

資料挖掘網路服務系統之探討

研究生：蔡旻宏

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

知識經濟時代的來臨使得企業界體認到知識累積與創新的重要，紛紛開始導入知識管理以提升營運效率與競爭優勢。近年來蓬勃發展的資料挖掘(Data Mining)相關技術，具有對大量資料分析處理的能力，因此成為企業在推動知識管理時的得力工具。然而企業在實際使用現有的資料挖掘軟體與服務時，經常面臨到以下的問題：(1)企業各部門間異質系統的整合與資料傳輸的相容性問題；(2)傳統資料挖掘軟體缺乏因應使用者的需求而進行流程彈性調整的能力；(3)軟體或服務的整合及維護成本過高。

為了解決以上所述之問題，本研究發展出一個以網路服務(Web Service)為基礎的「動態流程資料挖掘網路服務系統」(Dynamic Data Mining Web Service System; DDMWS System)架構，目的在於降低企業在導入資料挖掘技術時所需要的門檻。在網路服務的架構下，資料傳輸的格式均以具有開放性與結構化等優點的XML(可延伸性標示語言，Extensible Markup Language)格式文件傳遞，克服傳統資料挖掘流程中可能遭遇到的系統整合與相容性問題。而在本研究的架構之下，資料挖掘流程是由多個功能各異的網路服務之活動元件(Activity Component)所構成，得根據使用者的需求調整其組成順序；此外我們以BPEL4WS(網路服務業務流程執行語言，Business Process Execution Language for Web Service)作為描述流程相關細節的工具，提供一個系統化的方式記錄流程的各方面細節，以作為

後續修改或建立類似流程的參考。流程最終的資料挖掘模式與結果，在本系統中是以PMML(預測模型標記語言，Predictive Model Markup Language)描述之，便於使用者日後於其他軟體重新建置原有資料挖掘模式。

為了驗證本論文所提出之系統為一個合理且可行的資料挖掘系統概念，本研究透過一個簡單的資料挖掘流程，以及一個現實的信用卡資料庫作為案例，以探討本系統的效果。由這些應用中我們可以得到以網路服務為架構的資料挖掘流程可以簡單地降低資料挖掘的成本，並且符合企業對於資料挖掘的需要。

關鍵字：資料挖掘(Data Mining)、可延伸性標示語言(XML)、網路服務(Web Service)、網路服務業務流程執行語言(BPEL4WS)、預測模型標記語言(PMML)