

1. 資料採礦研究研究中心

聯絡方式

單位聯絡方式

- 電話：02-29387456
- 傳真：02-29361922
- 電子郵件：ting@nccu.edu.tw
- 網址：



中心簡介

● 緣起

政大商學院資料採礦研究中心於民國九十一年十月正式成立，隸屬於政大商學院，藉著院裡多位教授在資料採礦方面的專業學識，以及同仁們的努力經營耕耘下，已奠定本中心發展的基礎，成立以來，已累積相當的實務經驗，而專案內容涵蓋了學術性研究及市場行銷策略等不同領域，提供企業界最完整的資料採礦解決方案。

● 發展目標

本研究中心的發展特色如下：

- (1) 發掘企業商機：收集企業內部現有資訊及流程，確認那些資料能夠生產價值，然後投入資料採礦流程，將資料轉換為足以行動的資訊。
- (2) 提供建模諮詢：將商業問題解讀成一連串的資料採礦工作，利用資料採礦技術，從企業的龐大歷史資料中，找出隱藏的規律，建立精準的模型，提供重要資訊，做為決策時的重要參考依據。
- (3) 嚴謹教育訓練：舉辦資料採礦學術研討、專題座談及教育訓練事宜，提供政府、企業及社會各界參用。
- (4) 全方位的服務：與其他資料採礦/商業智慧顧問公司合作，結合學術研究領域與商業智慧建置規劃，提供企業界最完整的資料採礦方案服務。

● 未來發展

未來本中心會持續進行：

- (1) 對資料採礦的相關研究，藉由不斷的與外界企業單位、行政單位或是學術單位的合作，使得學術界與實務界相互合作，將資料採礦技術廣泛應用於各個方面。
- (2) 在商學院開設與資料採礦實務相關課程，讓更多學生了解資料採礦的內容，培養其興趣並建立穩固的基礎。
- (3) 舉辦研討或座談會，讓更多對資料採礦有興趣的學者可以相互交流交換意見，提升彼此的知識與經驗。同時也會致力於兩岸學術交流，兩岸可以在資料採礦上彼此交換經驗及研究成果，讓資料採礦深入各行各業。

所帶來的機遇，結合資料採礦的未來前景，我們認為至少可以在下面幾個領域裏發展及思考：
發展預測能力更强的模型

預測模型通常是企業最迫切需要的模型，因為企業常常有許多的困難等待解決，除非企業不想立足於市場。目前大多數資料採礦方法，僅僅通過對某段歷史數據的訓練，建立單一的預測模型，已經無法完成準確的預測任務。統計學中有一套比較完善的建模理論和分析方法，基於統計學和基於精確性的再訓練方法，當發現現存預測模型不再適用於當前數據時，對模型重新訓練，獲得新的權重參數，建立新的模型。傳統的統計方法過於強調假設精度，而目前無論改造過的統計方法，還是非統計的算法，由於放鬆了假定，又存在精度和信度方面的問題，統計在保證結果可靠性方面的努力還需加強。

發展更好的資料採礦模型

目前，大多資料採礦工具都和分類有關，在我們積極探索大量資料的方法的時候，是否想到數據增長的速度可能很快將超過正在研發的模型的極限，雖然以樹狀模型為基礎的分類方式，可以滿足 10 TB 或者 20TB 那麼大量的數據倉儲的詳細交易紀錄，但是這些模型是如何依賴於數據量的增加，並沒有仔細研究過。事實上，數據倉儲不可能為所有的變量設計數據立方體（cube），為有效利用有限的資源，必須進行屬性和維度的選擇。期望在基本的控制之下，使行銷和風險管理經理人員，能夠根據更大量的數據來建立起更具有預測能力的模型。

不斷成長的資料採礦標準

標準是指通過一致認同、約定俗成、或由最內行者認定而被確認為正確的任何事物、模型或準則。目前已經有一個由各相關廠商所構成的聯合組織，他們共同採用一種稱為「預測模型標記語言（Predictive Model Markup Language, PPML）」的可擴展置標語言 XML 格式來描述常見的預測模型。目標在於創造出一種模型，讓其它資料採礦，以及商業智能應用程序能夠直接採用，而不需要重新手動建立程序代碼。目前所提議使用的其它模型交換替代方案，例如微軟的 COM 或者 Object Management Group 的 CORBA 架構，都無法達到這項目標。統計早在質量控制方面就已經顯示出它在制訂行業標準方面的經驗，如何將這些思想轉移到資料採礦上面來，也是一個值得研究的課題。

面向資料庫發展的人才培養

資料採礦模型通常是由市場分析家以及統計學家，利用諸如 SAS、SPSS 或者 S+ 這類統計學相觀關軟體所產生出來的。為了產生對企業有用的解決方案，同時提高投資報酬率，這些資料採礦模型必須和企業的資訊管理系統的前後端整合在一起。但是，這些模型與算法的實施者却是由 Java 或者 C++ 程序設計員來完成。不幸的是，這兩種技術類型人員，身處彼此平行的世界之中（以不同的方式思考，使用彼此相異的語言），使得算法的開發和部署過程之間往往相當缺乏統籌與協調，投資報酬率本身是隱晦難明的。這裡也牽涉到未來的教育問題，不僅要了解統計方法的要點，更要掌握資料庫研發語言，成為集系統開發語言和統計學基本知識為一身的綜合人才。隨著資訊化水平的提高，統計學與資料採礦的應用平台漸漸統一。統計學已經具備相當完整的理論體系，它提供的是一個從了解事物到認識事物本質的科學的方法體系。我們堅信，隨著統計學與現代資訊技術的融合，藉著資料採礦對統計的強烈渴求的歷史契機，在方法的方面不斷進行新的探索，一定會為資料採礦未來的發展開闢一片新的天地。

● 中心主任：

鄭宇庭

學歷：美國明尼蘇達大學統計博士

專長：商業智慧、資料採礦、市場調查、抽樣方法

● 政大商學院資料採礦研究中心顧問陣容：

鄧家駒

學歷：美國印地安那大學博士

專長：風險評估與管理

余清祥

學歷：美國威斯康辛大學麥迪遜校區統計博士

專長：人口統計、決策理論、逐次分析

樓永堅

學歷：美國伊利諾大學香檳校區行銷博士

專長：消費者行為、流通業管理、價格及促銷策略、非營利組織行銷

別蓮蒂

學歷：美國普度大學消費者行為與零售管理博士

專長：消費者行為、消費者心理學、行銷研究、品牌管理

張愛華

學歷：美國伊利諾大學企業管理博士

專長：行銷策略、品牌管理、行銷研究、電子商務、科技行銷

余千智

學歷：美國德州大學奧斯汀校區工業工程/作業研究博士

專長：電子商務、專家決策支援系統、多媒體資訊系統

劉文卿

學歷：清華大學資訊科學博士

專長：企業整合模型、神經網路、行動商務系統、電子商務經營模式研究

蔡瑞煌

學歷：美國加州大學柏克萊分校作業研究博士

專長：類神經網路之開發、類神經網路系統在金融上的應用

研究領域

資料採礦研究

研究成果

一.學術研究與交流具體成果

- 1.2004 年 5 月在中國人民大學參加「成功企業數據挖掘暨數量化管理論壇會議」。
- 2.2006 年 4 月參加第二屆中國數據挖掘與商業智能研討會。



第二屆中國數據挖掘與商業智能研討會在中央人民大學逸夫會議中心

3.2006 年 4 月參訪天津財金大學。天津財金大學新聞報導如下:

台湾高校统计专业教师和研究生来我校访问

阅读次数：83 天津财经大学新闻 2006 年 4 月 1 日上午，台湾高校统计专业教师和研究生一行 40 余人来我校访问。校领导丁卫教授代表学校欢迎台湾师生并介绍了学校基本情况，肖红叶教授介绍了我校重中之重学科--统计学科的发展情况。台湾师生对我校的教学环境和学科发展给予了高度评价。双方增进了了解，为今后进一步扩大交流与合作开创了良好局面。

统计学系

2006 年 4 月 5 日



對方正在解釋天津財經大學電腦教室相關的情形

三.協助本院推展及提昇研究工作具體事項

1.開設商業智慧或資料採礦實務相關課程

多次在商學院開設資料採礦實務課程，讓更多學生可以學習資料採礦的技術，同時在EMBA 此課程也引起極大迴響。

2.與國立政治大學公共行政暨企業管理教育中心合作開設資料採礦學分班。

3.開設相關研討會

(1)資料採礦與創新線索研討會(2003)

主辦單位：政治大學商學院資料採礦研究中心、統計學系

贊助單位：政治大學商學院

(2)成功企業資料採礦（Data-Mining）暨數量化管理論壇

主題：資料採礦 在 金融業（金控）CRM 之案例及經驗

時間: 2003 年 2 月 28 日

四.執行成果對於社會實質貢獻

對於社會實質貢獻方面，大致可分為兩大部分:企業單位與行政單位。

企業單位:

1. 多次與 Microsoft 合作，舉辦資料採礦方法技術及企業應用培訓專班。至今已舉辦了 10 屆。時間分別為第一屆 94 年 09 月 24 日，第二屆 94 年 10 月 15 日，第三屆 94 年 11 月 19 日，第四屆 95 年 01 月 14 日，第五屆 95 年 02 月 18 日，第六屆 95 年 03 月 25 日，第七屆 95 年 04 月 22 日，第八屆 95 年 06 月 17 日，第九屆 95 年 09 月 23 日，第十屆 95 年 11 月 25 日。
2. 2003 年東方線上股份有限公司委託本中心，以資料採礦方法對 EICP 資料庫進行長期研究。
 - (1) 市場研究：針對目標產業市場，研究其市場規模、市場成長概況、市場內部競爭模式、市場特性分析以及消費者偏好之產品特性。
 - (2) 品牌研究：針對目標品牌，研究其品牌認知、品牌忠誠度、品牌成長、品牌市場佔有率以及消費者生活形態與人格特質。
3. 台灣人壽保單貸款的相關研究。
4. 彰化銀行的 Processing Data Mining with Clememtime。
5. 2005 年 EPSON Campaign 績效之利潤分析研究。
6. 第一銀行的 Customer Segmentation。
7. 2003 年臺灣智慧科技與應用統計學會智慧科技研究方法研習營。為了推動國內智慧科技研究方法經驗交流，協助各大專院校教師及企業人員從事智慧科技研究，並提昇

其實證研究能力與品質。

8. JCIC 資訊部 DW-DM 研習班教育訓練課程，財團法人金融聯合徵信中心委託。
9. 商業周刊臺灣企業人力資源大調查研究報告，商業週刊委託計畫。2005 年國泰世華銀行信用卡信用風險模型開發。合作專案預期將可產出信用卡申請客戶的信用風險評估模型（又稱信用評分模型；Credit Scoring Model）。所開發出來的模型如果能有效地篩選出償債能力不佳的個人信用卡客戶，就可以依此開發出一套信用風險系統，做出正確的核卡決策。
10. 和泰汽車顧客再購行為模式之預測。

政府行政單位:

1. 九十二、九十三、九十四年度鄉鎮市區別民眾對治安滿意度研究，行政院內政部警政署委託計畫。
2. 中小企業管理人才培訓需求研究，經濟部中小企業處北區中小企業研訓中心委託計畫。
3. 疾病管制局衛生教育宣導評估，台大公共衛生學院委託計畫。
4. 民眾使用網際網路狀況研究分析，行政院交通部委託計畫。
5. 台灣地區民眾衛教宣導後民眾對抗生素正確使用之認知、態度及行為研究報告，台大公共衛生學院委託計畫。

其 他

過去九年來，本中心在學術研究的具體成果可分為三方面：

壹、參加學術研討會，共發表論文 32 篇。

1. 鄭宇庭(2009), “企業信用風險模型”，2009 年資料採礦的未來展望研討會，2009 年 2 月 13 日，致理技術學院。
2. 鄭宇庭，徐慧玲(2008), “應用資料採礦技術建置符合新巴賽爾協定之企業信用風險模型”，2008 年海峽兩岸應用學術研討會，2008 年 11 月 9-15 日，首都經濟貿易大學。
3. 鄭宇庭(2008), “中小企業新巴賽爾協定之信用評等模型研究”，第二屆中國統計學年會，2008 年 10 月 24-26 日，浙江工商大學。
4. 謝邦昌，朱建平，鄭宇庭，董春，劉致琪(2008), “應用數據挖掘技術於多個數據庫連接與整合”，2008 國際應用統計學術研討會，2008 年 8 月 14-18 日，山東省應用統計學會，山東煙台。
5. 謝邦昌，朱建平，鄭宇庭，董春，孫莓婷(2008), “數據挖掘中的模型選取”，2008 國際應用統計學術研討會，2008 年 8 月 14-18 日，山東省應用統計學會，山東煙台。

6. 鄭宇庭(2008), “導入資料採礦技術於新巴塞爾協定下企業信用模型”, 第五屆中國數據挖掘與商業智能研討會暨海峽兩岸應用統計研討會, 2008年3月7-8日, 北京人民大學。
7. 鄭宇庭(2008), “金融數據挖掘和商業數據挖掘”, 海峽兩岸金融企業數據挖掘暨應用統計分析研討會, 2008年3月14-15日, 廈門大學。
8. 林伯修, 謝邦昌, 朱建平, 董春, 鄭宇庭(2007), “應用數據挖掘技術於銀行客戶分析之研究”, 2007統計教育國際高級論壇, 2007年8月6-9日, 內蒙古財經學院。
9. 陳佳樟, 鄭宇庭(2007), “資料採礦用於汽車客戶再購模型建置”, 第十六屆南區統計研討會, 2007年6月22、23日, 國立高雄大學。
10. 謝邦昌, 鄭宇庭(2007), “應用資料採礦技術於多個資料庫連結與整合”, 第十六屆南區統計研討會, 2007年6月22、23日, 國立高雄大學。
11. 鄭宇庭(2007), “DATA MINING 用於汽車客戶再購模型建置”, 2007資料探勘研討會, 2007年4月30日, 亞東技術學院。
12. 鄭宇庭(2007), “汽車客戶再購模型建置”, 第四屆中國數據挖掘與商業智能研討會, 2007年4月27日, 北京人民大學。
13. 謝邦昌, 鄭宇庭(2006), “統計學在數據挖掘與商業智能上之應用”, 2006海峽兩岸應用統計學術研討會, 2006年11月17-18日, 成都西南財經大學。
14. 鄭宇庭(2006), “兩岸統計平台之建立”, 海峽兩岸農村、農業暨地區統計問題研討會, 2006年11月2-5日, 湖南岳陽。
15. 鄭宇庭(2006), “智慧財產權刑事案件在台灣法院之審理”, 第一屆中國統計學年會, 2006年10月20-22日, 天津財經大學。
16. 鄭宇庭(2006), “The practice on the application of data mining”, 第五屆海峽兩岸統計與機率學術研討會暨2006中華機率統計學會年會及學術研討會, 2006年7月29、30日, 苗栗。
17. 鄭宇庭(2006), “金融數據挖掘和商業數據挖掘”, 2006中國數據挖掘與商業智能研討會, 2006年4月2、3日, 北京人民大學。
18. 楊惠如, 鄭宇庭, 謝邦昌(2005), “Data Systematic Purifying Analysis in Data Mining”, 第十四屆南區統計研討會暨、九十四年度中華資料採礦協會年會暨中華機率統計學會年會及學術研討會, 2005年6月25、26日, 國立成功大學。
19. 孫莓婷, 鄭宇庭, 謝邦昌(2005), “Model Selection in Data Mining”, 第十四屆南區統計研討會暨、九十四年度中華資料採礦協會年會暨中華機率統計學會年會及學術研討會, 2005年6月25、26日, 國立成功大學。
20. 劉致琪, 鄭宇庭, 謝邦昌(2005), “Several Databases are Linked and Combined Using

Data Mining Techniques”，第十四屆南區統計研討會暨、九十四年度中華資料採礦協會年會暨中華機率統計學會年會及學術研討會，2005年6月25、26日，國立成功大學。

21. 連婉君，鄭宇庭，蔣以仁(2004)， “上呼吸道疾病的藥物交互作用”，第十三屆南區統計研討會暨九十三年度中華資料採礦協會年會及學術研討會，2004年6月24、25日，台北亞太會館。
22. 蘇芷凡，鄭宇庭，蔣以仁(2004)， “心血管疾病之交互作用與用藥分析”，第十三屆南區統計研討會暨九十三年度中華資料採礦協會年會及學術研討會，2004年6月24、25日，台北亞太會館。
23. 廖彗嵐，鄭宇庭，蔣以仁(2004)， “糖尿病之藥物交互作用與用藥分析”，第十三屆南區統計研討會暨九十三年度中華資料採礦協會年會及學術研討會，2004年6月24、25日，台北亞太會館。
24. 黃雅芳，鄭宇庭，謝邦昌(2004)， “Imputation of value-added database in data mining”，第十三屆南區統計研討會暨九十三年度中華資料採礦協會年會及學術研討會，2004年6月24、25日，台北亞太會館。
25. 陳惠雯，鄭宇庭，謝邦昌(2004)， “Sampling methods of value-added database in data mining”，第十三屆南區統計研討會暨九十三年度中華資料採礦協會年會及學術研討會，2004年6月24、25日，台北亞太會館。
26. 包寶茹，鄭宇庭，謝邦昌(2004)， “Error index and model criteria for value-added database in data mining”，第十三屆南區統計研討會暨九十三年度中華資料採礦協會年會及學術研討會，2004年6月24、25日，台北亞太會館。
27. 林建言，鄭宇庭，謝邦昌(2004)， “Functional mapping of value-added database in data mining”，第十三屆南區統計研討會暨九十三年度中華資料採礦協會年會及學術研討會，2004年6月24、25日，台北亞太會館。
28. 謝晴仔，李珮榕，邱蔚群，李家旭，鄭宇庭(2003)， “Data Mining 在保險業之應用”，第十二屆南區統計研討會暨九十二年度中華機率統計學會年會及學術研討會，2003年6月26、27日，國立高雄大學。
29. 鄭宇庭，謝晴仔，李珮榕，邱蔚群，李家旭，何玉芝，林建言 (2002)， “資料探勘技術應用於企業資料庫之資料探索與決策”，九十一年統計學術研討會，2002年12月1日，台中東海大學。
30. 鄭宇庭，李珮榕，邱蔚群，李家旭 (2002)， “壽險公司附加保單銷售之目標消費群研究”，九十一年統計學術研討會，2002年12月1日，台中東海大學。
31. 鄭宇庭，謝晴仔，何玉芝，林建言 (2002)， “不同信用卡申請方式的持卡者異同研

究”，九十一年統計學術研討會，2002年12月1日，台中東海大學。

32. 鄭宇庭 (2001), “Data-Mining 在市場研究之應用”，第一屆海峽兩岸暨港澳地區中國企業管理案例學術研討會，2001年5月25、26日，大連。

貳、發表期刊論文，共 8 篇。

1. 鄭宇庭，夏學理(2008),“資料採礦流程在顧客關係管理中之應用”，*Journal of Data Analysis*, Vol.3, No. 1, . 167-174.
2. 鄒幼涵，黃登源，鄭宇庭(2007),“資料採礦之簡易系統—以流行病學為例”，*Journal of Data Analysis*, Vol.2, No. 3, . 139-170.
3. Yeh, R. L., C. Liu, B. C. Shia, Y. T. Cheng, Y. F. Hwang(2007), “Imputing Manufacturing Material in Data Mining”, *Journal of Intelligent Manufacturing*, Vol. 19, No. 1, 109-118. (SCI)
4. Rahman M.，朱建平，鄭宇庭 (2006), “預測誤差及變異效應之估計與分類”，*Journal of Data Analysis*, Vol.1, No. 3, . 113-129.
5. 郭良芬，鄭宇庭，謝邦昌(2006), “商業智慧分析在保險業之應用”，*核保學報*，No. 14, 275-300.
6. 鄭宇庭 (2006), “料採礦雛型系統”，*Journal of Data Analysis*, Vol.1, No. 1, 1-26.
7. 鄭宇庭，蘇志雄 (2002), “商業智慧的工具—資料採礦”，*輔仁管理評論*，Vol. 9, No. 2, 11-34.
8. 鄭宇庭，羅家蓉，陳怡伶 (2001), “資料採礦(Data Mining)雛形系統”，*統計學報*, Vol. 24, 50-67.

參、撰寫專書，共 4 本。

1. 韋端、鄭宇庭、鄧家駒、匡宏波、謝邦昌編著(2003)，*DATA MINING概述 – 以Clementine 7.0為例*，中華資料採礦協會出版。
2. 鄭宇庭、謝邦昌、易丹輝編著(2006)，*統計資料分析 – 以STATISTICA 7.0為例*，曉園出版社。
3. 謝邦昌、鄭宇庭、蘇志雄、郭良芬編著(2008)，*EXCEL在資料採礦之應用*，中華資料採礦協會出版。
4. 謝邦昌、鄭宇庭、蘇志雄、郭良芬編著(2008)，*Microsoft SQL Server在資料採礦與商業智慧之應用*，中華資料採礦協會出版。