教育研究與發展期刊 DOI 10.3966/181665042015091103001
第十一卷第三期 2015年9月 頁1-30

資料採礦模式於學校整併指標之應用與評估

林松柏 國立暨南國際大學教育政策與行政學系助理教授

摘 要

因應少子女化的衝擊，小型學校進行整併或裁撤已是必要策略之一，教育部遂於 2006 年 2 月 14 日提出小型學校發展評估指標，供各縣市政府參考運用。在運用指標進行分析時，若能有大數據的思維，並發展適切的資料採礦模式，將有助於各縣市政府進行學校整併。本研究的研究目的即探討如何基於大數據思維整合不同資料庫，將資料採礦技術運用於教育統計資料中，以利學校整併工作的執行。本研究依據教育部小型學校發展評估指標，整合現行不同資料庫針對個案縣市轄區內所有國民小學進行相關資料蒐集。本研究所運用的資料採礦模式有分類與迴歸樹、類神經網路、決策樹、支援向量機、貝氏網路等五種，研究結果發現五種模式具有正確率高與便於解讀的優點。依據研究結果，本研究提出學校整併應整合教育、人口與地理資料庫，並且應採實徵資料評估與實地訪察兩階段評估，而縣市政府或學校能夠運用本研究發展的操作型定義釐清有整併需求的學校名單或了解學校本身的相對位置。

關鍵詞：大數據、資料採礦、學校整併

林松柏電子郵件：splin@ncnu.edu.tw

Journal of Educational Research and Development
September 30, 2015 , Vol.11 No.3 pp. 1-30

Application and Assessment of Data Mining Models in School Consolidation Indicators

Sung-Po Lin
Assistant Professor, Department of Educational Policy and Administration National Chi Nan University

Abstract

Because of tendency of declining birthrate, it is seen as necessary to consolidate or abolish the small schools. The Ministry of Education then provided "Small School Development Evaluation Indicators" to county and city governments in February 2006. In depth analysis of the indicator data based on Big Data to develop data mining analysis model and operational definition of each indicator, is helpful for county and city governments consolidating small schools. This article aims to study how to integrate different databases based on Big Data thinking, and use data mining methods in education statistics, to facilitate school consolidation. According to the Ministry of Education indicators, this article integrated governance databases to collect the related data of all elementary schools. This article used supervised models, including Classification and Regression Tree, Neural Network, Decision Tree, Support Vector Machine, and Bayesian Network. The results reveal that five models have higher correction rate and are easy to read. According to the results, when consolidating small schools, education, population and geographic databases should be integrated. Besides, empirical data assessment and supervision should be adopted. The governance institution and each school can adopt operational definition of each indicator to calculate the relative position.

Keywords: big data, data mining, school consolidation

Sung-Po Lin's E-mail: splin@ncnu.edu.tw

壹、前言

繼農業革命、工業革命與數位革命後，當前正值資料革命。世界充斥著大量的資料，資料取得不再是問題，但凸顯要如何找出相關資訊的問題（Franklin & Andrews, 2012）。此時以「大數據」（big data，亦稱巨量或海量資料）來衡量每件事的思維扮演相當重要角色，大數據時代的到來意謂人類社會正值巨大轉變，亦促使社會行為科學的研究取徑呈現了嶄新面貌。一般調查研究方法多仰賴面對面訪談、電話訪談、郵寄問卷等蒐集實徵資料（Fowler, 2002），但隨著電腦科技、網際網路技術與雲端運算的發展，大型資料庫的使用已逐漸普及，也因其能夠蒐集更大量與多元的研究資料，故受到各界的重視。政府是最早巨量資料的蒐集者，但其使用效率並不彰顯，故美國、英國、歐盟與澳洲等國都有開放資料的措施與政策，結合企業或民間機構的運用，使資料價值最大化（Mayer-Schönberger & Cukier, 2013）。因此，運用政府所蒐集的統計資料庫，進行相關研究將是未來研究工作的發展趨勢。

在大數據的思維下，資料採礦（data mining）的技術常被使用。一般金屬的採礦流程依序包括考查地質、鑿坑道、炸開礦脈、運送與篩選礦石、處理精砂、提煉、製成金磚，而資料採礦的流程即如同採礦一般，必須進行資料考查、篩選數據變項、運用方法、建立模式等，即運用不同的統計方法，整合一套適切的流程，以提煉出具有品質與意義的分析結果（牛田一雄、高井勉、木暮大輔，1996；謝邦昌（主編），2005；Kennedy, Lee, Roy, Reed, & Lippman, 1997）。然而，在眾多的統計方法中，如何擷取適用的方法，並整併成適切的分析流程，形塑資料採礦模型，以挖掘出具有意義的訊息，將是最重要的課題。

因應少子女化的衝擊，小型學校進行整併或裁撤已是必要策略之一，教育部遂於 2006 年 2 月 14 日提出小型學校發展評估指標，供各縣市政府參考運用。然而，各縣市各校的背景差異甚大，難以單一指標套用於所有學校，而是應以周延多元的數據進行分析，但卻會產生分析資料變得龐雜的問題。因此，在進行這些指標資料分析時，若能有大數據的思維，並且發展適切的資料採礦模式，進一步定義各項指標，以及進行更明確的劃分，將有助於各縣市政府進行學校整併的工作。本研究的主要研究目的有三，茲分述如下：

一、探究大數據思維對教育研究的改變與運用；

二、了解如何運用資料採礦針對現行學校整併指標進行分析；

三、提出改善與充實現行學校整併工作的具體建議。

貳、文獻探討

本研究旨在探討如何在大數據的思維下，運用資料採礦的方法，並用於分析縣市的學校實徵資料以利學校整併的工作。在文獻探討一節中，將先針對大數據的方法論進行簡介，其次則是說明如何運用資料採礦分析模式，最後再論述學校整併工作的相關內容。

一、大數據思維對教育研究領域的意義

由於智慧型手機的普及、網路社群服務的增加，以及分析資料技術的提升，使得多樣化的非結構性資料可以預測與活用，此即為大數據應用的濫觴（宋吉永，2013）。基於大數據的思維，改變了只能依賴少量資料規模的傳統決策行為，而且也將會改變現有市場、組織，以及市民與政府間的關係等（Mayer-Schönberger & Cukier, 2013）。暴增的資料量使得當今世界變得感知化（instrumented）、物聯化（interconnected）和智能化（intelligent），所有的物體都能被感測，而感測的大量數據都被傳送處理，最後則是從龐雜的資料中分析出有用的資訊，以做出決策（胡世忠，2013）。

大數據具有巨量性（volume）、多樣性（variety）與即時性（velocity）三個特色（城田真琴，2013；Loshin, 2013）。此外，有鑑於資料真實性與可靠性的問題，第四個不確定性（veracity）也開始受到重視（胡世忠，2013）。在分析資料時，大數據有三項思維，首先是龐大資料的分析能力，其次是不求精準的雜亂資料，最後是重相關而非因果關係；巨量資料是指全部或儘量完整的資料，但也因此可能有不精確或格式不一的資料，而藉由相關分析的運算，能夠快速地篩選出最佳指標（Mayer-Schönberger & Cukier, 2013）。但相較於只在意資料的多寡，資料的多樣性與更新速度快等特性更加值得探討，並且巨量資料必須透過分析才有意義，而不只是單純地將資料累積在一起，所以巨量資料的分析必須透過觀察、量化，以及最後的演繹、推論階段，才能挖掘到有意義的結果（宋吉

永，2013）。而在蒐集大量資料時，也應思考是否存有哪些不確定因素，減少分析時雜訊的干擾（Silver, 2012）。此外，大數據分析一個很重要的思維是，大量的資料比更優秀的演算法有用，亦即統計上最適配的結果不是藉由複雜高深的演算法組合辨識輸入的資料，而是對大量儲存的正確資料進行分析（城田真琴，2013）。

大數據有賴於完整的資料庫，當前教育部或教育局處均提供完整的學校資料，並且教育部統計處亦提供相當完整的各校歷年統計資料。從這些資料中可以一窺臺灣教育發展的脈絡，而針對這些資料進行深入的分析，亦可發現未來的教育發展趨勢。除此之外，為了做出即時妥適的決策資訊，必須在短時間內從龐大的資料庫中挖掘（篩選）出合適的分析變項，並且還得因應不同的情境而考量不同的分析變項。而傳統的研究方法可能耗時較長，也無法即時改變分析變項，若要調整分析變項，則必須重新進行研究設計，所以此點是能藉由大數據思維分析資料庫來改善傳統研究方法的局限性。因此，以大數據思維做為教育決策的引領方向，在當前的科技發展背景下，已提供了適切的平臺。而如何在龐大的資料中，擷取適切的分析變項，則有賴於資料採礦的技術。

二、資料採礦分析模式在教育研究領域之運用

大數據思維對當今教育研究工作者應有其必要性，並且資料採礦的運用也隨著大型資料庫的建置逐漸普及，所以運用資料採礦的知能也相當重要。資料是新一代的石油，是當代最重要的資產，也是組織最重要的策略資產。即使如此，石油仍必須經過提煉才能成為能源，資料則是應經過資料採礦的分析才能成為資產。資料採礦整合了不同的統計分析技術，試圖從龐雜的數據中建構出有效的模式。在大數據的背景下，資料採礦的技術更顯得重要與適切。

資料採礦的功能主要有集群、分類，以及預測等，其模式通常分為描述性模式與預測性模式兩類，常用的模式包括分類模式、迴歸模式、群聚模式、關聯模式、序列模式、時間序列模式（張云濤、龔玲，2007），亦可分為探索、分析與整理等主要步驟，在操作過程中包含商業理解、資料理解、資料準備、塑模、評估，以及展開等階段（牛田一雄、高井勉、木暮大輔，1996）。亦有學者主張將資料採礦的流程分為問題定義、資料蒐集與選取、資料準備、採礦方法選擇、資料訓練／測試或演算法運用、最終模式評鑑與整合等步驟（Kennedy, Lee, Roy,

Reed, & Lippman, 1997）。謝邦昌（主編）（2005）認為資料採礦的步驟包括理解資料與進行的工作、獲取相關知識與技術、整合與查核資料、去除錯誤或不一致的資料、發展模式與假設、實際資料採礦工作、測試與檢核所採礦的資料、解釋與使用資料等步驟。資料採礦通常需要四個步驟，首先是定義問題，其次是資料選擇，第三是資料處理，最後是知識擷取（胡世忠，2013）。

簡言之，資料採礦的分析流程主要包括四個步驟，分別是定義問題、資料篩選、塑模與評估，以及模式整合與解釋等。首先，定義問題係先就研究議題進行界定，如同在採礦前所進行的地質考查，研究議題的界定亦可使研究者釐清分析的標的，並且蒐集合適的統計資料，以及透過文獻探討了解不同變項之間的關聯性；其次，資料篩選則是實際採礦的準備工作，如同在龐大的礦山中篩選出礦石般，此時研究者則要過濾可以使用的資料庫，整理統計資料並選擇可用的研究變項，進而對變項進行統計尺度的定義，提出變項的操作型定義；第三，塑模與評估則是將依據描述性或預測性模式，採用合適的統計分析方法，如同採礦階段中運用冶煉技術將礦石中的金屬提煉出來過程；最後，模式整合與解釋則是資料採礦的最後步驟，相當於採礦過程的製成金磚，研究者依據研究目的將不同的統計分析方法整合成為模式，並依據模式分析結果提出解釋。

雖然目前在教育領域運用資料採礦分析模式的相關研究並不多見，但國內外已有數篇值得參考的研究論文，但較多著重於預測學生的學業表現與中輟率等。例如，研究主題為預測大學新生學業表現（Vandamme, Meskens, & Superby, 2007）、藉由學生的學習資訊檢視其學習表現與方法（West, 2012）、了解學生學習改變情形與檢視其表現（Picciano, 2012）、探討閱讀素養的影響因素（張鈿富、林松柏，2012）、預測網上大學（online university）學生中輟率（Niemi & Gitin, 2012）、學生流失的偵測（邱宏彬、許依宸，2011）、中途輟學線上預警系統的建立（翁秉逸、謝明哲，2013）。其他亦有互動式多媒體學習系統在教學上的運用（Chrysostomou, Chen, & Liu, 2009）、學生在線上學習環境中的學習模式（Hung & Crooks, 2009）、改善師培生的教學訓練（Masunaga & Lewis, 2011）、發展學生學習投入態樣類型（Luan, Zhao, & Hayek, 2009），以及改善教師的教學（Cohen, 2003）等研究主題。然而，將資料採礦運用於學校整併的研究主題仍非常少見。

三、學校整併指標發展

由於人口結構快速的轉變，已對臺灣社會產生重大衝擊。人口結構呈現「少子女化」、「異質化」與「高齡化」的三化特徵，尤以少子女化對教育產生衝擊為烈，例如學校招生不足、學校空間閒置、教育投資浪費、城鄉教育落差等問題。為了解決生源不足，小校裁併的政策遂被提出並落實，然基於學生受教權、家長選擇權、教職員工作權、地方發展等問題，勢必不能輕易裁撤任何一所學校。因此，小校裁併需要審慎思考，並參酌周延的指標資料才能擬定合宜的作法。

借鏡美國學區整併的經驗，相關研究指出學校整併確實有助於學生的表現（Leacha, Paynea, & Chan, 2010），而且依據社會資本的概念，家長認為學校整併能具有一些正面的效益，子女能夠很快地適應轉變，並且有更多的朋友、競技場與發展機會（Surface, 2011）。面臨註冊人數下降與經費縮減等問題，凸顯出學校組織重整與整併的重要性，而學校整併與成效、經濟、學生成就、學校規模與社區認同有重要的關係（Bard, Gardener, & Wieland, 2006）。學校整併最大的挑戰在於教師的移動，故學校整併應顧及師生的意願，若能滿足師生的期待，那麼將有利於學校整併作業，例如提供學生更多元的課程與社交機會，減少教師的授課時數以利其準備專業發展的機會等（Wrobel, 2010）。此外，Adams 與 Foster（2002）針對肯塔基州（Kentucky）的學區整併進行研究，研究發現關鍵因素是學校財務狀況，而非學區大小，並且進行小型學校整併並不見得可以減少教育經費支出。然而，Zimmer、DeBoer 與 Hirth（2009）針對印第安納州（Indiana）學校整併的研究則指出，整併註冊學生數低於 2,000 人的小型學區將有明顯的效益。

最常被提出學校整併考量因素為學校規模，其中又以在學人口數為影響規模經濟之關鍵，亦即學生人數必須維持一定比例，才能產生學校規模經濟；國民小學經營規模不經濟應為 100 人以下，班級數 6 班以下學校（吳政達，2006）。除了學校規模外，仍應該考量地區人口狀況的差異，依據不同學校規模建構「合適規模」理論（林雍智，2006）。通常會被裁併的學校大多位處偏鄉，而學校被裁併使得居民的子女需要花更多的時間上下學，形成另一大負擔，那麼裁併校的結果可能使弱勢更弱勢（劉世閔，2012）。其實家長並不反對子女跨區到優質明星學校就讀，而是反對到另一個偏遠學校就讀，故若能提供優質的教育環境，以及

住宿或交通接送，取得家長的認同後，再逐步讓學校達到理想的經營規模，並使其他鄰近學校自然縮編，將不致於損害學生的基本受教權（張志明，2012）。此外，學校與社區發展息息相關，裁併任何一所學校應與社區居民討論協商，並取得社區居民的共識，才能獲得最大的助益（吳政達，2006；林雍智，2006）。

為了協助縣市政府評估小校裁併，教育部於 2006 年 2 月 14 日公布小型學校發展評估指標，評估指標分為一般性與特殊性，一般指標指學校的一般性條件，如學生數、社區結構、原校區之用途、社區對學校的依賴度等共 11 項，分別給予評分，如評分愈低，即表示學校愈應考慮進行整合；而特殊條件則指不宜整併的因素，如該鄉鎮只有一所小學、原住民地區學校或到鄰近學校交通有重大安全顧慮等，只要符合其中任何一項指標，學校不宜進行整合（中華民國監察院，2012）。此評估指標內容除了考量到學校規模、地區、交通與社區等因素外，亦涵蓋了其他因素，如校齡、學生趨勢、教室屋齡、學校用途、學校所在地環境等，可謂相當完整。然而，在實務運用可能會因為缺少具體明確的操作型定義，使得計算標準不一，例如學生數趨勢分為劇增、略增、平穩、略減、劇減等五個程度，但並沒有指出要以多長的時間為依據，亦無說明多少人以上或以下是屬於「劇」或「略」的程度，社區結構的指標內容亦有相同的問題。此外，像是社區對學校之依賴亦沒有一致的評估標準，容易使各校的認知解讀不同。指標分數則均採五等第的計分方式，使得各項指標的權重分數均是相同的，將無法凸顯不同指標之間的差異性，也可能消弭計分結果的差異程度而無法進行比較。

大數據的分析思維需要藉由完整的資料庫，以及適切的資料採礦分析模式。當前教育部與各縣市政府教育局／處均已提供豐富的學校資料，亦會逐年更新，已符合大數據的巨量性、多樣性與即時性的三個特色，再依據教育部小型學校發展評估指標與相關文獻探討，則能更快掌握分析的變項，降低蒐集資料過程中的雜訊干擾，避免不正確性的問題。而目前還有待釐清各指標的操作型定義與權重，以及發展適切的資料採礦分析模式。因此，本研究藉由大數據思維整合相關資料庫，並再將這些資料根據現行教育部小型學校發展評估指標轉化為具體明確的操作型定義，以及運用資料採礦模式分析各指標的權重分數。此外，學校整併相關研究大多係以學生人數做為主要的討論指標，但本研究則納入其他分析變項，透過明確具體的計算方式，有利於各縣市政府與學校能夠自行檢視。

參、實徵資料分析

本研究旨在以大數據的思維，運用資料採礦的方法，計算學校整併指標分數，在實徵資料分析則以個案縣市所轄的所有國民小學為例（因基於分析資料須保密之故，故以個案縣市匿名標示）。該縣市共有 20 至 30 個行政區，但因幅員遼闊，各行政區的發展差異甚大，更遑論不同行政區內的各級學校之間的差異性。個案縣市於 2013 年依據教育部小型學校發展評估指標針對縣市內的所有國民中小學進行整併評估，除了原本教育部的指標外，還加入了教學品質（例如：各校在當年度的學科能力檢測結果，以五等第級分計算，分數愈高代表教學品質愈佳）與學校財務狀況（例如：各校的人事費用、退休撫卹、其他基本需求、先期計畫及收支對列數等經費的合計除以學生人數所得之每生單位成本）兩項指標，接著則依據分析結果組成專案小組，針對被列入風險警戒的學校進行視導，以即早因應少子女化的衝擊。此外，個案縣市的各校統計資料能夠透過公開的統計資訊網取得，除了有利於研究分析外，各校也能夠自行計算指標分數，了解本身的相對位置。因此，研究者針對個案縣市轄區內所有國民小學進行相關資料蒐集，再運用不同的資料採礦方法進行分析，並比較不同模式的分析結果。以下將說明實徵資料的來源、內容，以及分析方法。

一、分析資料來源

大數據的特徵包括巨量性與多樣性，此需要倚賴資料庫的內容，但只憑單一的資料庫是無法取得所有的資料，所以必須將不同的資料庫進行比對與整合。這也符合在大數據時代中，整體會比部分更有價值的觀點，亦即結合多個資料集的價值會大於只用單一的資料集（Mayer-Schönberger & Cukier, 2013）。然而，也不應該漫無目標地囊括所有資料庫，如此將會失之焦點，無法凸顯問題的重點，並且可能會使模式有過度學習的問題，誤把雜訊當資訊，對於資料已顯示的訊息做過多的假設（Silver, 2012）。因此，在進行實徵資料分析的第一個階段，即先定義研究問題，如同應先確認礦脈的位置，才能進行採礦的道理。此外，先定義好問題，也能減少雜訊過多，且避免干擾分析品質的缺點。

本研究針對學校整併指標內容進行分析，考量的因素有學生人數、學校所在地區、交通狀況，以及社區認同等，而相關的資料庫應包括學校資料、師生資

料、人口資料，甚至是地區資料，這些資料庫均公開於各縣市政府的全球資訊網統計資料頁面，並且會定期更新。本研究以學校為單位，再根據相關資料變項進行資料庫的整合，包括「縣市政府教育統計資訊網」，以及「縣市政府人口統計資訊網」兩種資料庫。前者提供的資料主要包括各級學校資料、教職員人數、學生人數、學區，以及學校背景等，本研究資料蒐集時間為 2013 年 12 月至 2014 年 4 月間；後者提供的資料則有里鄰戶數與人口數、單一年齡人口數、人口結構、遷入遷出統計、出生死亡統計等，資料蒐集時間為 2014 年 2 月。此外，針對地區性的資料庫，研究者係以 Google 蒐尋引擎逐筆輸入各校地址加以記錄，自行建構地區性資料庫，資料蒐集的時間為 2013 年 12 月至 2014 年 1 月間。

二、分析資料篩選

如何克服過多的雜訊，在進行指標篩選時應有所依據，通常有賴文獻理論的支持，而本研究主要以教育部小型學校發展評估指標為依據，此指標係教育部委託專家學者所建構的指標，並且由各縣市所運用。如此一來將有所依歸，也可以減少雜訊的問題。教育部小型學校發展評估指標區分為一般性指標與特殊性指標，前者又包括 11 項指標內容，後者則有四項，總共是 15 項指標內容。一般性 11 項指標中的「學生數」、「學生數趨勢」、「校齡」與「整合後之學校是否需再增建教室及充實設備」的統計數據能夠直接在縣市政府教育統計資訊網取得或換算；「社區結構」與「社區對學校之依賴度」則是整合縣市政府教育統計資訊網，以及縣市政府人口統計資訊網兩者資料庫的統計數據；「距公立學校距離」與「與鄰近學校間有無公共交通工具」則是藉由 Google 蒐尋引擎取得統計數據；「小型學校大部分教室屋齡」、「原校區之用途」與「其他（如歷史背景）」三項指標無法從現有的資料庫中取得，故將此三項視為補充性指標，即待分析結果完成後，再視需要整併的小型學校蒐集相關的資料，以進一步評估整併的可能性。四項特殊性指標「該區只有一所小學或國中」、「原住民地區學校」、「到鄰近學校交通有重大安全顧慮(如經過土石流危險區域)」與「其他」則是因符合其中一項即減少整併的可能性，故本研究將此四項指標整併為特殊性指標，並不列入原始指標分數計算，亦不列入分析模式中。以下將分別說明八項列入分析模式的指標內容操作型定義，以及原始指標分數的計算方式：

（一）學生數

個案縣市各行政區各國民小學 2013 學年全校學生數，原訂的評估指標分數計算方式係以人數為基準，分為 81 人以上、61～80 人、41～60 人、21～40 人，以及 20 人以下，人數愈高，指標分數愈高，依序為 5 分至 1 分。而在資料採礦分析中，則不將人數轉換為區間變項，仍以各校學生數做為連續型的輸入變項（同樣命名為「學生數」）。

（二）學生數趨勢

個案縣市的縣市政府教育統計資訊網並沒有提供學生數趨勢的單一變項，故研究者依據各國民小學 2010～2013 學年第一學期的一年級學生數（新生數），計算平均增長率（T），本研究的計算方式如下：

$$T = \frac{\Sigma \frac{P_x - P_{x-1}}{P_{x-1}}}{N_{year}}$$

註：PX=單一年度的個數；PX-1=前一年度的個數；Nyear=年度個數。

為比較各校的趨勢係數幅度，研究者在計算各校的平均增長率後，再以相對比較的概念計算各校的百分等級分數，並轉化為五等第分數。指標分數 5 分代表學生數趨勢為劇增，4 分為略增，3 分為平穩，2 分為略減，1 分為劇減。而在資料採礦分析中，為更充實該指標的內涵，故輸入各校在 2010～2013 學年的新生人數與平均增長率做為連續型的輸入變項，並分別命名為「新生人數」與「學生數趨勢」。

（三）社區結構

社區結構的操作型定義為各國民小學學區人口數結構趨勢，因當前的資料庫並沒有提供此變項的數據，故研究者首先在縣市政府教育統計資訊網蒐集各校的學區，接著在縣市政府人口統計資訊網中，蒐集 2010 年 12 月至 2013 年 1 月各行政區各里的人口數，最後則是計算各校學區的人口數趨勢，計算公式如同學生數趨勢。在資料採礦分析中，為更充實該指標的內涵，故輸入各校學區 2010 年至 2013 年的平均人口數與平均增長率，分別命名為「社區人口數」與「社區結構趨勢」，做為連續型的輸入變項。

（四）距公立學校距離

因縣市政府教育統計資訊網與縣市政府人口統計資訊網兩個資料庫中均無提供此指標的內容，故研究者以各國民小學所在行政區為範圍，利用 Google 地圖的規劃路線功能，依次查詢兩校間的行車距離，選取最短距離為計算標準。若兩校之間的行車距離在 3 公里以上，指標分數則為 5 分；介於 2.1～3 公里，分數為 4 分；介於 1.6～2 公里，分數為 3 分；介於 1～1.5 公里，分數為 2 分；在 1 公里以內，分數為 1 分。在資料採礦分析中，則輸入兩校間的最短行車距離，並將之命名為「學校距離」做為連續型的輸入變項。

（五）與鄰近學校間有無公共交通工具

在此指標的資料蒐集上，同樣透過 Google 蒐尋引擎取得。以各校所在行政區為範圍，利用 Google 地圖的規劃路線功能，查詢兩校間的距離是有無大眾交通工具的乘運點。若距離最近的兩校之間沒有大眾交通工具，則指標分數為 5 分，有則是 1 分。在資料採礦分析中，則輸入間斷型的變項，0 代表無公共交通工具，1 則代表有公共交通工具，輸入變項名稱為「交通工具」。

（六）校齡

資料來源係在縣市政府教育統計資訊網中，查詢各國民小學的創校日期，再以 2014 年為基準計算各校的校齡。若校齡在 81 年以上，指標分數為 5 分；介於 61～80 年，分數為 4 分；介於 41～60 年，分數為 3 分；介於 21～40 年，分數為 2 分；在 20 年以下，分數為 1 分。在資料採礦分析中，則輸入各校的校齡，並同樣命名為「校齡」做為連續型的輸入變項。

（七）整合後之學校是否需再增建教室及充實設備

因在資料庫中均無明確的變項，故研究者係計算各校學生人數與現有普通教室數的比率，並依其計算五等第分數，做為本指標的操作型定義內容。比率等第高者代表需要增建教室及充實設備，指標分數得分愈高，比率等第低者則代表不需要，即指標分數得分愈低。在資料採礦分析中，則輸入各校學生人數與現有普通教室數的比率，做為連續型的輸入變項，變項名稱則命名為「整合比率」。

（八）社區對學校之依賴度

此指標在資料庫中亦無明確的變項，故研究者蒐集 2010 年 12 月至 2013 年 1 月各行政區的各里單一年齡人口數，包括 6 歲至 11 歲的六筆資料，再分別計算各校一年級至六年級學生人數所占的比率，即各校在各學區的平均招生率，接著

再依等第分數分為高、中與低，得分為 5 分、3 分及 1 分。在資料採礦分析中，則輸入各校的平均招生率，並將之命名為「學區平均招生率」做為連續型的輸入變項。

三、分析方法塑模與評估

本研究在資料採礦分析模式中所使用的預測變數為個案縣市委託研究計畫報告的評估結果名單，其評估的原則主要有三，分別是法規依據、各區權重分數排序，以及學生數趨勢推定，並依據各校入學區招生率、未來新生人數推估、裁併校權重分數，以及加入教學品質與學校財務狀況兩項指標，進行整體性考量。分析的方法是以巢狀 if-then-else 的陳述規則進行評估，評估結果區分為「風險警戒」、「潛在風險」、「逐步成長」、「維持穩健」與「特殊背景」等五類。其中風險警戒代表學校人數少於 60 人，權重分數排序為行政區後 25%，並且學生數趨勢或社區結構趨勢為減少者；潛在風險是學校人數低於 180 人，招生率不足 75%，權重分數排序為後 50%，並且學生數趨勢或社區結構趨勢為減少者；逐步成長是該校權重分數排序為前 25%，學生數趨勢或社區結構趨勢為增加，未來新生人數較目前為高，並且生師比已超過 16 人，教育成本不高者；維持穩健則是排除在其他名單之外的學校，表示在短時間內，該校有較少的機率面臨裁校、整併或增設的考量。評估結果名單中列為「風險警戒」有四所，「潛在風險」有 14 所，「逐步成長」有三所，「維持穩健」有 198 所，其他則是「特殊背景」。因特殊背景並不列入本研究的分析範圍，故預測變數的名單只有四類，分析的學校總數為 219 所。

在塑模階段本研究所運用的模式有分類與迴歸樹（Classification and Regression Tree, CART）、類神經網路（Neural Network）、決策樹（Decision Tree）、支援向量機（Support Vector Machine, SVM）、貝氏網路（Bayesian Network）等五種，均屬於監督式學習（supervised learning）的方法，係根據輸入變數分析與預測變數之間的關聯。但五種模式的統計與演算法基礎、適用變項屬性與數量、分類規則解讀與呈現均有不同的特點與差異，以下將分別說明：

（一）分類與迴歸樹

分類與迴歸樹的目標變數可以是連續性的數值資料，也可以間斷型的類別資料（張云濤、龔玲，2007），係以反覆運算的方式，由根部開始建立二元分支

樹，直到達到樹節點的同質性達到預設標準或終止條件為止，其產生的基本演算法是貪婪演算法（greedy algorithm），針對分類與迴歸樹分割所需最小雜質改變量模型採 Gini 索引法（廖述賢、溫志皓，2011）。相較於決策樹可以產生多元的分支樹，分類與迴歸樹僅能建立二元分支樹，雖然對解讀分類規則相當便利，但也限制了分類規則的精確度。

（二）類神經網路分析

類神經網路分析可視為推算預測類的改良技術（謝邦昌（主編），2005），係模擬大腦神經細胞的運作方式，由一些高度連結的處理單元組成一動態的運算系統，透過不斷自我調整，使輸入的資訊在經過神經元運算後得到預設的輸出結果（曾憲雄、蔡秀滿、蘇東興、曾秋蓉、王慶堯，2005）。類神經網路多運用在資料變數間的關係為非線性模式，當輸入變數有交互作用效果，類神經網路亦能自動偵測，但資料分析結果模式卻有不易判讀的缺點，也因為每次原始訓練隱藏層的節點原始權值是隨機產生，所以會產生不同的分析結果，而限制了模式的精確度（謝邦昌（主編），2005；Hämäläinen & Vinni, 2011）。

（三）決策樹

決策樹是最廣為人知的一種分類方法，其以樹枝狀呈現各種輸入變數影響輸出變數的一種預測模型，並且依據不同情境的輸出變數，建立輸入變數的分類規則（謝邦昌（主編），2005）。決策樹的中間節點為測試條件，分支為條件測試結果，葉節點則是分類結果（Hämäläinen & Vinni, 2011）。最重要的目的係根據樣本所建立的決策樹，用以預測新資料樣本的分類（曾憲雄等人，2005）。決策樹所產生的分類規則具有容易理解，以及可以處理連續和間斷變項的優點。然而，決策樹對連續變項的屬性預測較困難，而當類別太多時，會有誤差亦隨之增加的缺點（張云濤、龔玲，2007）。決策樹的分類規則若太簡單則有決策的資料量不足問題，但分類規則若太複雜，則會對資料過度反應，造成無法準確預測的問題，即產生過度學習的缺點，此缺點和分類與迴歸樹是相同的。

（四）支援向量機

支援向量機是用來做分類或迴歸的方法，利用已訓練的模型預測尚未分類的資料，是屬於監督式學習的方法。目的是在高維度空間中找到一個最佳超平面（thickest hyperplane）作為二類的分割，並使分類間隔最大，其特點是能同時最小化經驗誤差與最大化幾何邊緣區，對於解決小樣本、非線性、高維度和局部極

小點有較好的解決能力（廖述賢、溫志皓，2011；Hämäläinen & Vinni, 2011）。支援向量機與類神經網路分析有相同的限制，分析結果不容易解讀，並且也不容易選擇適宜的估計參數，在運用上較不容易（Hämäläinen & Vinni, 2011）。

（五）貝氏網路

貝氏網路又稱信任網路（belief network）是一種機率圖型模型，是以條件機率為基礎，建構具有方向性非循環的有向圖（directed acyclic graph, DAG），能將特定領域中不確定性組合成模型（廖述賢、溫志皓，2011）。貝氏網路圖包含節點與連結，前者表示欲研究的變項，後者代表變項之間的相互關係。貝氏網路是以整體性的觀點調整網路，當機率值需要被調整時，所有相關節點都能根據條件機率同時被調整，即在推論的過程中，若有新的資訊時，便能立即反應並計算所有事件的發生機率（張云濤、龔玲，2007）。當分析變項較多或變項之間的相關較大時，貝氏網路分類效率的表現並不佳，但如果變項較少或相關較小時，貝氏網路的分類效率則較佳。

為比較不同的資料採礦分析模式的結果，本研究將再進行模式評估，主要分為模式內部評估與外部評估兩種方式，內部評估方式係將資料採礦分析模式預測的結果對照個案縣市委託研究計畫報告的評估結果名單，即建立分析矩陣以驗證分析模式的準確率，塑模與評估途徑如圖1。首先是分析資料的輸入，即圖1的分析資料.sav圓形圖示；接著是就分析資料進行定義，為圖1中的 Type 六邊形圖示；其次是進行模式分析，並且產生分析結果，如圖1的評估名單六邊形圖示，以及評估名單鑽石形的圖示；再次則針對評估結果進行模式評估，分別是建立分析矩陣（圖1的評估名單×$R-評估名單的四邊形圖形）、分析準確率（圖1的 Analysis 四邊形圖形），以及進行平均數檢定（圖1的 Means 四邊形圖形）。

圖1 資料採礦分析模式之塑模與評估途徑

註：分析模式以分類與迴歸樹為例。

外部評估係以教育部公布的小型學校發展評估指標分數加總做為效標分數，再以模式分析結果的名單做為分類標準，以單因子變異數分析（Analysis of variance, ANOVA）檢定各類名單的效標分數差異。此外，在指標權重建構的一致性分析，則計算肯德爾和諧係數（Kendall's coefficient of concordance），了解不同模式分析結果的一致性。

肆、研究結果與討論

研究者依序運用五種資料採礦分析模式進行實徵資料的分析，以下將分別說明分析結果，以及模式評估的結果，並進行綜合討論。

一、五種模式分析結果

本研究所運用的模式有分類與迴歸樹、類神經網路、決策樹、支援向量機、貝氏網路等五種，以下將分別說明五種模式的分析結果，以及輸入變項的重要性。

分類與迴歸樹的分析結果總共產生三個層級，輸入變項的重要性分別是學生數的 .464、社區人口數為 .168，以及其他輸入變項的重要性均為 .052。類神經網

路的分析結果則列出所有輸入變項的重要性，依序為學生數趨勢（.245）、校齡（.138）、整合比率（.133）、社區人口數（.107）、學區平均招生率（.093）、交通工具（.076）、學生數（.077）、新生人數（.072）、學校距離（.050）與社區結構趨勢（.008）。決策樹的分析結果產生四個層級，輸入變項的重要性分別是學生數的 .566、社區人口數為 .279、學區平均招生率為 .109，以及學生數趨勢的重要性為 .046，其他均低於 .001。支援向量機的分析結果可以得到所有輸入變項的重要程度，依序為整合比率（.212）、社區結構趨勢（.155）、新生人數（.129）、學生數（.121）、社區人口數（.113）、學生數趨勢（.107）、學區平均招生率（.054）、交通工具（.049）、校齡（.038）與學校距離（.022）。貝氏網路則僅顯示三個輸入變項的重要程度，分別是校齡的 .913、學生數趨勢的 .064，以及交通工具的 .023。表1顯示五種模式分析結果的各項輸入變項重要程度，以及計算輸入變項的等級平均數。進行輸入變項重要程度的一致性檢定，發現校正後的肯德爾和諧係數為 .238（因有相同等級出現，故肯德爾和諧係數必須經過校正），卡方值為 10.725（自由度為 9），並未達 .05 的顯著水準，表示五種模式所計算的 10 項輸入變項重要程度並沒有一致性。分類與迴歸樹和決策樹兩種模式均將學生數與社區人口數兩項列為最重要的輸入變項，但其他模式並非如此。新生人數與社區結構趨勢在支援向量機的重要程度較高，但在其他模式中則未列入前三項中，此種情形亦見於貝氏網路中。校齡與學生數趨勢兩項輸入變項在貝氏網路的重要程度較高，此結果僅與類神經網路較一致，但在其他模式中則未列入前三項中。

表1 各模式輸入變項之重要程度分析

指標名稱	輸入變項名稱	模式				
		分類與迴歸樹	類神經網路	決策樹	支援向量機	貝氏網路
學生數	學生數	.464(1)	.077	.566(1)	.121	.000
學生數趨勢	新生人數	.052	.072	.000	.129(3)	.000
	學生數趨勢	.052	.245(1)	.046	.107	.064(2)
社區結構	社區人口數	.168(2)	.107	.279(2)	.113	.000
	社區結構趨勢	.052	.008	.000	.155(2)	.000

註：括號內數字為重要性排序，分數愈低表愈重要，反之則否。

續下頁

表1 各模式輸入變項之重要程度分析(續)

指標名稱	輸入變項名稱	模式				
		分類與迴歸樹	類神經網路	決策樹	支援向量機	貝氏網路
距公立學校距離	學校距離	.052	.050	.000	.022	.000
與鄰近學校間有無公共交通工具	交通工具	.052	.076	.000	.049	.023(3)
校齡	校齡	.052	.138(2)	.000	.038	.913(1)
整合後之學校是否需再增建教室及充實設備	整合比率	.052	.133(3)	.000	.212(1)	.000
社區對學校之依賴度	學區平均招生率	.052	.093	.109(3)	.054	.000

註：括號內數字為重要性排序，分數愈低表愈重要，反之則否。

二、模式評估結果

針對五種模式的分析結果進行評估（如表2），首先是內部評估，由五種模式的分析矩陣可以發現正確率最高的模式為決策樹（97.26%），其次分別為分類與迴歸樹（95.43%）、支援向量機（94.52%）與貝氏網路（93.61%），最末是類神經網路（92.69%）。惟支援向量機與貝氏網路兩者能夠區分四種名單類型，而分類與迴歸樹和決策樹均未區分出逐步成長，類神經網路則是只區分出潛在風險與維持穩健兩種。在外部評估方面，五種模式的 F 值均達 .05 顯著水準，即五種模式所評估名單類型的效標分數均有差異。依據內外部評估結果，均顯示五種分析模式均能夠有效區辨不同的學校整併類型。附錄一列出五種分析模式結果與原始分類名單有差異的學校名單，以進一步了解各分析模式的準確度。值得特別注意的是學校代碼為 077 的學校，原始分類名單評估其為潛在風險，但有四種分析模式的結果均評估為維持穩健，呈現分析模式有高估的情形。另外，學校代碼 158 的學校係歸類為維持穩健，但亦有四種分析模式的結果為潛在風險，學校代碼 032 則是逐步成長，但分析模式結果為維持穩健，顯示分析模式為低估的情形。因此，應針對此三所學校再進一步釐清或列入個案討論。

表2 模式評估結果

模式	分類名單	風險警戒	潛在風險	維持穩健	逐步成長	正確率	F值
分類與迴歸樹	風險警戒(4)	3	0	1	-	95.43%	5.357*
	潛在風險(14)	0	12	2	-		
	維持穩健(198)	2	2	194	-		
	逐步成長(3)	0	0	3	-		
類神經網路	風險警戒(4)	-	1	3	-	92.69%	8.741*
	潛在風險(14)	-	5	9	-		
	維持穩健(198)	-	0	198	-		
	逐步成長(3)	-	0	3	-		
決策樹	風險警戒(4)	3	0	1	-	97.26%	6.400*
	潛在風險(14)	0	13	1	-		
	維持穩健(198)	0	1	197	-		
	逐步成長(3)	0	0	3	-		
支援向量機	風險警戒(4)	2	1	1	0	94.52%	6.993*
	潛在風險(14)	0	8	6	0		
	維持穩健(198)	1	2	195	0		
	逐步成長(3)	0	0	1	2		
貝氏網路	風險警戒(4)	3	1	0	0	93.61%	6.015*
	潛在風險(14)	0	12	2	0		
	維持穩健(198)	0	11	187	0		
	逐步成長(3)	0	0	0	3		

註：分類名單欄位後括號內的數值表示原始分類名單的次數；-代表分析結果未有此項類別。
*p< .05.

三、綜合討論

教育部或各縣市政府教育局／處較能夠掌握龐大的資料庫，有利於憑藉資料分析的結果推展學校整併的政策。學校整併的考量因素主要有學校規模、學生人數、學校所在地區、交通狀況、社區認同，以及財務狀況等（吳政達，2006；林雍智，2006；張志明，2012；Adams & Foster, 2002; Bard, Gardener, & Wieland, 2006）。然依據分析結果，五種模式所呈現的各項輸入變項重要性排序並不一致，但前述文獻探討所歸納的考量因素均有涉及。

依據分析結果，可以從三種比較觀點評估五種模式，分別是正確率、分類結果的周延性，以及輸入變項重要程度的完整性。正確率最高的模式為決策樹，

較不理想的是類神經網路，但其正確率仍然高達 92.69%。惟正確率較高的決策樹和分類與迴歸樹兩個模式的分類結果僅區分出風險警戒、潛在風險與維持穩健三種結果，缺少逐步成長的結果。因此，雖然兩種模式的正確率較高，但在分類結果的周延度上略顯不夠。而在輸入變項重要程度的分析上，此兩種模式都只有學生數與社區人口數兩項輸入變項的重要程度明顯高於其他指標，而其他輸入變項的重要程度均一致，顯然在區分輸入變項重要性上也較簡單。然而，此兩種模式能夠輸出簡明易懂分類規則的優點是其他模式較難取代的。附錄二呈現分類與迴歸樹的分析結果，由分類規則即可以清楚看出先依據學生數大於 131.5 人的學校，其歸類為維持穩健與逐步成長的機率較高，小於 131.5 人的學校則再分別依據學生數趨勢與社區人口數區分為潛在風險與風險警戒的名單。吳政達（2006）指出國民小學經營規模不經濟應為 100 人以下，此論述與本研究的分類與迴歸樹的分析結果 131.5 人有些許差異。可能的原因是本研究係針對個案縣市的資料進行分析，而各縣市的背景不一，若以其他縣市為分析對象，則會出現不同的結果，亦符合本研究提出學校整併應依不同的縣市背景為考量，不應有相同的標準。

分類結果較為周延的模式有支援向量機與貝氏網路，此兩種模式均能區分四種分類結果，但貝氏網路的分類結果在潛在風險與維持穩健兩項中有較大的落差，應屬於維持穩健的名單誤判為潛在風險的次數較高。此外，從輸入變項的重要程度來看，貝氏網路的輸入變項重要程度分析僅能明顯區分出三個輸入變項，並且其中兩項輸入變項也不在其他模式的前三項重要輸入變項中。類神經網路的分類結果僅有潛在風險與維持穩健兩種，但在輸入變項的重要程度分析上頗為完整，能夠列出每個輸入變項的重要程度。

綜上所述，若要以模式正確率與便於解讀考量，那麼決策樹和分類與迴歸樹是較佳的選擇。而若以模式結果的周延性與輸入變項重要程度的完整性為考量，則以支援向量機模式為佳。貝氏網路的分析結果較為周延，但輸入變項重要程度的分析較不理想，校齡的重要程度太高，並且應該也不是首要考量的指標。類神經網路的分析結果則是在正確率與周延性的表現較不理想，並且因為每次原始訓練隱藏層的節點原始權值是隨機產生，所以使得分析結果並不相同，使得類神經網路的分析結果較不適用本研究中。

伍、結論與建議

依據研究目的與研究結果，本研究提出下列結論與建議，首先是學校整併評估應考量的資料庫、資料採礦運用的結果，以及整併評估的執行階段等三項結論；其次，則是針對大數據方法論、資料採礦方法的運用，以及學校整併指標內容等三項建議。

一、結論

（一）學校整併應以大數據思維整合教育、人口與地理資料庫

學校整併考量因素相當多元，應有效整合不同的資料庫。本研究以學校為單位，整合教育、人口與地理三種資料庫的內容，而這些資料庫的內容均是公開便於取得且定期更新，為學校整併提供客觀的分析依據，並有利於學校本身能夠自行檢視。此舉除了符應大數據時代的到來，亦改善傳統研究方法的侷限性，為教育決策提供不同的思考方向。

（二）資料採礦能在龐雜資料中挖掘出有意義的指標項目與重要程度

當前的世界充斥龐雜的資料，運用資料採礦分析模式有利於挖掘隱藏在其中的資訊，並且形成知識與智慧。在教育領域已運用資料採礦進行研究，但運用於學校整併方面的研究仍未多見。本研究即利用此途徑並結合各種資料庫，釐清學校整併指標的操作型定義內容與權重，雖然五種資料採礦分析模式的分析結果有所不同，但都有相當程度的正確率，其中決策樹和分類與迴歸樹兩種模式更有分類規則清楚易讀的優點，而支援向量機能夠判別四種分類名單，以及列出所有輸入變項的權重，提供了最完整的資訊。

（三）學校整併應採實徵資料評估與實地訪察兩階段評估

學校整併涉及因素與影響層面甚鉅，絕非依據大數據思維運用資料採礦模型即能一蹴可幾，而是應審慎進行評估，故本研究提出實徵資料評估與實地訪察兩階段評估模式。首先，依據本研究對教育部小型學校發展評估指標的進一步定義，共分為一般性、補充性與特殊性三種指標種類。一般性指標包括學生數、學生數趨勢、社區結構、距公立學校距離、與鄰近學校間有無公共交通工具、校齡、整合後之學校是否需再增建教室及充實設備、社區對學校之依賴度等八項指標，這些指標能夠從相關資料庫取得，故運用於實徵資料評估階段。其次，補充

性指標包括小型學校大部分教室屋齡、原校區之用途，以及其他（如歷史背景）等三項指標，但這些指標數據無法從現有資料庫取得，故運用於實地訪視階段。特殊性指標包括該區只有一所小學或國中、原住民地區學校、到鄰近學校交通有重大安全顧慮（如經過土石流危險區域），以及其他等四項，因學校只要符合其中一項即大幅減少整併的可能性，故不列入整併的名單中。

茲就個案縣市代碼 017 學校為例說明如何進行兩階段評估。首先，在實徵資料評估階段為蒐集各校的實徵資料，表3呈現該校的背景資料，再依據本研究所發展的指標操作型定義進行計算，該校的指標分數為 22，在個案縣市中屬於偏低的分數。而依據五種資料採礦的分類結果，分類與迴歸樹和決策樹模式為風險警戒，類神經網路、支援向量機與貝氏網路的分類為潛在風險（如附錄一）。因此，在實徵資料評估階段中，可考量將此校列入未來可能面臨整併的重點學校名單中。其次，則由縣市政府教育主管機關組成專案小組進行實地訪察評估，評估的重點係針對該校的大部分教室屋齡、原校區用途、歷史背景、財務狀況等指標，再進行是否應整併的整體性評估。

表3 個案縣市代碼017學校評估資料

輸入變項名稱	學生數	新生人數	學生數趨勢	社區人口數	社區結構趨勢	學校距離	交通工具	校齡	整合比率	學區平均招生率
變項資料	35～40人	少於10人	-0.288	1,400～1,500人	-0.005	3公里	無	60～65年	6.333	.590
指標名稱	學生數	學生數趨勢		社區結構		距公立學校距離	與鄰近學校間有無公共交通工具	校齡	整合後之學校是否需再增建教室及充實設備	社區對學校之依賴度
指標分數	2	1		2		4	5	4	2	2

二、建議

（一）應基於大數據思維建構長期且即時性的教育資料庫

具備大數據的思維能夠在制定教育政策時，基於較為客觀的統計數據分析結

果，而避免過於主觀的偏見，故教育政策決策者與教育行政執行者都應具有大數據的思維。而各項教育政策的執行，都應有適切的指標指引，資料採礦則適用於各種龐雜的統計資料分析，故以資料採礦做為發展決策依據係可行的策略。而各種資料的蒐集工作就變得非常重要，此時教育行政機關則應發揮此功能。通常政府較有能力建置大型的資料庫，但分析資料的任務可以交給專家進行，提供行政機關決策的參考依據。大型的教育資料庫並不僅只是大，還要能夠具有即時性，並且所蒐集的資料時間點應以月或日為單位，而非現行大多以年為單位。例如，學生的轉學情形、社區人口的出生或流動狀況，就應以月或日為記錄單位，如此可以觀察更多與即時的變化趨勢，有利於學校或教育行政機關提早因應。其次，除了蒐集儲存大量的資料外，亦應建置便於即時輸入的資料庫平台，使學校或教育行政人員能夠方便輸入各種資料。大數據的思維並不是要取代專家決策，而是統計分析結果只能提供決策的佐證，如同本研究以大數據思維分析哪些學校是歸類於應裁併校的名單，但仍依據督學的實地訪察評估，才能作出更適宜的判斷。

（二）學校與縣市政府可運用資料採礦模式有效評估學校整併的需求

本研究運用的資料採礦模式有分類與迴歸樹、類神經網路、決策樹、支援向量機、貝氏網路等五種，研究結果均有相當程度的正確率。在實務運用上，本研究建議可同時運用五種模式，並比較此五種模式的分析結果，若分析結果不一致，則應再蒐集補充性指標資料，並進行實地訪察，使決策更為周延。就學校運用方面，建議可參考本研究的分類與迴歸樹分析結果路徑（附錄二）了解本身的定位，縣市政府則可依據本研究的操作型定義，並配合支援向量機或類神經網路的各項權重分數，計算所轄各校的分數，釐清可能面臨整併的重點學校。此外，運用資料採礦的研究人員應有健康的態度來對待電腦，而不應該期望機器來取代思考，即各種模式的分析結果僅是暫時的，並非最終的決策結果。

（三）學校整併作業應再考量更多元廣泛的指標項目

本研究以資料採礦的方法分析小型學校發展評估指標的重要程度，在正確率、周延性與完整性的表現上，五種資料採礦方法各有優缺點。然而，當前現行的教育資料庫應當持續蒐集多年期與多元化的統計資料，以利發揮大數據的巨量性特色，並能依據本研究所發展的分析模式改善學校整併作業。本研究係以拋磚引玉的角色提出適宜的分析途徑，做為將來進一步深入分析的踏腳石。未來可以依據本研究的分析模式，持續輸入更新的統計數據，以利模式進行更深入的機

器學習，改善模式的準確度。研究發現進行學校整併考量的重要因素包括學生人數、社區結構與學校設備，但若能再加入更多元的指標項目，相信將更貼切反應該所學校的特色。例如，記錄學生每學期學習表現資料，以建構長期性資料庫，做為學校辦學績效的指標項目，或者建構家長與社區民眾對學校的交流平臺，計算各校跨學區就讀人數，亦能夠充實社區對學校依賴度的指標內涵，以及蒐集學校的年度經費預算，計算每生單位成本做為學校財務狀況的依據。而各校亦可利用本研究對各項指標的操作型定義與權重，計算評估每一年的分數，了解學校本身的定位與改變情形。教育行政機關亦可利用本研究的定義來建置學校整併的資料庫，並且發展相關的演算法快速評估可能面臨裁併的學校，以即時介入輔導改善或協助轉型。除了有統計量化資料的客觀支持外，甚至避免不當的政治因素干擾，也能更快掌握重點學校名單，提高行政效率。

大數據的思維與資料採礦在教育決策上的應用仍有限制，諸如大數據並不長於分析因果關係，而只能呈現相關情形，容易受到雜訊的干擾而使模式產生過度學習。此外，並非所有的因素均能量化投入分析模式中，利害關係人的介入與情感因素即是典型的例子。就實務面來說，現行統計資料庫仍有待累積更長期的資料，以及蒐集以個人變項為主的資料，但卻會涉及《個人資料保護法》與是否受到政府監控的議題。再者，目前資料採礦模式的運用並不普及，並非所有的政策制定者或行政人員均熟悉這些方法，對資料的篩選與結果的解讀可能會有困擾。一項教育決策影響甚鉅，攸關國家整體的發展，實不可輕忽。大數據的思維與資料採礦的運用雖然可以提供一個參考方向，但仍不宜全然仰賴資料分析的結果，應進行通盤整體性的規劃才能形成較佳的決策。

參考文獻

中華民國監察院（2012）。**「國中小學廢併校後之閒置校舍活化成效與檢討」專案調查研究**。取自https://www.cy.gov.tw/AP_HOME/Op_Upload/eDoc/調查報告/102/1020001081010833345.pdf

牛田一雄、高井勉、木暮大輔（1996）。**資料採礦利用 Clementine 使用手冊**。臺北市：鼎茂。

吳政達（2006）。少子化趨勢下國民中小學學校經濟規模政策之研究。**教育政策論壇，9**（1），23-41。

宋吉永（2013）。**Big data：讓你看見真實欲望**。陳姿穎（譯）。臺北市：精誠資訊。

林雍智（2006）。日本實施中小學校整併的情形對我國之啟示。**教育行政與評鑑學刊，1**，135-157。

邱宏彬、許依宸（2011）。資料採礦在學生流失偵測上之應用。**資訊管理研究，11**，83-99。

城田真琴（2013）。**Big data 大數據的獲利模式**。鐘慧真與梁世英（譯）。臺北市：經濟新潮社。

胡世忠（2013）。**雲端時代的殺手級應用：Big data 海量資料分析**。臺北市：天下雜誌。

翁秉逸、謝明哲（2013，10月）。**應用資料採礦技術於中途輟學線上預警系統之實現**。TANET 2013 臺灣網際網路研討會，教育部資訊及科技教育司，臺中市。

張云濤、龔玲（2007）。**資料探勘原理與技術：Data mining、AI、Algorithm**。臺中市：五南。

張志明（2012）。實施國中小裁併校必須考量的政策思維。**臺灣教育評論月刊，1**（14），5-6。

張鈿富、林松柏（2012）。資料採礦分析 PISA 閱讀素養之影響因素。**教育政策論壇，15**（4），95-128。

曾憲雄、蔡秀滿、蘇東興、曾秋蓉、王慶堯（2005）。**資料探勘**。臺北市：旗標。

廖述賢、溫志皓（2011）。**資料探勘理論與應用：以 IBM SPSS Modeler 為範例**。新北市：博碩文化。

劉世閔（2012）。績效與平等之風雲又起：小校裁併之我見。**臺灣教育評論月刊，1**（14），1-4。

謝邦昌（主編）（2005）。**資料採礦與商業智慧：SQL server 2005**。臺北市：鼎茂。

Adams, J. E., Jr., & Foster, E. M. (2002). District size and state educational costs: Should consolidation follow school finance reform? *Journal of Education Finance, 27*(3), 833-855.

Bard, J., Gardener, C., & Wieland, R. (2006). Rural school consolidation: History, research summary, conclusions, and recommendations. *The Rural Educator, 27*(2), 40-48.

Chrysostomou, K., Chen, S. Y., & Liu, X. (2009). Investigation of users' preferences in interactive multimedia learning systems: A data mining approach. *Interactive Learning Environments, 17*(2), 151-163.

Cohen, F. (2003). Mining data to improve teaching. *Educational Leadership, 60*(8), 53-56.

Fowler, F. J., Jr. (2002). *Survey research methods* (3rd ed.). Thousand Oaks, CA: Sage.

Franklin, D., & Andrews, J. (2012). *Mega change: The world in 2050*. Hoboken, NJ: John Wiley & Sons.

Hämäläinen, W., & Vinni, M. (2011). Classifiers for educational data mining. In C. Romero, S. Ventura, M. Pechenizkiy, & R. S. Baker, (Eds.), *Handbook of educational data mining* (pp. 57-74). Boca Raton, FL: CRC.

Hung, J.-L., & Crooks, S. M. (2009). Examining online learning patterns with data mining techniques in peer-moderated and teacher-moderated courses. *Journal of Educational Computing Research, 40*(2), 183-210.

Kennedy, L., Lee, Y., Roy, van B., Reed, C. D., & Lippman, R. P. (1997). *Solving data mining problems through pattern recognition*. Upper Saddle River, NJ: Prentice-Hall.

Leacha, J., Paynea, A. A., & Chan, S. (2010). The effects of school board consolidation and financing on student performance. *Economics of Education Review, 29*, 1034-1046.

Loshin, D. (2013). *Big data analytics: From strategic planning to enterprise integration with tools, techniques, NoSQL, and graph.* Waltham, MA: Morgan Kaufmann.

Luan, J., Zhao, C.-M., & Hayek, J. C. (2009). Using a data mining approach to develop a student engagement-based institutional typology. *IR Applications, 18*. Retrieved from http://www.eric.ed.gov/contentdelivery/servlet/ERICServlet?accno=ED504332

Masunaga, H., & Lewis, T. (2011). Self-perceived dispositions that predict challenges during student teaching: A data mining analysis. *Issues in Teacher Education, 20*(1), 35-49.

Mayer-Schönberger, V, & Cukier, K. (2013). *Big data: A revolution that will transform how we live, work, and think.* New York, NY: Houghton Mifflin Harcourt.

Niemi, D., & Gitin, E. (2012, October). *Using big data to predict student dropouts: Technology affordances for research.* Paper presented at the International Association for Development of the Information Society (IADIS) International Conference on Cognition and Exploratory Learning in Digital Age (CELDA), Madrid, Spain.

Picciano, A. G. (2012). The evolution of big data and learning analytics in American higher education. *Journal of Asynchronous Learning Networks, 16*(3), 9-20.

Silver, N. (2012). *The signal and the noise: Why so many predictions fail-but some don't*. New York, NY: Penguin.

Surface, J. (2011). Assessing the impact of twenty-first century rural school consolidation. *International Journal of Educational Leadership Preparation, 6*(2), 1-13.

Vandamme, J.-P., Meskens, N., & Superby, J.-F. (2007). Predicting academic performance by data mining methods. *Education Economics, 15*(4), 405-419.

West, D. M. (2012). *Big data for education: Data mining, data analytics, and web dashboards*. Retrieved from http://www.brookings.edu/research/papers/2012/09/04-education-technology-west

Wrobel, S. L. (2010). A phenomenological study of rural school consolidation. *Journal of Research in Rural Education (Online), 25*(2), 1-19.

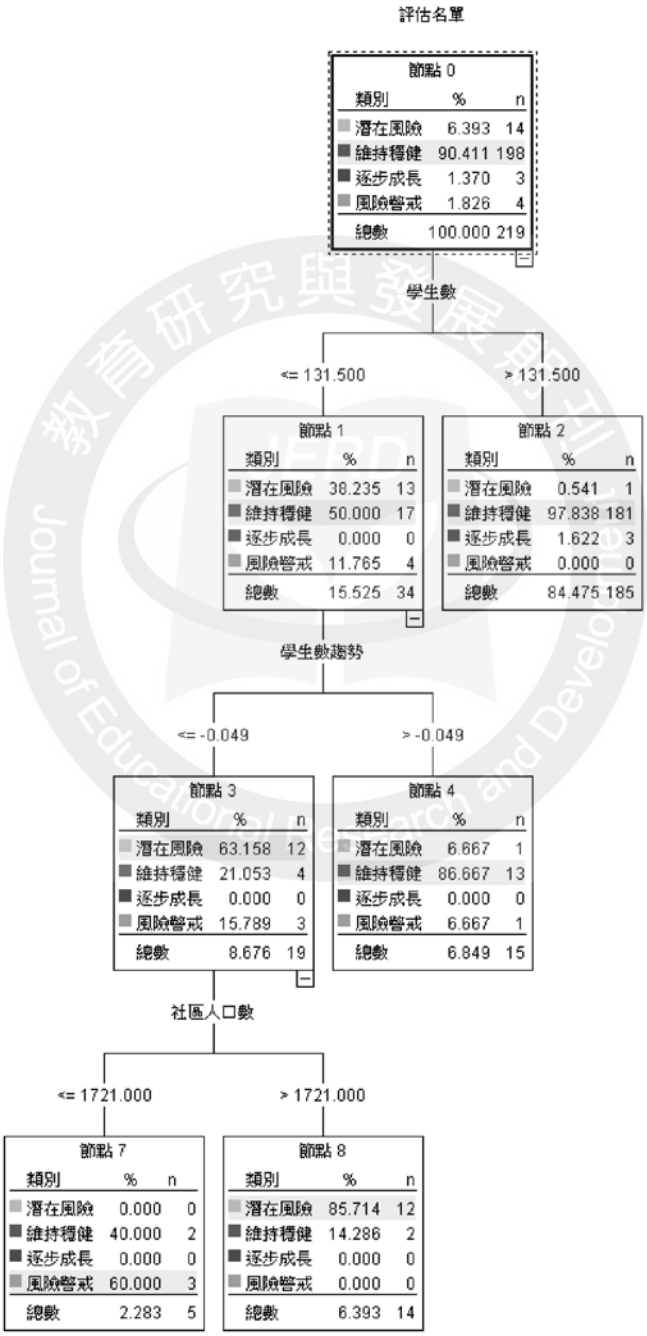
Zimmer, T., DeBoer, L., & Hirth, M. (2009). Examining economies of scale in school consolidation: Assessment of Indiana school districts. *Journal of Education Finance, 35*(2), 103-127.

附錄一資料採礦分析模式評估結果與評估名單有差異之學校名單

學校代碼	指標分數	評估名單	分析模式				
			分類與迴歸樹	類神經網路	決策樹	支援向量機	貝氏網路
017	22	風險警戒	-	潛在風險	-	潛在風險	潛在風險
048	23	風險警戒	-	維持穩健	-	-	-
224	24	風險警戒	-	維持穩健	-	維持穩健	-
226	26	風險警戒	維持穩健	維持穩健	維持穩健	-	-
047	23	潛在風險	-	維持穩健	-	-	-
050	23	潛在風險	-	維持穩健	-	維持穩健	-
090	23	潛在風險	-	維持穩健	-	維持穩健	-
165	27	潛在風險	-	維持穩健	-	-	-
207	25	潛在風險	-	維持穩健	-	-	-
077	26	潛在風險	維持穩健	維持穩健	維持穩健	維持穩健	-
172	26	潛在風險	-	維持穩健	-	維持穩健	維持穩健
056	28	潛在風險	維持穩健	維持穩健	-	維持穩健	-
169	28	潛在風險	-	維持穩健	-	維持穩健	維持穩健
044	20	維持穩健	-	-	-	風險警戒	-
178	21	維持穩健	-	-	-	潛在風險	潛在風險
158	23	維持穩健	潛在風險	-	潛在風險	潛在風險	潛在風險
054	25	維持穩健	-	-	-	-	潛在風險
094	25	維持穩健	風險警戒	-	-	-	潛在風險
187	25	維持穩健	-	-	-	-	潛在風險
206	25	維持穩健	-	-	-	-	潛在風險
112	26	維持穩健	-	-	-	-	潛在風險
202	26	維持穩健	-	-	-	-	潛在風險
036	27	維持穩健	風險警戒	-	-	-	-
117	27	維持穩健	-	-	-	-	潛在風險
148	28	維持穩健	-	-	-	-	潛在風險
031	29	維持穩健	-	-	-	-	潛在風險
204	30	維持穩健	潛在風險	-	-	-	-
084	29	逐步成長	維持穩健	維持穩健	維持穩健	-	-
001	31	逐步成長	維持穩健	維持穩健	維持穩健	-	-
032	32	逐步成長	維持穩健	維持穩健	維持穩健	維持穩健	-

註：-代表分析模式的評估結果與評估名單的分類一致。

附錄二分類與迴歸樹分析結果



教育研究與發展期刊
JERD
Journal of Educational Research and Development