

應用序列樣式探勘技術建構一個時間間隔序列最佳化分類模型

研究生：簡君如

指導教授：蔡介元教授

元智大學 工業工程與管理研究所

摘要

在現今競爭激烈與環境變化快速的商業環境中，企業若能確實掌握顧客的消費行為模式，將會比其他競爭者更有立足與生存的機會。因此，企業管理者必須全盤了解顧客的消費行為，並在適當的時間提供其適當的服務。然而，企業內部的顧客資料庫通常龐大且紊亂，使得顧客行為分析的工作困難度大幅提高。資料探勘技術中廣為人知的「序列模式探勘」與「序列分類」技術將可解決上述問題。「序列模式探勘」技術因可偵測出資料庫中較常發生的序列模式，而「序列分類」技術藉由其所建立之序列模型，可預測一新的序列最有可能所屬的類別，故常被應用於許多實例當中。在以往的研究中，討論有關時間序列的分類方法非常少，但是，涵蓋時間訊息之序列確實能提供給管理者較一般序列更多且更有效的訊息。因為，在兩條內含項目集合皆相同的序列中，若每項項目之間的時間間隔長度不同，有可能為兩個不同類型的序列，若沒有考慮時間間隔的問題，將會做出此兩條序列為相同序列的判斷。

為此，本研究針對時間間隔序列(time-interval sequences) 提出了一個兩階段的分類模型，幫助管理者做出更有效的決策以滿足客戶需求。第一階段採用I-PrefixSpan演算法，從資料庫中挖掘出頻繁的時間間隔序列模式，將較常出現的客戶行為模式找出；第二階段為時間間隔序列的分類方法，包含序列相似性測量、分類模型之建立，以及模型最佳化三個部分。本研究並應用一個簡易的案例及一個大型資料庫-北風資料庫測試所提出的時間間隔分類方法，實驗結果證明本研究所提出的時間間隔分類方法是為可行且有效率的。

關鍵詞：時間間隔序列樣式探勘 時間間隔序列分類 序列相似度量測 粒子群最佳化演算法