

變數精準粗略集之理論與應用

研究生：許志華

指導教授：蘇朝墩 博士

國立交通大學

工業工程與管理學系

摘要

變數精確粗略集理論是資料探勘的重要工具之一，已廣泛應用於不同領域的知識獲取。然而，變數精確粗略集理論卻無法應用於資料含有連續型屬性的分類問題，它需要一個將能將屬性離散化的方法來進行資料的前置處理。此外，變數精確粗略集理論缺乏一個適當的方法來決定精確參數(β)值以確定其最簡化屬性集合(β -reducts)。本論文提出一個稱為「擴充的 Chi2」的新演算法，此演算法以 Chi2 演算法為基礎來發展，並改善了 Chi2 演算法無法由訓練樣本決定預先定義的錯誤分類率(δ)的問題。本論文也提出一個根據精確參數來選擇最簡化集合的方法，這方法首先利用資料錯誤率的最小上界來決定精確參數值，並利用所獲得的精確參數值來尋找資訊系統的子集合；接著計算每一個子集合的分類品質並利用分類品質的量測移除子集合中多餘的屬性，而刪除多餘屬性的子集合即 β 的最簡化屬性集合。

本論文利用決策樹軟體 See 5 分析五筆數值資料。分析結果顯示所提出的「擴充的 Chi2」演算法之績效優於 Chi2 演算法。論文中也利用一個簡單的範例說明所提出的最簡化屬性集合選擇方法如何實施，並分析一個實際的醫學案例將實驗結果和類神經網路進行比較，實驗結果顯示本論文所提出的方法有較好的績效。

最後，一個通訊產業的應用案例被分析，利用本論文所修正的變數精確粗略集理論來刪減行動電話製造程序中多餘的無線電頻率測試項目。實驗結果顯示，無線電頻率測試項目顯著的減少，而利用這些剩餘的測試項目進行後續分析，結果顯示產品檢驗的準確率非常接近原先未進行測試項目刪減前的測試程序；此外，與決策樹相比較，變數精確粗略集理論也有較好的績效。

關鍵詞：資料探勘，粗略集理論，最簡化屬性集合，離散化，Chi2 演算法

