搖。從過去傳統的聯考制度、死背強記的填鴨式教育、文憑價值取向；演變到今日強調的多元入學、適性教

育、帶著走的能力觀……，雖然各種社會價值觀不斷在改變，但課業成績的表現在學習成果與努力的評斷上依舊保有一定程度的客觀性。

根據內政部統計處的統計資料結果顯示，去年民國98年出生嬰兒計十九萬一千三百一十人、粗出生率8.3%，分別較前年民國97年減少3.7%及減少0.4%，再創新低〔4〕。少子化的結果，以致內人服務的學校學生流失的情形有逐年增加的趨勢，在面臨教師減班超額以及新生招生不足的雙重壓力下，該校積極走訪學區急欲了解新生入學意願與舊生轉學背後的原因。分析後發現，家長決定子女就讀的關鍵在於學校是否能夠讓學生未來具有較佳的競爭力，至於如何看出競爭力一點，多數家長直言，在他們心目中，子女的成績表現就是判斷未來升學能否有較佳競爭力的一個重要依據。

以上原因，讓我有興趣應用研究所課程中學到的資料採礦（Data Mining）技術來分析學生成績表現的念頭，與教授幾次討論覺得可行之後，便確立此研究就是日後論文的題目。

第二節 研究目的

本研究採用資料採礦工具裡的關聯法則（Association Rule）演算法來找出各科成績間的學習關聯，一來因為研究資料的筆數不多；二來資料內容很少增減異動，所以用關聯法則裡最常用的Apriori演算法來做資料分析是最簡潔有效的方法。

以下是本研究預期所要達到的目的：

壹、藉由資料採礦技術分析學科成績，找出低成就學生在各學科之間

成績表現的關聯性，藉此做出正確方向的補救教學。

貳、找出潛藏在學科成績間的有意義規則，用來做為調整課程內容、提高學生學習成效之依據，藉此提昇辦學績效、減少學生流失。

第三節 研究範圍與限制

研究範圍以嘉義市某國小中年級段的學童為對象，該校位於嘉義市西區，屬中型學校。學生來源多為嘉義市，其餘少數為鄰近縣市（嘉義縣、台南縣市、雲林縣……等）越區就讀的學子。樣本數分散平均，對本研究採礦資料庫的建立有相當的參考價值。研究限制方面，因為當初資料取得僅設定在單一國小中年級段學童部份，因此研究結果不見得能適用於全體年段（一到六年級）甚至概化到他校。

第四節 論文架構

本研究共分成五章，各章節內容之安排概略說明如下：

第一章 緒論

分成四節，逐步說明本研究之研究背景與動機、研究目的、研究範圍與限制、以及本論文架構。

第二章 文獻探討

分成四節，前三節就課業表現定義及相關研究、資料採礦定義及類型應用、關聯法則做相關文獻的探討。末節則摘錄兩篇跟本研究主題相近的論文做回顧，重點說明其研究結果與發現。

第三章 研究方法

分四節，針對研究對象與資料來源、研究流程、資料處理以及資料採

礦工具SQL SERVER 2005逐一做概略介紹。

第四章 資料分析

分三節，針對資料做敘述性統計，整理出一些人（次）數分佈表，之後進行資料採礦之關聯法則分析，經不斷反覆篩選與觀察，得出一些有意義的規則。

第五章 結論與建議

分二節，針對關聯法則分析所得到的規則進行整理，推敲、解釋背後可能的原因，最後提出結論與建議。

第二章 文獻探討

第一節 課業表現的定義及相關研究

壹、課業表現的定義

儘管在詞彙表達上有些許不同，但是課業表現、學習表現、學業成績或學習成就……等字眼所要表達的概念其實都是一致的，即是學生在學校學科上的學習結果；或是透過學習歷程所獲得較持久性的結果〔19〕。黃富順〔13〕認為，學業成就是指學生在學校裡，經由一定的課程、教材，透過學習後所獲得的知識和技能，通常是以學校考試成績或由學業測驗上所獲得的分數代表之。Brown（1981）認為，所謂學業成就是指經由正式的課程、教學設計之特殊教育經驗，所獲得的知識、理解和技能，亦即個體經由特殊教學而獲得的某些訊息和純熟技能〔8〕。依據上述學者的表述，顯而易見，於是學習成就就有了所謂狹義與廣義之區別，廣義的學習成就泛指各種在校期間的學生學習紀錄資料，如作業、平時測驗、期中考試、期末考試等；狹義的學習成就則指各學科的學習成績，或各學科綜合後的平均學期成績〔6〕。最後，就教育實質意義的角度來看，學習成就其實就是學生透過學習，用以增進其生活適應與身心發展之表現。

至於本研究所謂的課業表現一詞，採取的是狹義的學習成就觀點。研究資料來源為嘉義市某國小三、四年級學生，在95、96兩學年度的國語、英語、數學、社會、自然與科技……等五學科的各科平均分數。

貳、影響課業表現的因素與相關研究

整理近年來國內、外學者對課業表現的看法與論述後，影響學生學業成績的因素大致可分為兩大部份。

一、個人因素方面

包含個體的生理與心理因素，如：學習動機、性別、情緒、態度、智力、興趣、價值、經驗、期望等等。以下僅針對一項對本研究重要的因素加以說明。

（一）性別因素

多數研究均顯示出性別與課業成績之間確有相關性存在。Stickney與Fitzpatrick〔29〕研究指出，柯爾曼對美國學校做大規模的研究，發現性別是影響學業成績的因素之一，且女生優於男生。在國內張春興依據國、內外的文獻指出，小學階段時女生的語文能力比男生好，男生的數理推理和空間能力比女生好，但整體成績表現女生優於男生，在國小低、中年級，此種狀況更為明顯，而且台灣與美國都是如此〔12〕。陳美娥〔9〕的研究發現，不同性別的國小兒童，其國語科成績有顯著之差異，且女生優於男生。綜合上述研究可以發現，性別的確是影響課業成績的重要因素之一。

二、環境因素方面

環境因素包含了家庭社經地位、父母管教態度與重要他人（如：家長、教師）的期望……等等因素。以下也僅針對一項較重要的因素加以說明。

（一）家庭社經地位（包括父母教育程度、父職業以及家庭收入）

根據許多實證研究發現，家庭社經地位是影響文化資本的重要背景因素，進而影響個人的教育成就。Chen, Lee與Stevenson〔24〕以縱貫研究法長期觀察美國、台灣與日本之國小一年級學童共729人，發現兒童早期的認知能力和家庭社經地位的關係非常一致。Hao和Burns〔26〕分析亞洲移民與美國學生之學業成績差異的研究指出，社會資本是造成各國學生學業成績差異的重要因素，他們發現中國及韓國移民學生在學習過程中，由於與父母互動的品質與頻率較高，增強了父母的教育期望，進而提高子女的學業成績。國內研究方面，謝孟穎〔18〕採問卷調查和訪談法探究家長社經背景與學生學業成績之關聯，研究亦證實家長社經地位和學生學業成績之間確實有相關性，並發現家長的教育期望、家庭文化刺激等都是影響學生學業成績的重要中介變項。也有學者提出家庭讀書環境對於升學與成績表現有重要影響，如：父母買課外讀物的頻率、擁有個人書桌或書櫃、書架的項數，以及補習……等等；家庭社經地位越高者，家庭讀書環境越佳，補習的機會也越多，而有助於提高成績與升學機會〔11〕。譚康榮〔20〕在「誰家小孩學習成就最高？哪群學生心理最不健康？台灣教育長期追蹤資料庫的初步發現」一文中也指出，對台灣的中學生而言，父母教育程度對學生學習成就的影響，比貧富差距對學習成就的影響更為重要；這個現象與美國的發現極為相似且更為明顯。

總結來說，在學業成就方面，不同的性別、家庭社經地位、家庭氣氛、

父母管教方式與學童的學業成就有顯著之差異（〔10〕；〔15〕；〔16〕）。而在社會大眾對家庭社經地位低的學生普遍期待性較低，加上這些學生在比較缺乏可以仿效的模範和重要他人的期望的雙重影響之下，在課業表現上也會顯得相對消極。

第二節 資料採礦的定義、類型及應用

壹、資料採礦的定義

“Data Mining”一詞在華人世界的中文刊物、文獻中有著各式各樣的中譯名稱，如：資料探勘、資料挖掘、資料開採、數據挖掘、資料採擷等，本論文採用的是中華資料採礦協會（Chung-Hua Data Mining Society, CDMS）的說法，把Data Mining譯名為「資料採礦」，依中華資料採礦協會〔3〕指出：Data Mining最早是由人稱資料採礦之父的Usama Fayyad在1991年首次於他的博士論文中所提出，目的是為了要從龐大的維修資料中找出規則。此外，和資料採礦一樣常聽到的字眼還有KDD（Knowledge Discovery in Database），不過一般來說，還是資料採礦一詞較為大眾所常用。

談過以上這些，究竟資料採礦的定義為何？以下擷錄幾位資料採礦大師對資料採礦所下的定義，供大家參考：

- 一、從資料庫中挖掘出不明確、前所未知以及潛在有用的資訊的過程〔27〕。
- 二、資料探勘乃使用自動或半自動的方式，對龐大的資料作分析，為了從資料庫中找出有意義的關係或規則〔28〕。
- 三、資料探勘是一種新的且不斷循環的決策支援分析過程，它能夠從

組合在一起的資料中，發現出隱藏價值的知識，以提供給企業專業人員參考〔23〕。

四、資料探勘是指由已存在的資料中探勘出新的事實及發現專家尚且不知的新關係〔25〕。

貳、資料採礦技術的類型及應用

綜合前述幾位大師的定義，我們不難看出，資料採礦就是一段發現有意義規則的過程。資料採礦既然是用來找尋有意義的規則，因此，可以將規則的型態區分為幾種模式〔5〕：

- 一、分類 (Classification)：基本上資料採礦的原形就是預測，預測類別變數的過程稱之為「分類」。
- 二、推估 (Estimation)：推估是預測連續數值，透過已知的屬性，推估未知的連續數值的走向與趨勢。
- 三、群集化 (Cluster)：根據相似性，將相似的事物分群，使得複雜的資訊大幅簡化，這個步驟稱之為「群集化」。
- 四、同質分組 (Affinity Group)：從歷史資料中，找出哪些事件總是相伴發生，我們稱之為關聯法則 (Association Rule) 或者是購物籃分析 (Basket Analysis)。
- 五、序列 (Sequential)：透過序列，可以找出事物「先後」發生的順序，我們稱這樣的規則為時序規則 (Sequential Pattern)。
- 六、描述 (Description)：資料採礦的過程中，除了分析的預測模型之外，很重要是在分析與處理資料的過程中，透過資料視覺化以

及觀察來找出許多有意義的規則。

在現今社會中，資料採礦的應用範圍非常廣泛。根據Two Crows Corp.最近的調查顯示，Data Mining主要的三個應用方式都在市場推廣方面，分別是：Customer Profiling、Targeted Marketing、以及 Market-Basket Analysis〔2〕。在Customer Profiling方面，希望找出客戶的一些共同的特徵，藉此預測哪些人可能成為我們的客戶，以幫助行銷人員找到正確的行銷對象；Market-Basket Analysis用來幫助零售業者瞭解客戶的消費行為，譬如哪些產品客戶會一起購買，或是客戶在買了某一樣產品之後，在多久之內會買另一樣產品等等。客戶關係的管理則是Data Mining的另一個常見的應用方式，因為畢竟找一個新客戶的成本要比留住一個原有客戶的成本要高出許多。近來，電話公司、信用卡公司、保險公司、股票交易商、以及政府單位對於詐欺行為的偵測都很有興趣，Data Mining可以找出可能的詐欺交易，減少損失；財務金融業可以利用Data Mining來分析市場動向，並預測公司的營運以及股價走向。

第三節 關聯法則

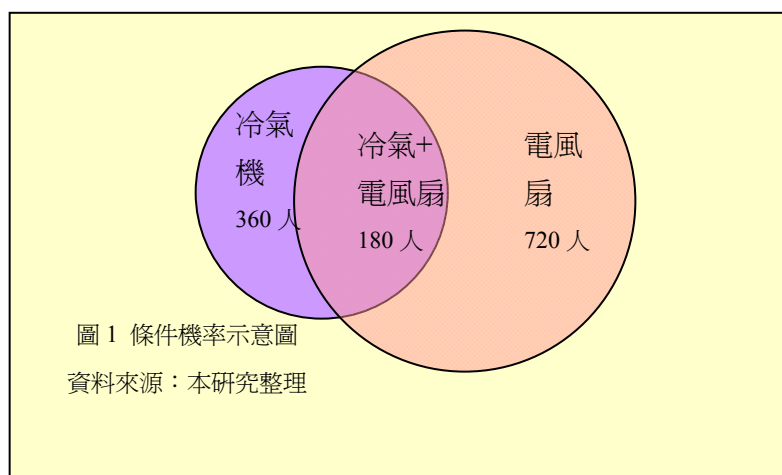
關聯法則原理

關聯法則大概是目前資料採礦中最淺顯易懂、最普遍使用的一種基本技術，該技術由Agrawal〔22〕等幾位學者所提出，目的就是希望能從一群龐大交易的資料項目及屬性間，找出彼此之間相關的規則，以了解顧客的消費行為，進而做為交叉銷售或套餐組合等商業銷售行為的應用。

關聯法則的基本規則如下：

購買甲商品也會同時購買乙商品的有百分之P

我們試用「圖1 條件機率示意圖」來說明這條規則是如何計算出來的。假設這個月份在大賣場裡，購買冷氣機的人有360人，購買電風扇的人有720人，而兩者都有買的共有180人。



依據上述條件，我們可以利用條件機率的公式算出底下兩條關聯規則：

購買冷氣也同時購買電風扇的有50%($180/360=0.5$)

購買電風扇也同時購買冷氣的有25%($180/720=0.25$)

這個規則我們稱為可信度(C Confidence)，意思就是買甲商品也會同時買乙商品的機率有多少？機率越高，表示可信賴的程度越高。而關聯法則除了這個評估指標以外，尚有另一個評估指標—支持度(Support)，意思為甲商品及乙商品都有的交易筆數占總交易筆數的百分比。以本例子來說，倘若在這個月的所有1000筆銷售資料裡，購買冷氣也同時購買電風扇的有300筆，那支持度有3%；而購買電風扇也同時購買冷氣的有100筆的話，則支持度只有1%。

這兩個指標通常被用來做為評估規則是否成立的標準。支持度和可信度的門檻設定過高會不容易產生規則，遺漏可能的重要關聯規則；但是

門檻設定值過低又會產生許多雜亂不可靠的規則。所以，支持度和可信度的設定值還需要靠分析者的經驗〔21〕。

第四節 相關研究之回顧與探討

本節簡單整理兩篇有關運用資料採礦技術研究學生學習表現文章的結果，以供本研究第五章比較印證使用。

壹、鄭景邁〔17〕以資料探勘技術分析國小五、六年級學童學業成績，焦點放在學習狀況較差的低成就學童身上，利用關聯規則演算法找出學童各學科成績間的交互關係，最後提供探勘結果給學校做為補救教學策略的參考。資料來源為國語、英語、數學、自然、社會等五個主要學科，將取得的原始資料合併、剪裁、轉換成適合探勘的資料庫後，研究發現：

一、國語科的學習狀況對數學、自然、社會三科有明顯的影響，所以可找出包含在這三條關聯規則中的學童（國語 \cap 數學、國語 \cap 自然、國語 \cap 社會），對國語科投入較多比例的補救教學資源。

二、信賴度最高的一條關聯規則是「國語、自然 \rightarrow 社會」，配合觀察其他的關聯規則發現自然科出現的機會相當高，自然科的學習成就明顯影響其他科目，所以也應考慮增加自然科的補助教學資源。

貳、林金火〔7〕應用資料探勘技術分析高中一、二年級學生成績，也是要從成績不佳的學生身上，利用關聯規則演算法找出學生各

學科成績間的交互關係。資料來源為高中課程的國文、數學、英文、物理、化學、歷史、地理等七科，將原始資料經合併、剪裁及轉換後，研究發現：

- 一、國文、歷史及歷史、地理的是互為有效的關聯規則，這說明學生的史地文學的學習成效是有關聯的。
- 二、在第三、四條關聯規則中物理、化學是相互關聯的，這更說明高中生的學習上已有分科學習關聯的事實。
- 三、在 L_2 的候選項目集分析中，數學、英文兩科被刪除，在留下的國文、物理、化學、歷史、地理分析中，只得到六條有效的關聯規則，可見高中生的各科學習的相互影響是較中小學為小的。

第三章 研究方法

本章研究方法分為四小節介紹：第一節先說明研究對象及資料來源；第二節介紹本研究的研究流程；第三節介紹資料處理的步驟；最後，第四節簡介本研究資料採礦工具軟體SQL SERVER 2005。

第一節 研究對象及資料來源

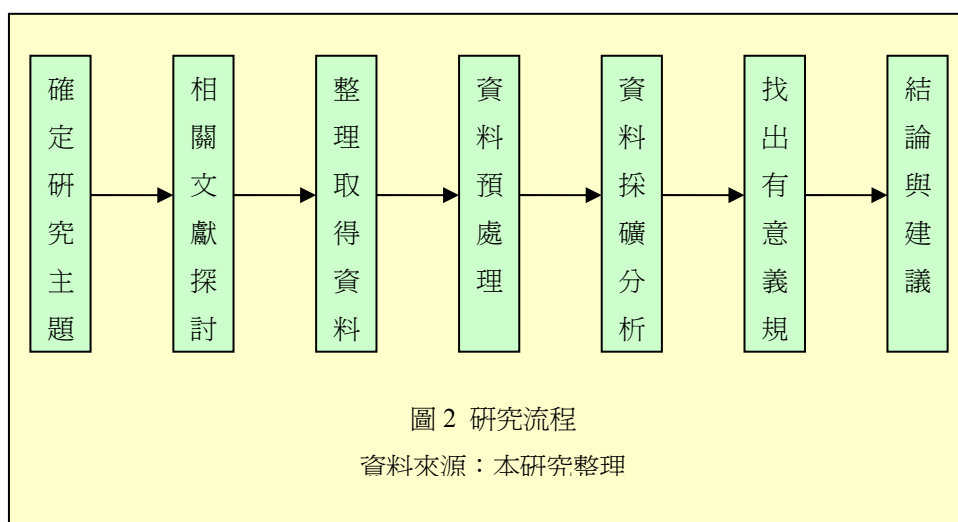
本研究是以嘉義市某國小95、96兩學年度三、四年級學童為研究對象，共計441人。該校位於嘉義市西區，屬中型學校。學生來源多為嘉義市，其餘則為鄰近縣市（嘉義縣、台南縣市、雲林縣……等）越區就讀的學子。研究資料來源為該校三、四年級學童在95、96兩學年度之國語、英語、數學、社會、自然與科技等五科的成績。

第二節 研究流程

本研究以嘉義市某國小中年級學童之國語、英語、數學、社會、自然與科技等五科成績，做為採礦資料庫，然後採用關聯法則演算法為採礦工具，分析學科成績，找出潛藏在學科成績間的有意義規則。研究流程（詳見「圖 2 研究流程」）依序為：確定研究主題、相關文獻探討、整理取得資料、資料預處理、資料採礦分析、找出有意義的規則、結論與建議，其內容說明如下：

壹、確定研究主題：確定主題與整個研究進行的方向後，開始著手資料的蒐集。

- 貳、相關文獻探討：閱讀尋找跟研究主題相關的文獻，著手整理背景知識的同時，並從中找尋適合本研究的研究工具。
- 參、整理取得資料：資料陸續取得的同時，和教授討論先前蒐集資料的完整性並做適當的補強。
- 肆、資料預處理：為求正確地使用資料採礦技術，避免不適當的資料對分析結果造成不利影響（garbage in,garbage out），此步驟目的在於將原始資料經由刪除、整合及轉換純化資料，形成可用資料表。同時，了解輸入變項的資料表欄位特性，以便進行資料分析之變數設定。
- 伍、資料採礦分析：運用關聯法則演算法進行資料分析，透過反覆不斷地建立、觀察評估模型，試圖取得有意義的樣式與規則。
- 陸、找出有意義的規則：取得有意義的樣式與規則後，嘗試去解釋、推敲隱藏在規則背後的真正原因。
- 柒、結論與建議：找出成績背後潛在的真正原因，做為調整課程內容、提高學生學習成效之參考。



第三節 資料處理

原始資料來源為嘉義市某國小三、四年級學生95、96學年度學籍簿上的成績資料。資料取得上，在實際拜訪並說明其研究用途後，除其中一班導師不願提供外，最後共取得15班計441筆資料。觀察後發現，欄位數目計有：學年度、學期、年級、班級、座號、姓名、國語、英語、鄉土語言、語文平均、數學、自然與生活科技、社會、藝術與人文、健康與體育、綜合活動、學習領域總成績、日常生活表現、學期總成績、選修課程(1)、選修課程(2)、選修課程(3)、選修課程(4)、應出席日數、事假日數、病假日數、曠課日數、缺席總日數、導師評語……等十餘個欄位。

接著，開始進行資料預處理（Pre-Processing）的過程。此部份分成三階段，分述如下：

壹、資料純化：考量到不是所有欄位均對採礦過程有用，因而此階段只留下國語、英語、數學、社會、自然與科技等五科的欄位值，其餘欄位均刪除，並且把學年度、年級、班級、座號四項欄位值合併命名成「學號」一欄，以減低資料複雜度。

貳、資料整合：此外，跟教授討論過後，教授建議加入性別一欄以探討男女學習表現上是否有差異？因此，再把學生輔導記錄簿上的性別欄位值補上，與學號、國語、英語、數學、社會、自然與科技等六個欄位值先存放在同一張Excel工作表中。接著，一般在設計商業交易的資料表時，多設計成主檔和明細檔這種一對多的關聯表（Relations）格式，而SQL SERVER 2005在做資料採礦的過程中，對於資料表的輸入也是有「案例」和「巢狀表」這樣的類似設計。

考量本研究資料的來源並不複雜，因此採用與SQL SERVER 2005能相容的Excel表格方式將資料存放在單一工作表中（詳見「表1 整合過的Excel資料表」）。

表1 整合過的Excel資料表

學號	性別	國語	英語	數學	社會	自然與科技
954101	男	95	95	92	96	94
954102	男	97	95	91	97	95
954103	男	83	92	76	83	83
954104	男	96	95	91	95	96
954105	男	91	94	83	87	90
®954106	男	89	76	81	92	89
954107	男	93	91	95	96	95
954108	男	86	80	68	83	86
954109	男	62	59	38	70	70
954110	男	80	87	70	78	76
僅以 10 筆代表，以下省略						

資料來源：本研究整理

參、轉化成資料庫可用的訓練組資料表S1：參照國際分級指標的計分方法，將成績分為頂標、前標、均標、後標及底標〔14〕，算出各科目的五標分級標準後，製成表格（詳見「表2 各科目成績的分級標準」）。然後把成績大於等於頂標的訂為A級分、小於頂標且大於等於前標的訂為B級分、小於前標且大於等於均標的訂為C級分、小於均標且大於等於後標的訂為D級分、小於後標且大於等於底標的訂為E級分、小於底標的訂為F級分，再製成「表3 六級距分類表」。最後，依據「表3 六級距分類表」之規則把「表1 整合過的Excel資料表」轉換成資料表S2（詳見「表4 資料表

S2」)。

表2 各科目成績的分級標準

級距	國語	英語	數學	社會	自然與科技	說明
頂標	97	97	96	98	97	由高到低，分數排名第12%的
前標	96	96	95	97	96	由高到低，分數排名第25%的
均標	90	87	86	90	89	由高到低，分數排名第50%的
後標	81	73	72	80	78	由高到低，分數排名第75%的
底標	75	66	63	73	72	由高到低，分數排名第88%的

資料來源：本研究整理

表3 六級距分類表

類別 等級	A	B	C	D	E	F
≥頂標	√					
<頂標 ≥前標		√				
<前標 ≥均標			√			
<均標 ≥後標				√		
<後標 ≥底標					√	
<底標						√

資料來源：本研究整理

表4 資料表S2

學號	性別	國語	英語	數學	社會	自然與科技
954101	男	C	C	C	C	C
954102	男	A	C	C	B	C
954103	男	D	C	D	D	D
954104	男	B	C	C	C	B
954105	男	C	C	D	D	C
954106	男	D	D	D	C	C
954107	男	C	C	B	C	C
954108	男	D	D	E	D	D
954109	男	F	F	F	F	F
954110	男	D	C	E	E	E
僅以 10 筆代表，以下省略						

資料來源：本研究整理

再按照常態分佈的概念，將成績分成三群，高分的一群、中分的一群，低分落後的一群。把分數由高到低，分數排名大於等於25%以上的為高分組（即A級分+B級分的總和）；排名小於25%大於等於75%以上為中分組（即C級分+D級分的總和），小於75%以下者為低分組（即E級分+F級分的總和），來製成「表5 訓練組資料表S1」。

表5 訓練組資料表S1

學號	性別	國語	英語	數學	社會	自然與科技
954101	男	中分組	中分組	中分組	中分組	中分組
954102	男	高分組	中分組	中分組	高分組	中分組
954103	男	中分組	中分組	中分組	中分組	中分組
954104	男	高分組	中分組	中分組	中分組	高分組
954105	男	中分組	中分組	中分組	中分組	中分組
954106	男	中分組	中分組	中分組	中分組	中分組
954107	男	中分組	中分組	高分組	中分組	中分組
954108	男	中分組	中分組	低分組	中分組	中分組

954109	男	低分組	低分組	低分組	低分組	低分組
954110	男	中分組	中分組	低分組	低分組	低分組
僅以 10 筆代表，以下省略						

資料來源：本研究整理

第四節 資料採礦工具SQL SERVER 2005簡介

本研究採用微軟公司的SQL SERVER 2005軟體為資料採礦工具，其演算法有：決策樹、貝氏機率分類、時序群集、時間序列、群集、線性迴歸、羅吉斯迴歸、類神經網路、關聯規則等九種演算法。

SQL Server 2005同時也是一個功能完備的資料庫平台，利用整合式商業智慧（BI）工具，提供企業級資料管理功能。SQL Server 2005資料庫引擎提供更安全、可靠的儲存環境給關聯式和結構式資料，讓您能夠建置並管理高可用性、高效能資料應用程式。透過豐富的功能集（詳見「圖3 SQL Server 2005資料平台的配置」）、與現有系統的互通性，以及例行工作的自動化，SQL Server 2005為各種規模的企業提供完整的資料解決方案〔1〕。



圖3 SQL Server 2005資料平台的配置

資料來源：Microsoft中文網站〔1〕

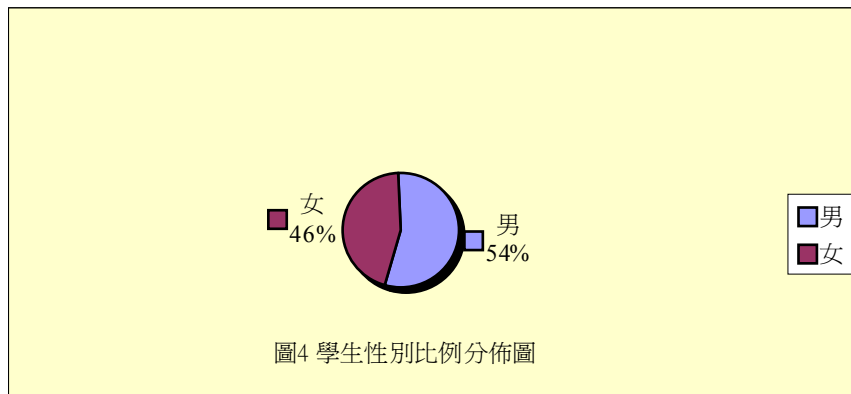
第四章 資料分析

本章共分三節，第一節為資料特性分析，對取得的資料先做一敘述性統計，整理成各種人（次）數分佈表，綜合說明觀察的現象；第二節說明關聯法則模型的建置過程；第三節為實驗結果的數據及說明。

第一節 資料特性分析

壹、性別資料分析：

441筆資料中，男生佔236人，女生佔205人。從「圖4 學生性別比例分佈圖」中可看出比例約54：46，人數差距約8%。



資料來源：本研究整理

貳、以三級距分級標準來分析各科成績：

採百分位數做為訂定高分組、中分組、低分組的標準，所謂高、中、低分組（詳見「表6 三級距分類表」），是把學生成績分數由高到低排列，排名大於等於第25%以上的訂為高分組（即A級分+B級分的總和）；小於25%大於等於75%以上的訂為中分組（即C級分+D級分的總和），小於75%以下者訂為低分組（即E級分+F級分的總和）。

表6 三級距分類表

等 級	類 別	高分組	中分組	低分組
	百分位數 $\geq 25\%$	√		
	百分位數 $< 25\%$ ， $\geq 75\%$		√	
	百分位數 $< 75\%$			√

資料來源：本研究整理

接下來，我們會觀察比較各科高分組、中分組、低分組男女整體人數分佈的情形。不過，因為全體的男女人數並不一致，因此比較的基準不是絕對的比率，而是採相對的增益（Lift）。增益值，就是將子體發生機率除以母體的原始發生機率所得到的比值。增益值 > 1 表示發生機率高於母體。因此，該條件會被視作為傾向發生。

一、國語科成績分析：

從「表7 國語科成績人數分佈表」觀察、對照後，結果發現：高分組部份，女生人數增益值為1.32，其值大於1代表該條件會傾向發生，代表此組得高分的女生相對較多；低分組部份，男生人數增益值為1.18，其值大於1代表該條件會傾向發生，代表此組得低分的男生相對較多。

表7 國語科成績人數分佈表

國語科成績級距	高分組	中分組	低分組		
分組總人數	88	323	30	441	全體總人數
男生人數	34	183	19	236	男生總人數
女生人數	54	140	11	205	女生總人數
男生人數的增益值	0.72	1.06	1.18		
女生人數的增益值	1.32	0.93	0.79		

資料來源：本研究整理

二、英語科成績分析：

從「表8 英語科成績人數分佈表」觀察、對照後，結果發現：高分組部份，女生人數增益值為1.37，其值大於1代表該條件會傾向發生，代表此組得高分的女生相對較多；低分組部份，男生人數增益值為1.20，其值大於1代表該條件會傾向發生，代表此組得低分的男生相對較多。

表8 英語科成績人數分佈表

英語科成績級距	高分組	中分組	低分組		
分組總人數	66	330	45	441	全體總人數
男生人數	24	183	29	236	男生總人數
女生人數	42	147	16	205	女生總人數
男生人數的增益值	0.68	1.04	1.20		
女生人數的增益值	1.37	0.96	0.76		

資料來源：本研究整理

三、數學科成績分析：

從「表9 數學科成績人數分佈表」觀察、對照後，結果發現：高分組部份，女生人數增益值為1.16，其值大於1代表該條件會傾向發生，代表此組得高分的女生相對較多；低分組部份，女生人數增益值為1.05，其值大於1代表該條件會傾向發生，代表此組得低分的女生相對較多。

表9 數學科成績人數分佈表

數學科成績級距	高分組	中分組	低分組		
分組總人數	65	335	41	441	全體總人數
男生人數	30	185	21	236	男生總人數
女生人數	35	150	20	205	女生總人數
男生人數的增益值	0.86	1.03	0.96		
女生人數的增益值	1.16	0.96	1.05		

資料來源：本研究整理

四、社會科成績分析：

從「表10 社會科成績人數分佈表」觀察、對照後，結果發現：高分組部份，女生人數增益值為1.21，其值大於1代表該條件會傾向發生，代表此組得高分的女生相對較多；低分組部份，女生人數增益值為1.05，其值大於1代表該條件會傾向發生，代表此組得低分的女生相對較多。

表10 社會科成績人數分佈表

社會科成績級距	高分組	中分組	低分組		
分組總人數	78	324	39	441	全體總人數
男生人數	34	182	20	236	男生總人數
女生人數	44	142	19	205	女生總人數
男生人數的增益值	0.81	1.05	0.96		
女生人數的增益值	1.21	0.94	1.05		

資料來源：本研究整理

五、自然與科技科成績分析：

從「表11 自然與科技科成績人數分佈表」觀察、對照後，結果發現：高分組部份，女生人數增益值為1.04，其值大於1代表該條件會傾向發生，代表此組得高分的女生相對較多；低分組部份，男生人數增益值為1.06，其值大於1代表該條件會傾向發生，代表此組得低分的男生相對較多。

表11 自然與科技科成績人數分佈表

自然與科技科成績級距	高分組	中分組	低分組		
分組總人數	86	318	37	441	全體總人數
男生人數	48	167	21	236	男生總人數
女生人數	38	151	16	205	女生總人數
男生人數的增益值	1.04	0.98	1.06		
女生人數的增益值	0.95	1.02	0.93		

資料來源：本研究整理

參、說明敘述性分析後觀察到的結果：

整理此小節裡的各科成績男女分佈情形，製成「表12 各科成績男女人數分佈之增益值表」，可以觀察出幾個現象：

表12 各科成績男女人數分佈之增益值表

增益值	高分組	中分組	低分組
男生人數的增益值（國語）	0.72	1.06	1.18
女生人數的增益值（國語）	1.32	0.93	0.79
男生人數的增益值（英語）	0.68	1.04	1.20
女生人數的增益值（英語）	1.37	0.96	0.76
男生人數的增益值（數學）	0.86	1.03	0.96
女生人數的增益值（數學）	1.16	0.96	1.05
男生人數的增益值（社會）	0.81	1.05	0.96
女生人數的增益值（社會）	1.21	0.94	1.05
男生人數的增益值（自然）	1.04	0.98	1.06
女生人數的增益值（自然）	0.95	1.02	0.93

資料來源：本研究整理

- 一、從高分組來看；國語科、英語科、數學科與社會科高分組的女生人數增益值均 >1 ，其值大於1代表該條件會傾向發生，相對來講，代表這四科目高分組整體人數為女生比男生多；而自然與科技科高分組人數則為男生比女生多。
- 二、從中分組來看；國語科、英語科、數學科與社會科中分組的男生人數增益值均 >1 ，其值大於1代表該條件會傾向發生，相對來講，代表這四科目中分組整體人數為男生比女生多；而自然與科技科中分組人數則為女生比男生多。
- 三、從低分組來看；國語科、英語科、與自然科低分組的男生人數增益值均 >1 ，其值大於1代表該條件會傾向發生，相對來講，代表

這三科目低分組整體人數為男生比女生少；而數學科與社會科低分組人數則為女生比男生少。

第二節 關聯法則模型的建置過程

壹、連接資料庫

Step1：首先，點選「SQL Server Management Studio」連結資料庫引擎後，新增一個名為Score的資料庫（詳見「圖5 新增資料庫」）。



圖5 新增資料庫

資料來源：本研究整理

Step2：選定Score資料庫，按滑鼠右鍵選擇「工作」→「匯入資料」，資料來源選擇Microsoft Excel格式，確定檔案路徑之後匯入Excel資料表Database_B.xls（詳見「圖6 匯入Excel資料表」）。

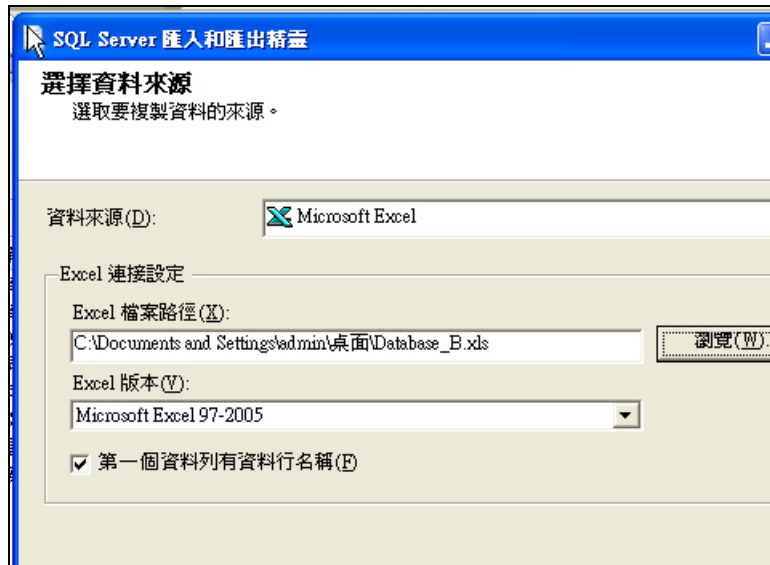


圖6 匯入Excel資料表
資料來源：本研究整理

Step3：選取來源資料表，檢視並設定好資料表S1\$內適當的資料類型（詳見「圖7 設定適當的資料類型」），成功連結後按「關閉」，到此完成整個連結Score資料庫的流程。



圖7 設定適當的資料類型
資料來源：本研究整理

貳、建置關聯法則的採礦結構

Step1：首先，點選「SQL Server Business Intelligence Development Studio」，
建立一個名為Score的採礦專案（詳見「圖8 新增採礦專案」）。

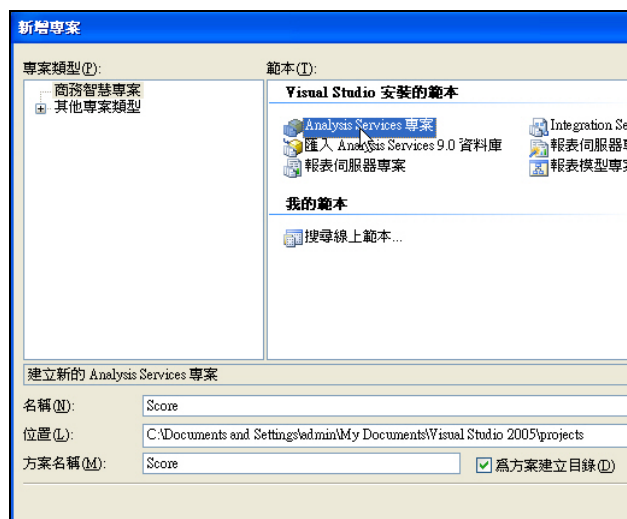


圖8 新增採礦專案
資料來源：本研究整理

Step2：點選「方案總管」內的「資料來源」，按滑鼠右鍵選擇「新增資料來源」，接著連接先前建立的「Score」資料庫（詳見「圖9 連接Score資料庫」），完成資料來源的設定。

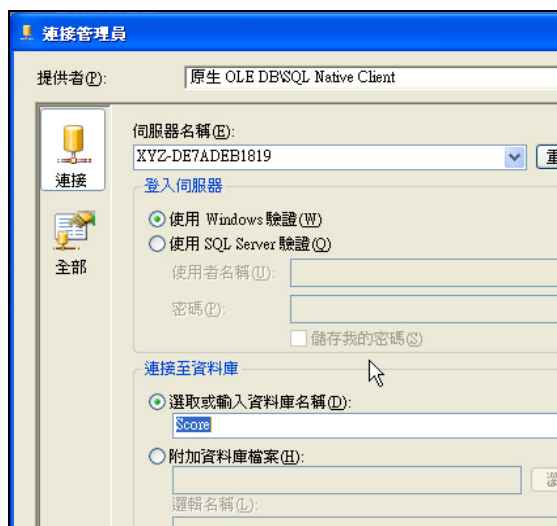


圖9 連接Score資料庫
資料來源：本研究整理

Step3：再點選「方案總管」內的「資料來源檢視」，按滑鼠右鍵選擇「新增資料來源檢視」，選取並連接名為dbo.S1\$資料表的關聯式資料來源（詳見「圖10 連接dbo.S1\$資料表」），完成資料來源檢視的設定。



圖10 連接dbo.S1\$資料表
資料來源：本研究整理

Step4：最後，點選「方案總管」內的「採礦結構」，按滑鼠右鍵選擇「新增採礦結構」，選取「Microsoft關聯規則」資料採礦技術（詳見「圖11 選取關聯規則採礦技術」），然後選取資料來源檢視裡的Score資料表，在「輸入資料表」裡勾選「案例」一項（本研究雖無另外分開資料做巢狀表，但在SQL SERVER 2005裡，單用案例表來跑資料一樣是可行的），指定好培訓資料的輸入和輸出選項（詳見「圖12 勾選『輸入』和『可預測』的選項」），確定資料類型無誤後，即完成新增採礦結構的動作。接下來，就可以開始下一節的規則尋找過程。



圖11 選取關聯規則採礦技術
資料來源：本研究整理



圖12 勾選「輸入」和「可預測」的選項
資料來源：本研究整理

參、採礦工具畫面及標籤頁功能介紹

介紹「圖13 採礦工具畫面」裡幾項關聯法則常用標籤頁的功能：

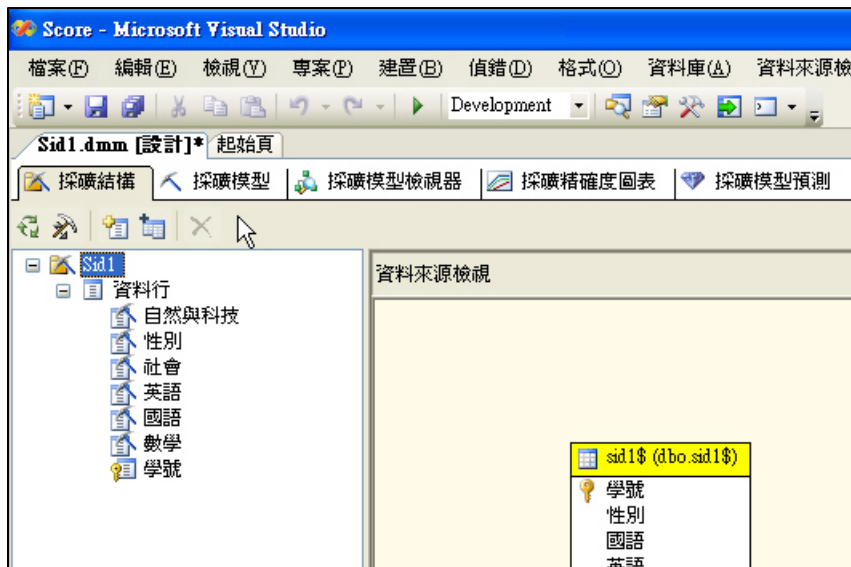


圖13 採礦工具畫面

資料來源：本研究整理

一、採礦模型：此標籤頁面裡（詳見「圖14 採礦模型標籤頁」），可以針對同一資料來源來建立或刪除不同的採礦模型，並處理採礦模型結構。

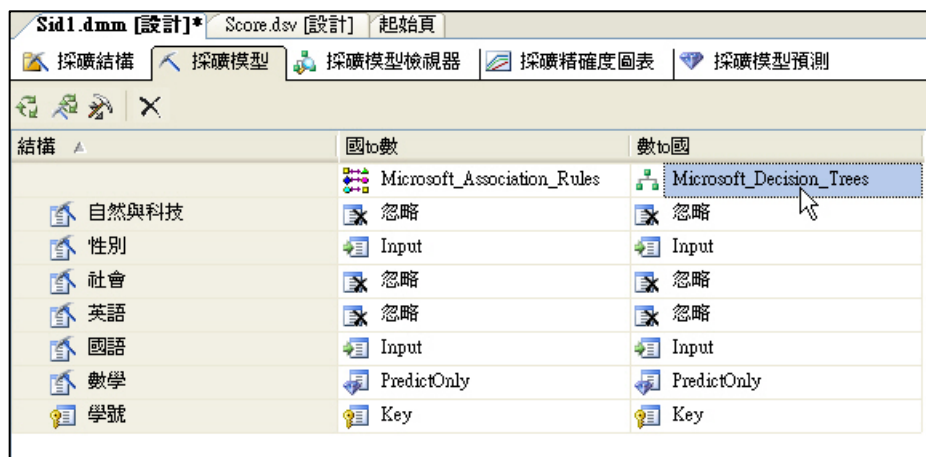
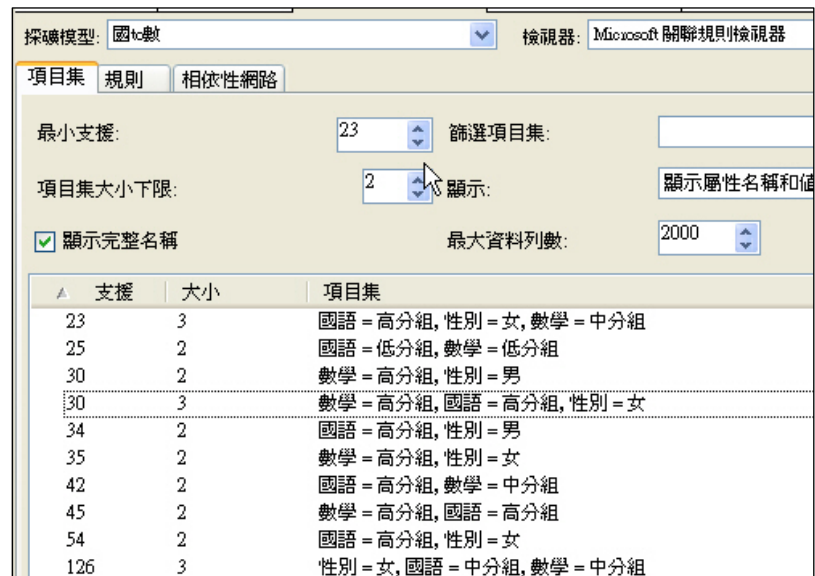


圖14 採礦模型標籤頁

資料來源：本研究整理

二、採礦模型檢視器：此標籤頁面裡有3個子標籤，功能分別為：

(一) 項目集 (詳見「圖15 採礦模型檢視器—『項目集』標籤頁」)：可以設定「最小支援」(案例數)和「項目集大小下限」，也可以打入關鍵字來篩選項目集，來減少項目集的數目以便觀察。



支援	大小	項目集
23	3	國語 = 高分組, 性別 = 女, 數學 = 中分組
25	2	國語 = 低分組, 數學 = 低分組
30	2	數學 = 高分組, 性別 = 男
30	3	數學 = 高分組, 國語 = 高分組, 性別 = 女
34	2	國語 = 高分組, 性別 = 男
35	2	數學 = 高分組, 性別 = 女
42	2	國語 = 高分組, 數學 = 中分組
45	2	數學 = 高分組, 國語 = 高分組
54	2	國語 = 高分組, 性別 = 女
126	3	性別 = 女, 國語 = 中分組, 數學 = 中分組

圖15 採礦模型檢視器—「項目集」標籤頁

資料來源：本研究整理

(二) 規則 (詳見「圖16 採礦模型檢視器—『規則』標籤頁」)：可以設定「最小機率」和「最小重要性」，也可以鍵入關鍵字來篩選，減少規則數目以便找出有意義的規則。

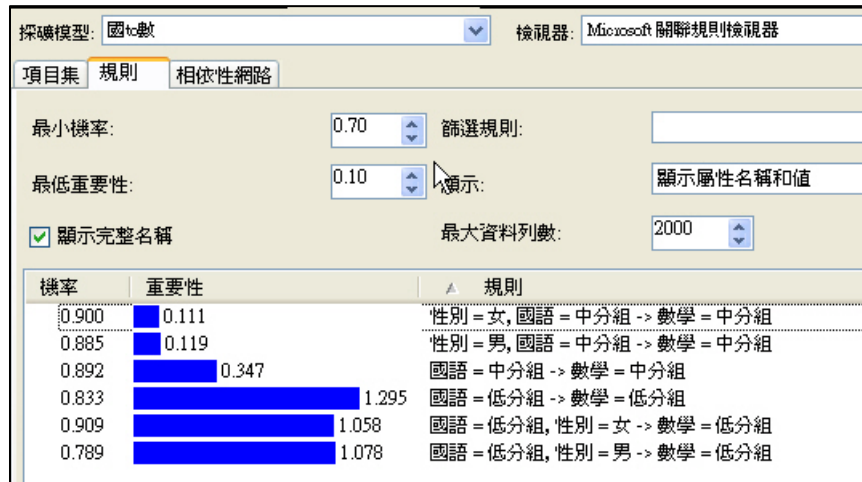


圖16 採礦模型檢視器—「規則」標籤頁

資料來源：本研究整理

(三) 相依性網路 (詳見「圖17 採礦模型檢視器—『相依性網路』標籤頁」)：可以調整左側的捲軸拉桿，來觀察關聯性連結的強度，以便找出有意義的現象。

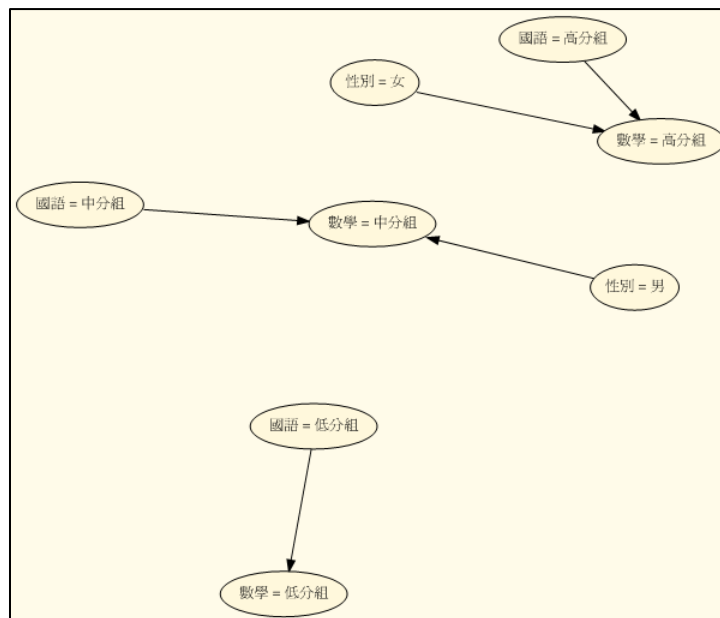


圖17 採礦模型檢視器—「相依性網路」標籤頁

資料來源：本研究整理

第三節 規則尋找的結果及說明

壹、參數的設定與調整

關聯規則能容易解釋其產生的規則，且可完整呈現變數之間的關聯，但篩選規則的條件設定很重要，條件太鬆散可能造成結果多且凌亂；相反地，如果條件太嚴苛可能忽略掉一些罕見變數的有趣樣式。所以，在慢慢調整「最小支援度」、「最小機率」和「最低重要性」的同時，每個關聯規則的數目約控制在10條左右。

本研究參數設定如下：

- 一、最小支援度初始設定在 ≥ 23 ，以資料總筆數441的5%做粗估，如有需要視採礦狀況做調整。
- 二、最小機率初始設定在 ≥ 0.7 （70%），此取法是參考諸多相關文獻的建議來設定的，如有需要視採礦狀況做調整。
- 三、最低重要性初始設定在 ≥ 0.1 ，如有需要視採礦狀況做調整。
- 四、「性別」則設定每次都與其他科目一同做Input選項，觀察是否會與科目相互的學習有所關聯。

貳、開始做各個採礦模型關聯預測：

「國語→數學」關聯預測：國語、性別設為「輸入」，數學設為「預測」。採礦模型設定如下（詳見圖18「國語→數學」採礦模型設定）：

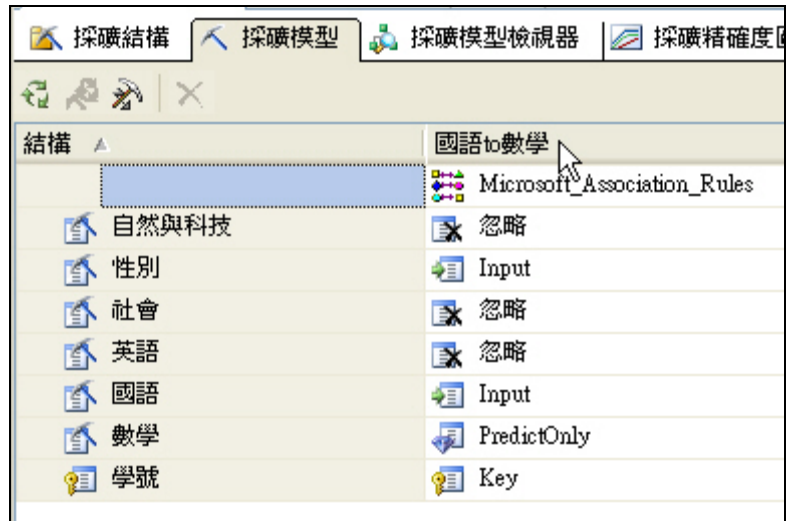


圖18 「國語→數學」採礦模型設定

資料來源：本研究整理

項目集的觀察如下（詳見圖19 「國語→數學」項目集）：

▲	支援	大小	項目集
	23	3	國語 = 高分組, 性別 = 女, 數學 = 中分組
	25	2	國語 = 低分組, 數學 = 低分組
	30	2	數學 = 高分組, 性別 = 男
	30	3	數學 = 高分組, 國語 = 高分組, 性別 = 女
	34	2	國語 = 高分組, 性別 = 男
	35	2	數學 = 高分組, 性別 = 女
	42	2	國語 = 高分組, 數學 = 中分組
	45	2	數學 = 高分組, 國語 = 高分組
	54	2	國語 = 高分組, 性別 = 女
	126	3	性別 = 女, 國語 = 中分組, 數學 = 中分組

圖19 「國語→數學」項目集

資料來源：本研究整理

規則的觀察如下（詳見圖20 「國語→數學」有效規則）：

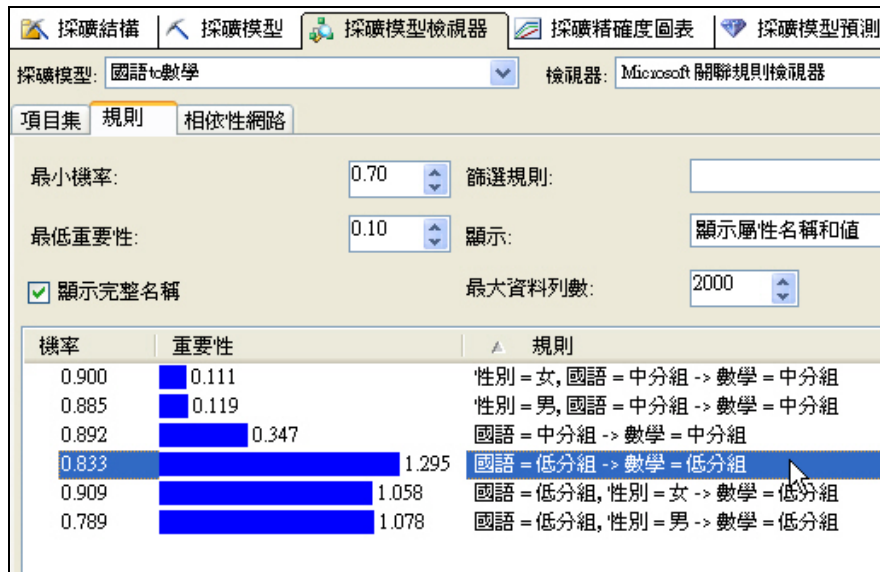


圖20 「國語→數學」有效規則

資料來源：本研究整理

相依性網路的觀察如下（詳見圖21 「國語→數學」相依性網路）：

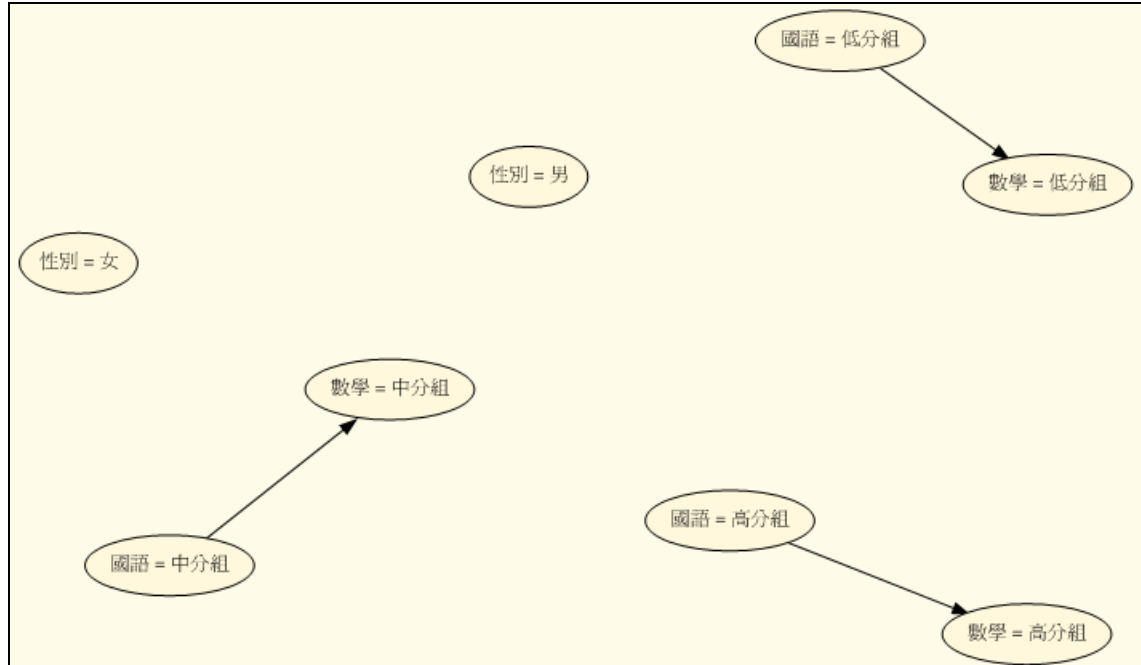


圖21 「國語→數學」相依性網路

資料來源：本研究整理

規則的發現：

- 一、在項目集的觀察中，發現性別在項目集大小下限選2的時候就未達最小支援度，故被刪除；同樣地，在相依性網路的標籤頁面裡，關聯性強度的拉桿調到1/2的位置，性別亦未見與國語、數學有相依性連結的存在，可見「性別」對國語和數學兩學科之間的學習關聯性並無明顯的關聯。
- 二、國語=低分組->數學=低分組的機率是0.833，符合初始最小機率設定值0.7，重要性1.295也是最高的，而支援度25也符合最小支援度23的要求，故取出這條規則。

以下開始，各科的相互關聯規則就仿照「國語→數學」關聯預測的流程做預測，過程就不一一贅述，做成（表13 各科成績間的關聯規則總表）。

表13 各科成績間的關聯規則總表

關聯規則	支援度	機率	重要性	規則	性別關聯
國語→數學	25	0.833	1.295	國語=低分組->數學=低分組	無明顯
數學→國語	25	0.610	1.608	數學=低分組->國語=低分組	無明顯
國語→社會	25	0.833	1.350	國語=低分組->社會=低分組	無明顯
社會→國語	25	0.641	1.630	社會=低分組->國語=低分組	無明顯
國語→自然	24	0.800	1.363	國語=低分組->自然=低分組	無明顯
自然→國語	24	0.649	1.570	自然=低分組->國語=低分組	無明顯
國語→英語	22	0.773	1.092	國語=低分組->英語=低分組	無明顯
英語→國語	22	0.489	1.335	英語=低分組->國語=低分組	無明顯
數學→社會	32	0.780	1.586	數學=低分組->社會=低分組	無明顯
社會→數學	32	0.821	1.512	社會=低分組->數學=低分組	無明顯
數學→自然	32	0.780	1.711	數學=低分組->自然=低分組	無明顯
自然→數學	32	0.865	1.536	自然=低分組->數學=低分組	無明顯
數學→英語	28	0.683	1.178	數學=低分組->英語=低分組	無明顯
英語→數學	28	0.622	1.244	英語=低分組->數學=低分組	無明顯

社會→自然	32	0.821	1.734	社會=低分組->自然=低分組	無明顯
自然→社會	32	0.865	1.633	自然=低分組->社會=低分組	無明顯
社會→英語	26	0.667	1.124	社會=低分組->英語=低分組	無明顯
英語→社會	26	0.578	1.213	英語=低分組->社會=低分組	無明顯
自然→英語	27	0.730	1.186	自然=低分組->英語=低分組	無明顯
英語→自然	27	0.600	1.334	英語=低分組->自然=低分組	無明顯
國、數→社	24	0.960	1.384	國語=低分組,數學=低分組 ->社會=低分組	無明顯
社→國、數	24	0.615		社會=低分組-> 國語=低分組,數學=低分組	無明顯
國、數→自	23	0.920	1.394	國語=低分組,數學=低分組 ->自然=低分組	無明顯
自→國、數	23	0.622		自然=低分組-> 國語=低分組,數學=低分組	無明顯
國、數→英	20	0.800	1.097	國語=低分組,數學=低分組 ->英語=低分組	無明顯
英→國、數	20	0.444		英語=低分組-> 國語=低分組,數學=低分組	無明顯
國、社→數	24	0.960	1.332	國語=低分組,社會=低分組 ->數學=低分組	無明顯
數→國、社	24	0.585		數學=低分組-> 國語=低分組,社會=低分組	無明顯
國、社→自	23	0.920	1.394	國語=低分組,社會=低分組 ->自然=低分組	無明顯
自→國、社	23	0.621		自然=低分組-> 國語=低分組,社會=低分組	無明顯
國、社→英	20	0.800	1.097	國語=低分組,社會=低分組 ->英語=低分組	無明顯
英→國、社	20	0.444		英語=低分組-> 國語=低分組,社會=低分組	無明顯
國、自→數	23	0.958	1.309	國語=低分組,自然=低分組 ->數學=低分組	無明顯
數→國、自	23	0.560		數學=低分組-> 國語=低分組,自然=低分組	無明顯
國、自→社	23	0.958	1.357	國語=低分組,自然=低分組 ->社會=低分組	無明顯

社→國、自	23	0.589		社會=低分組-> 國語=低分組,自然=低分組	無明顯
國、自→英	19	0.792	1.077	國語=低分組,自然=低分組 ->英語=低分組	無明顯
英→國、自	19	0.422		英語=低分組-> 國語=低分組,自然=低分組	無明顯
國、英→數	20	0.909	1.224	國語=低分組,英語=低分組 ->數學=低分組	無明顯
數→國、英	20	0.487		數學=低分組-> 國語=低分組,英語=低分組	無明顯
國、英→社	20	0.909	1.265	國語=低分組,英語=低分組 ->社會=低分組	無明顯
社→國、英	20	0.512		社會=低分組-> 國語=低分組,英語=低分組	無明顯
國、英→自	19	0.864	1.266	國語=低分組,英語=低分組 ->自然=低分組	無明顯
自→國、英	19	0.513		自然=低分組-> 國語=低分組,英語=低分組	無明顯
數、社→國	24	0.750	1.635	數學=低分組,社會=低分組 ->國語=低分組	無明顯
國→數、社	24	0.8		國語=低分組-> 數學=低分組,社會=低分組	無明顯
數、社→自	28	0.875	1.545	數學=低分組,社會=低分組 ->自然=低分組	無明顯
自→數、社	28	0.756		自然=低分組-> 數學=低分組,社會=低分組	無明顯
數、社→英	23	0.719	1.101	數學=低分組,社會=低分組 ->英語=低分組	無明顯
英→數、社	23	0.511		英語=低分組-> 數學=低分組,社會=低分組	無明顯
數、自→國	23	0.719	1.559	數學=低分組,自然=低分組 ->國語=低分組	無明顯
國→數、自	23	0.766		國語=低分組-> 數學=低分組,自然=低分組	無明顯
數、自→社	28	0.875	1.466	數學=低分組,自然=低分組 ->社會=低分組	無明顯

社→數、自	28	0.717		社會=低分組-> 數學=低分組,自然=低分組	無明顯
數、自→英	24	0.750	1.138	數學=低分組,自然=低分組 ->英語=低分組	無明顯
英→數、自	24	0.533		英語=低分組-> 數學=低分組,自然=低分組	無明顯
數、英→國	20	0.714	1.422	數學=低分組,英語=低分組 ->國語=低分組	無明顯
國→數、英	20	0.666		國語=低分組-> 數學=低分組,英語=低分組	無明顯
數、英→社	23	0.821	1.291	數學=低分組,英語=低分組 ->社會=低分組	無明顯
社→數、英	23	0.589		社會=低分組-> 數學=低分組,英語=低分組	無明顯
數、英→自	24	0.857	1.393	數學=低分組,英語=低分組 ->自然=低分組	無明顯
自→數、英	24	0.648		自然=低分組-> 數學=低分組,英語=低分組	無明顯
社、自→國	23	0.719	1.559	社會=低分組,自然=低分組 ->國語=低分組	無明顯
國→社、自	23	0.766		國語=低分組-> 社會=低分組,自然=低分組	無明顯
社、自→數	28	0.875	1.399	社會=低分組,自然=低分組 ->數學=低分組	無明顯
數→社、自	28	0.682		數學=低分組-> 社會=低分組,自然=低分組	無明顯
社、自→英	23	0.719	1.101	社會=低分組,自然=低分組 ->英語=低分組	無明顯
英→社、自	23	0.511		英語=低分組-> 社會=低分組,自然=低分組	無明顯
社、英→國	20	0.769	1.454	社會=低分組,英語=低分組 ->國語=低分組	無明顯
國→社、英	20	0.666		國語=低分組-> 社會=低分組,英語=低分組	無明顯
社、英→數	23	0.885	1.274	社會=低分組,英語=低分組 ->數學=低分組	無明顯

數→社、英	23	0.561		數學=低分組-> 社會=低分組,英語=低分組	無明顯
社、英→自	23	0.885	1.377	社會=低分組,英語=低分組 ->自然=低分組	無明顯
自→社、英	23	0.621		自然=低分組-> 社會=低分組,英語=低分組	無明顯
自、英→國	19	0.704	1.379	自然=低分組,英語=低分組 ->國語=低分組	無明顯
國→自、英	19	0.633		國語=低分組-> 自然=低分組,英語=低分組	無明顯
自、英→數	24	0.889	1.299	自然=低分組,英語=低分組 ->數學=低分組	無明顯
數→自、英	24	0.585		數學=低分組-> 自然=低分組,英語=低分組	無明顯
自、英→社	23	0.852	1.306	自然=低分組,英語=低分組 ->社會=低分組	無明顯
社→自、英	23	0.589		社會=低分組-> 自然=低分組,英語=低分組	無明顯

資料來源：本研究整理

把以上「表13 各科成績間的關聯規則總表」裡不符合最小支援度或最小機率的部份去除後，整理成「表14 各科成績間的關聯規則總表—整理一」，再來討論規則間有無關聯性。

表14 各科成績間的關聯規則總表—整理一

關聯規則	支援度	機率	重要性	規則	性別關聯
國語→數學	25	0.833	1.295	國語=低分組->數學=低分組	無明顯
國語→社會	25	0.833	1.350	國語=低分組->社會=低分組	無明顯
國語→自然	24	0.800	1.363	國語=低分組->自然=低分組	無明顯
數學→社會	32	0.780	1.586	數學=低分組->社會=低分組	無明顯
社會→數學	32	0.821	1.512	社會=低分組->數學=低分組	無明顯
數學→自然	32	0.780	1.711	數學=低分組->自然=低分組	無明顯
自然→數學	32	0.865	1.536	自然=低分組->數學=低分組	無明顯
社會→自然	32	0.821	1.734	社會=低分組->自然=低分組	無明顯
自然→社會	32	0.865	1.633	自然=低分組->社會=低分組	無明顯
自然→英語	27	0.730	1.186	自然=低分組->英語=低分組	無明顯

國、數→社	24	0.960	1.384	國語=低分組,數學=低分組 ->社會=低分組	無明顯
國、數→自	23	0.920	1.394	國語=低分組,數學=低分組 ->自然=低分組	無明顯
國、社→數	24	0.960	1.332	國語=低分組,社會=低分組 ->數學=低分組	無明顯
國、社→自	23	0.920	1.394	國語=低分組,社會=低分組 ->自然=低分組	無明顯
國、自→數	23	0.958	1.309	國語=低分組,自然=低分組 ->數學=低分組	無明顯
國、自→社	23	0.958	1.357	國語=低分組,自然=低分組 ->社會=低分組	無明顯
數、社→國	24	0.750	1.635	數學=低分組,社會=低分組 ->國語=低分組	無明顯
國→數、社	24	0.8		國語=低分組-> 數學=低分組,社會=低分組	無明顯
數、社→自	28	0.875	1.545	數學=低分組,社會=低分組 ->自然=低分組	無明顯
自→數、社	28	0.756		自然=低分組-> 數學=低分組,社會=低分組	無明顯
數、社→英	23	0.719	1.101	數學=低分組,社會=低分組 ->英語=低分組	無明顯
數、自→國	23	0.719	1.559	數學=低分組,自然=低分組 ->國語=低分組	無明顯
國→數、自	23	0.766		國語=低分組-> 數學=低分組,自然=低分組	無明顯
數、自→社	28	0.875	1.466	數學=低分組,自然=低分組 ->社會=低分組	無明顯
社→數、自	28	0.717		社會=低分組-> 數學=低分組,自然=低分組	無明顯
數、自→英	24	0.750	1.138	數學=低分組,自然=低分組 ->英語=低分組	無明顯
數、英→社	23	0.821	1.291	數學=低分組,英語=低分組 ->社會=低分組	無明顯
數、英→自	24	0.857	1.393	數學=低分組,英語=低分組 ->自然=低分組	無明顯

社、自→國	23	0.719	1.559	社會=低分組,自然=低分組 ->國語=低分組	無明顯
國→社、自	23	0.766		國語=低分組-> 社會=低分組,自然=低分組	無明顯
社、自→數	28	0.875	1.399	社會=低分組,自然=低分組 ->數學=低分組	無明顯
社、自→英	23	0.719	1.101	社會=低分組,自然=低分組 ->英語=低分組	無明顯
社、英→數	23	0.885	1.274	社會=低分組,英語=低分組 ->數學=低分組	無明顯
社、英→自	23	0.885	1.377	社會=低分組,英語=低分組 ->自然=低分組	無明顯
自、英→數	24	0.889	1.299	自然=低分組,英語=低分組 ->數學=低分組	無明顯
自、英→社	23	0.852	1.306	自然=低分組,英語=低分組 ->社會=低分組	無明顯

整理完後還有36條規則，顯然太多，難以做比較觀察。於是，把最小機率調整成0.8，再去掉13條未達到標準的規則，我們就剩下的23條規則來做最後的歸納分析（詳見表「15 各科成績間的關聯規則總表－整理二」）。

表15 各科成績間的關聯規則總表－整理二

關聯規則	支援度	機率	重要性	規則	性別關聯
國語→數學	25	0.833	1.295	國語=低分組->數學=低分組	無明顯
國語→社會	25	0.833	1.350	國語=低分組->社會=低分組	無明顯
國語→自然	24	0.800	1.363	國語=低分組->自然=低分組	無明顯
社會→數學	32	0.821	1.512	社會=低分組->數學=低分組	無明顯
自然→數學	32	0.865	1.536	自然=低分組->數學=低分組	無明顯
社會→自然	32	0.821	1.734	社會=低分組->自然=低分組	無明顯
自然→社會	32	0.865	1.633	自然=低分組->社會=低分組	無明顯
國、數→社	24	0.960	1.384	國語=低分組,數學=低分組 ->社會=低分組	無明顯
國、數→自	23	0.920	1.394	國語=低分組,數學=低分組	無明顯

				->自然=低分組	
國、社→數	24	0.960	1.332	國語=低分組,社會=低分組 ->數學=低分組	無明顯
國、社→自	23	0.920	1.394	國語=低分組,社會=低分組 ->自然=低分組	無明顯
國、自→數	23	0.958	1.309	國語=低分組,自然=低分組 ->數學=低分組	無明顯
國、自→社	23	0.958	1.357	國語=低分組,自然=低分組 ->社會=低分組	無明顯
國→數、社	24	0.8		國語=低分組-> 數學=低分組,社會=低分組	無明顯
數、社→自	28	0.875	1.545	數學=低分組,社會=低分組 ->自然=低分組	無明顯
數、自→社	28	0.875	1.466	數學=低分組,自然=低分組 ->社會=低分組	無明顯
數、英→社	23	0.821	1.291	數學=低分組,英語=低分組 ->社會=低分組	無明顯
數、英→自	24	0.857	1.393	數學=低分組,英語=低分組 ->自然=低分組	無明顯
社、自→數	28	0.875	1.399	社會=低分組,自然=低分組 ->數學=低分組	無明顯
社、英→數	23	0.885	1.274	社會=低分組,英語=低分組 ->數學=低分組	無明顯
社、英→自	23	0.885	1.377	社會=低分組,英語=低分組 ->自然=低分組	無明顯
自、英→數	24	0.889	1.299	自然=低分組,英語=低分組 ->數學=低分組	無明顯
自、英→社	23	0.852	1.306	自然=低分組,英語=低分組 ->社會=低分組	無明顯

參、最後歸納分析：

- 一、國語單科或國語與另一科一起影響數學、社會、自然與科技的關聯規則次數出現最多，出現10次；其次是社會8次、自然與科技8次、數學6次。可見，國語學不好，數學、社會、自然與科技也

跟著學不好；反觀，數學、社會、自然與科技學不好，國語未必學不好。可見國語科的確對其他三科的學習是有極大關聯的。由這些現象我們可以推論國語科屬於基礎工具學科，語文能力不佳的確會對知識的吸收與理解造成困難，因此，對其他學科的學習是會產生某種程度上的影響。

二、自然與科技和社會單科或與另一科一起影響其他科目的關聯數目也不少，這跟我們補救教學裡只偏重在國語、數學兩科做補強的舊有認知觀念相較，顯然有相當大的落差。

三、英文科就單科對其他科的關聯數目來看，是找不到符合最低支援和最低機率的規則，足見對其他四科的關聯性不大。

四、性別納入學科互相學習的考量後，並未找出有高相關的關聯性，可見學科之間的相互學習不因性別不同而有明顯差異。

第五章 結論與建議

第一節 研究結論

本研究採礦的科目計有國語、英語、數學、社會、自然與科技五科，最小支持度設定在23（5%）、最小機率設定在0.8（80%）。反覆建模、調整輸入變數，同時搭配使用關聯法則採礦裡「規則」和「相依性網路」兩個標籤頁做觀察、篩選。最後，就得到較有意義的23條關聯規則，嘗試歸納解釋看到的現象：

- 壹、性別一欄與其他五個科目搭配做輸入選項（Input），嘗試預測性別是否會對科目之間的相互學習產生關聯？。以常理來看，一般人普遍認為男生的數理邏輯能力較佳，女生的語文應用能力較好，在相關文獻與本研究第四章第一節中，指的都是單科成績的表現，也都獲得許多實證。但經由本研究關聯法則採礦的結果，是看不出有任何高相關性的，也就是說，性別不會對科目之間的相互學習產生明顯的關聯，這是頗出乎預料的。
- 貳、國語科成績不好（低分組），對數學、自然與科技、社會三科（均為低分組）的成績也有顯著關聯，這個結果符合我們一般期待。這種情況最有可能的解釋是：國語科歸類在基礎學習工具一類，算是各科學習必備的基礎知識。因此，整體語文的聽、說、讀、寫能力表現差的人，對於文章內容理解的能力一定也較差。所以，對於其他三個科目的學習表現一定也有相當程度的關聯，進而影響學習成效。

參、英語未達到最小支持度23或最小機率0.7的條件，因此找不到比較有意義的規則，可見英語成績的高低對其他四科關聯不大。推究可能原因，有可能是國小目前的課程內容和評量的題目中，較少有運用到英文詞彙的機會；或者是這些學童本身的英語學習能力就不佳。因此，才沒有產生任何較有意義的規則。

肆、自然與科技和社會對其他科目學習的關聯性也不小，也發現數學、自然、社會三科彼此間的學習互動極深。

第二節 研究建議

根據本研究資料採礦的結果，提出幾點建議：

壹、對於國語不好，數學也不好；國語不好，自然與科技也不好；國語不好，社會也不好的這3條關聯規則的現象，我們集合符合這些條件的學童來加強語文的聽、說、讀、寫能力，對其他科目的學習成績必能有所提昇。

貳、對於國語科成績不好，數學、自然與科技、社會等三科成績也會不好的這一顯著關聯，本研究建議在未來課程規劃上，該對國語課的授課節數做提昇的調整；此外，若有補救教學的配套，順序也應該把國語科列為優先考慮。全面提昇語文能力的結果，將有助於提昇全體學童對其它科目學習的表現。而且，在成績明顯提昇的同時，也有助於減少在學學童流失、增加新生入學意願。

參、「家庭社經地位」、「學習態度」等一些變數，本研究因為資料取得的關係並未納入研究的考量，建議對此議題有興趣者，可採

問卷做輔助，取得這些變數值來做後續深入之研究。

參 考 文 獻

一、中文部份

- [1] Microsoft 中文網站 Microsoft SQL Server：什麼是SQL Server 2005？
<http://www.microsoft.com/taiwan/sql/prodinfo/overview/what-is-sql-server.msp>
- [2] UniMiner Data Mining 資料採礦介紹
<http://www.uniminer.com/center01.htm>
- [3] 中華資料採礦協會
<http://www.cdms.org.tw/xoops2/html/modules/news/>
- [4] 內政部九十九年第六週內政統計通報(98年嬰兒出生狀況統計)
http://www.moi.gov.tw/stat/news_content.aspx?sn=3829
- [5] 尹相志，SQL Server 2005 Data Mining資料採礦與Office 2007資料採礦增益集，台北，悅知文化，民國96年。
- [6] 余民寧，「影響學習成就因素的探討」，教育資料與研究，73期，11-24頁，民國95年。
- [7] 林金火，「資料探勘技術應用於中學生成績之研究」，立德管理學院應用資訊研究所碩士論文，民國96年1月。
- [8] 邱佳椿，「高職電機電子群學生學習態度與學業成就之路徑模式研究」，國立臺灣師範大學工業教育研究所碩士論文，民國92年。
- [9] 陳美娥，「國小學習遲緩兒童業成就之關係研究」，國立台北市立師範學院初等教育學系碩士論文，民國85年。
- [10] 陳雅雯，「中部地區國小高年級學童自我概念、學習動機與學業成就關係之研究」，台中師範學院諮商與教育心理研究所碩士論文，民國92年。
- [11] 孫清山、黃毅志，「補習教育、文化資本與地位取得」，台灣社會學刊，第十九期，95-139頁，民國92年。
- [12] 張春興，教育心理學—三化取向的理論與實踐，台北，東華，民國83年。
- [13] 黃富順，「影響國中學生學業成就的家庭因素」，台灣師範大學教育研究所集刊，16卷0期，383-486頁，民國63年。
- [14] 溫侑柯，「應用資料探勘之關聯法則探討大學入學成績對在學成績的影響—以資管系為例」，南華大學資訊管理研究所碩士論文，民國95年。
- [15] 楊景森，「國中學生家庭氣氛家長參與及學業成就之研究」，彰化師範大學教育研究所碩士論文，民國92年。
- [16] 蔡政忠，「國中學生父母管教方式、自我調整學習策略與學業成

- 就之關係研究」，彰化師範大學教育研究所碩士論文，民國93年。
- [17] 鄭景邁，「資料探勘技術於學童補救教學之研究」，立德管理學院應用資訊研究所碩士論文，民國96年。
- [18] 謝孟穎，「家長社經背景與學生學業成就關聯性之研究」，國立嘉義大學國民教育研究所碩士論文，民國91年。
- [19] 簡茂發，「高級中學學生家庭社經背景、教師期望與學業成就之關係」，台灣師範大學教育研究所集刊，26卷0期，1-97頁，民國72年。
- [20] 譚康榮，「誰家小孩學習成就最高？哪群學生心理最不健康？—台灣教育長期追蹤資料庫的最新發現」，中央研究院學術諮詢總會通訊，第13卷第1期，86-91頁，民國93年。
- [21] 鐘金燕，「運用資料採礦技術於創新服務MOD採用之研究」，台北大學資訊管理研究所碩士論文，民國95年。



二、英文部份

- [22] Agrawal, R. and Srikant, R., "Fast Algorithms for Mining Association Rules", Proc. of the 20th VLDB Conference Santiago, pp. 487-499, 1994.
- [23] C. Kleissner, "Data mining for the enterprise", Proc. of the Thirty-First Hawaii International Conference on System Sciences,(HICSS-31), Volume 7, pp.295-304, 1998.
- [24] Chen, C. S., Lee, S. Y., and Stevenson, H. W. , *Academic achievement and motivation of Chinese students: A cross-national perspective*3., In S. Lau (Ed.), Growing up the Chinese way: Chinese child and adolescent development. Hong Kong: The Chinese University, 1996.
- [25] Grupe, F.H. and Owrang, M. M., "Database Mining Discovering New Knowledge And Cooperative Advantage", Information Systems Management, Vol.12, NO.4, pp.26-31, Fall 1995.
- [26] Hao, L., and Bonstesad-Bruns, M. , "Parent-child differences in educational expectations and the academic achievement of immigrant and native students.", *Sociology of Education*,7, 175-198, 1998.
- [27] Frawley, W.J., Paitetsky-Shapiro, G. and Matheus. C. J., "Knowledge Discovery in Databases: An Overview", *Knowledge Discovery in Databases*, California, AAAI/MIT Press, pp.1-30, 1991.
- [28] Berry, M.J., and Linoff, G., *Data Mining Techniques for Marketing, Sales, and Customer Support*, USA, John Wiley, 1997.
- [29] Stickney, B. D., and Fitzpatrick, J. , *Coloma's inequality twenty years later: The origins, the issues and the implications*, Paper presented at the annual meeting of the American Educational Research Association. U.S.: Colorado, 1987.