

解決序列分類問題的複合式方法

研究生：陳致融

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

近年來，複合式的序列分類方法已受到許多學者的關注。在複合式序列分類方法中，以序列樣式探勘為基礎的序列分類方法被認為是最有效解決複雜序列分類問題。雖然先前研究已經證明序列樣式探勘基礎之序列分類方法的優點，但仍然存在許多問題，例如產生冗餘的序列樣式、不適當的序列樣式相似性的衡量及難以分類的序列問題。因此，本研究提出一種有效的兩階段序列樣式探勘之序列分類方法來解決上述問題。第一階段，在序列樣式探勘過程中，冗餘的序列樣式被刪除，而產生簡潔的序列樣式，並被視為第二階段的代表特徵樣式。第二階段中，一種序列相似性的衡量方法被提出來用於計算序列和樣式之間的局部相似性。最後，發展粒子群最佳化 AdaBoost (PSO-AB) 序列分類器，能夠有效提升序列分類的準確性。PSO-AB 序列分類器中，AdaBoost 方法能「聚焦於」那些較難分類的序列樣式上，並自適應調結改變較難分類的序列樣式的權重分佈，將前一個分類器分錯的樣式用來訓練下一個分類器。粒子群最佳化演算法能優化分類器中的樣式與類別權重，以提升分類的準確性。另外，為了解決行動商務序列分類問題，本研究將發展一套 LIT closed 序列樣式探勘之序列分類方法，將涵蓋行動商務序列之三種資訊，如顧客參觀的商店 (Location)、購買的商品(item)及拜訪的時間(Time)。LIT closed 序列樣式探勘方法能夠有效的刪除冗餘的 LIT 序列樣式，並挖掘出顧客的 LIT closed 行為樣式。這些 LIT closed 行為樣式被視為主要特徵，並用來建構分類器。

關鍵詞：序列分類、封閉探勘演算法、粒子群最佳化演算法、自適性調升法