

認識資料探勘

◎ 統卒

資料探勘(Data Mining)是統計領域中充滿挑戰的新學門，在電子記錄龐大存檔資料中採掘有用訊息，取代不少傳統資訊蒐集方式。

老師利用機率甚低的複選題相同答錯率頻次，發現某些隔鄰同學有群聚作弊傾向，進而調整座位，讓公平與誠實建立於無形的環境中，但沒有一位學生知道。

警大一位教授在汽機車竊案之研究中，對某些員警高破案率好奇，正擬面訪其技巧與經驗，俾編教戰手冊前夕，無意間發現：查獲案車牌經常重複。遂以報竊案件之車牌號重整資料庫，不少老舊機車總是失而復得，最多者八年達百次以上。少數員警藉親友跨縣市或跨轄區報竊以交叉製造破案績效之自導自演弊端始告結束。

超市或便利商店早在十年前即已利用電腦盤存控制舖貨種類，但選擇每日最佳上架時點及擺設位置，以提升坪效，則是近年利用 Data Mining 的結果。

台積電以產程電腦記錄查核不良品之產出機組、機台、時段、溫溼度等細微過程進行交叉分析，提升良率，行之已久。

健保局利用每年上億筆之醫病、給付紀錄，分析國民健康趨勢、調整醫療資源分配及改善財務，亦見初效。

Data Mining 海闊天空，在企業界已被廣泛應用。公部門稍嫌落後，除健保檔外，財稅檔、社會保險檔、金融信用檔及社會福利檔等都是大礦場，有待探勘及開發。