

Design An Efficient Incremental Data Mining Algorithm Based on Pre_large Technique

游坤明, 林仕奇

資訊工程學系

資訊學院

yu@chu.edu.tw

摘要

本研究主要針對關聯式法則的漸進式探勘演算法進行研究，關聯式法則的漸進式探勘演算法即是希望能夠藉由保留某些上一次探勘後的相關資訊，能夠不需要再度處理舊有資料便能夠維持關聯式法則的正確性與有效性。本研究運用尋訪速度快的二元樹結構，以及用以降低項目在高頻與非高頻之間變動機率的準高頻(Pre_large)觀念，在盡量減少掃描資料庫的次數情況下，提出一個有效率的、可靠的漸進式探勘演算法-Pre_large Descending frequent pattern Binary-tree Algorithm(PDBA)，PDBA不但可以節省相當的探勘時間成本，並可以維持關聯式法則的有效性。除了提出PDBA演算法外，在研究中並比較PDBA演算法與前人所提出的漸進式探勘演算法(DFPBT、AFPIM)，經由一個實例的實驗結果發現，在此實例中PDBA演算法的表現皆比DFPBT及AFPIM為優，可以節省至少23%的探勘時間。

關鍵字：資料探勘、關聯式法則、樹狀結構、漸進式、準高頻