

基於降冪高頻項目二元樹之快速資料探勘演算法

游坤明, 莊文宗, 江啟宏

資訊工程學系

資訊學院

yu@chu.edu.tw

摘要

資料探勘是一種從大量資料庫中萃取有用資訊、知識的技術，資料探勘技術應用的範圍相當廣泛，挖掘出資料之間的關聯式法則就是其中的一種應用。當所累積下來的資料不是很龐大的時候，資料探勘可以很有效率的探勘出資料間的規則，但隨著企業所累積的資料量越來越龐大，資料探勘的工作也就顯得非常耗費時間，因此如何有效率完成資料探勘便成為很重要的研究課題。有鑑於以往關聯式法則相關研究所提出的探勘方式需要相當多的執行時間，本論文提出一種有效率的資料探勘方式演算法-降冪頻繁樣式二元樹演算法(Descending Frequent Patterns Binary Tree Algorithm)，它僅需掃描資料庫二次即可建構出完整的高頻項目樹狀搜尋結構，而且使用此樹狀結構與維持正確的項目集合次數所用的合併左子樹(Merge Left Child Tree)方式，不但可以很快速地探勘出關聯式法則，並且可以大量地節省探勘過所需的記憶體使用量。在實驗中，本論文也利用實際的資料針對本演算法進行效能分析，實際驗證演算法的效能。

關鍵字：資料探勘、關聯式法則、樹狀結構、降冪頻繁樣式二元樹