

## 6. 結論與未來研究方向

資料探勘對每個公司來說是一種重要的策略性的的計畫，而將之列為高度機密，所以要調查各家公司到底用資料探勘來做什麼樣的事其實相當不容易。根據 Two Crows Corp. 最近的調查顯示，資料探勘主要的三個應用方式，如我們所預期的都在市場推廣方面，分別是：Customer Profiling、Targeted Marketing（客戶關係的管理）、以及 Market-Basket Analysis。在 Customer Profiling 方面，我們希望找出客戶的一些共同的特徵，希望能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。行銷人員就可以只針對這些名單寄發廣告資料，以降低成本，也提高行銷的成功率。Market-Basket Analysis 主要是用來幫助零售業者瞭解客戶的消費行為。客戶關係的管理則是可以由一些原本的客戶，後來卻轉而成為競爭對手的客戶群中，分析其特徵，再根據這些特徵到現有客戶資料中找出有可能轉向的客戶，然後公司必須設計一些方法將他們留住，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。

所以在此篇論文中主要是利用資料探勘提供兩種不同方法的關聯法則，無論在布林值的關聯法則或複合維度的關聯法則上均能準確預測出兩個或兩個以上的特徵因子區間。在複合維度關聯法則時建立多個關聯法則，不但可以提供購屋者有更多的選擇外，更能提供房地產業者可以針對消費族群擬定好的企畫案，吸引更多潛在消費族群。但在決定關聯法則之前的準備工作較為繁瑣，除了樣本的特性外，若特徵因子區間分割過於繁多，反而會影響關聯法則建立。要如何將特徵因子區間做一個妥善的分割，進而得到有效度及信度的關聯法則，這個勢必成為一個非常重要的問題。

而在關聯法則的運用方面，除了銀行間的大樣本客戶資料及超級市場中販賣物品的擺設外，應該要加以延伸擴大，不應只居限於金融及產品擺設方面。在公司中各個部門的工作分配、工廠的生產流程應如何配置流程方能得到最好的效益，這些都可以使用關聯法則來得知最好的效益。而在關聯法則的誤差方面，由於在建立關聯法則時，在分割特徵因子區間及訂定最小支持度及最小信賴度均會影響到最終的關聯法則，所以如何將關聯法則的誤差最低也是接下來得面對的問題。而未來的研究方向應該不能只單純探討單一區間的關聯法則必須加以延伸擴大，多個區間的關聯法則，若單是使用線性迴歸模式來解決，非但無法解決問題，就算得到了完整的迴歸模型也不適用。

但若要建構出有效度的關聯法則，在區間大小上必須作更為準確的分析。還有在樣本特性方面，若能收集擁有同一特性的樣本將會求得較為有效度的關聯法則外，若能在不同特性的樣本中依然可以循著規則找出有效的關聯法則，將是在資料探勘的一大突破。除此之外，在資料是否因為空間、時間的因素有著一定的週期性，而週期性是否也影響著關聯法則的建立，這也是需要去斟酌衡量的。而雖然目前有許多配套的軟體可解

決這一類的問題，而且可以迅速的將關聯法則建立，如 SAS 公司的 Enterprise Miner、IBM 公司的 Intelligent Miner 和 SPSS 公司的 Clementine，但若上述的問題無法解決的話，所得出的關聯法則也將會流於形式，就如同線性迴歸模式一樣。

未來加強的研究主題包括上述所提之各種限制，在文獻上也不是沒有提出解決方案，例如，Han 與 Fu（1995）就曾提出一次發掘多商品項目之間關聯的作法，也曾考慮到在此過程中利用一些特性來提升效率，Ozden（1998）也對在週期發掘中效率問題提出三種解決方案。在 Han, Dong 與 Yin（1999）的研究中也提出演算法來一次發掘所有週期（而不需要使用者指定參數）及考慮非嚴格週期的問題。然而這些研究均有待進一步結合在一起才能有效地解決目前時間、空間、商品種類等維度同時考慮的問題，所以未來的工作來應該是除了致力於做有效的分類和如何將不同特性的樣本歸納出合理的關聯法則外，還要多方面探討資料採礦中商品的週期性。因為房地產方面除了距離因素外，仍然夾帶著空間時間的變換是否影響房地產的地價。所以唯有打破不同特性樣本及空間、時間的因素才能得知更多潛在的特徵及更多有效的關聯法則。而唯有得知更多潛在的特徵才能更準確的提供人類更多更好的資訊。