

一個高效能多核心平行資料探勘演算法

張大鈞, 游坤明, 周嘉奕

資訊工程學系

資訊學院

yu@chu.edu.tw

摘要

近年來由於資訊技術的進步，企業產生的資料量也隨之快速成長，因此利用資料探勘技術在大量的資料中找尋有意義的資訊成為了一項重要的議題。資料探勘最常使用的方法是關聯式法則，其中以 Apriori 演算法為代表，但 Apriori 在探勘高頻項目集合時會產生大量的候選項目集合並需要多次掃描資料庫驗證高頻與否。因此本論文在多核心處理器架構上提出 MultiCore-Apriori (MCA) 演算法以加速探勘效率。再者 MCA 以每一階高頻項目為單位做資料分配可以有效平衡資料負載。本論文使用 OpenMP 函式庫實作 MCA 於多核處理器系統，並使用 IBM synthetic data 驗證 MCA 的效率。實驗結果中可以觀察到，雙核心的處理器最高有 1.9 倍的效能，四核心最高有 3.8 倍的效能。

關鍵字：