

應用基因演算法及權重項目法於關聯法則挖掘之研究

研究生：龔書賢

指導教授：蔡介元 博士

元智大學工業工程與管理研究所

摘 要

關聯規則演算法是資料探勘 (Data Mining) 中一項相當重要且實用的技術。過去的文獻中，關聯規則大多應用於購物籃分析上，用以找出不同項目間的關聯性。故若能有效地利用關聯規則的技術，便能找出大量資料中所隱含的資訊，來提供決策者更好的決策依據及提高企業的競爭力。然而，大部分的關聯規則應用上，通常僅考慮不同項目組合所產生的關聯性，並未充分的利用資料庫中相關的項目屬性，來提昇挖掘出之關聯規則的品質。此外，管理者總要花費許多的時間來決定挖掘過程中，萃取出關聯規則所需的最小支持度與最小信心度門檻值的設定方式，故如何快速且客觀地設定最小門檻值來得到有意義的規則便相當重要。

本論文中將提出一挖掘關聯規則的新演算法，其透過暫存支持度與資料叢集索引的概念來挖掘二元型態交易資料的關聯規則。同時藉由權重項目法的概念來反映不同項目間的重要性，並將之應用在啟發式方法—基因演算法中適應度函數與目標規則函數上。透過基因演算法演算速度快且屬於廣域搜尋的特性，及目標規則函數與權重值為考量，來評估不同關聯規則間的價值，進而產生一客觀的最小門檻值設定方式建議，以期能快速地挖掘出重要且有意義的關聯規則，供企業管理者進行決策分析。

為了驗證本研究所提出之方法為一有效之關聯規則挖掘法，本研究透過數個範例資料，以及一實際的信用卡消費案例資料之應用分析，探討在實際應用上對關聯規則產生結果的影響。由此些應用中可知，將權重項目法應用在基因演算法之適應度函數與目標規則函數的方式，確實能有效的評估規則間的價值與重要性，且藉由基因演化應用確實能快速且客觀地提供一合適的最小門檻值設定建議，來提昇挖掘關聯規則的品質與效率。因此，由上述應用結果可證明本研究所提出的方法，應用在挖掘關聯規則上確實具有其實用性。

關鍵字：關聯規則、權重項目法、基因演算法、暫存支持度、資料叢集