工業技術研究院
Industrial Technology
Research Institute

視覺式資料探勘實作文本爬文分析與統計實作

■ 課程簡介

台灣企業或政府組織開始對於大數據採掘方面相當重視，希望藉由該相關之工具及資訊技術提升企業組織之獲利、降低投資風險。面對資料型態多元，如公司內部表格式結構資料庫資料；一般描述性工作日誌及服務紀錄等文本資料分析；訪談聲音之紀錄；圖片資料等資料分析亦可經處理後進行後續分析。資料科學的分析是藉由資料統計及探勘技術為基礎，**主要目的就是在大數據庫中找出有價值的隱含情報資訊，透過中文文字探勘及資料分析技術來分析及歸納出有用的結構資訊，作為決策的依據。**

■ 課程特色

主要透過大數據之分析，提供眾多資料採掘分析演算法，包含羅吉斯回歸、SVM、貝氏網路、決策數、類神經、群集、模式測試、自然語意分析來做大數據採掘。本課程將運用 PolyAnalyst 視覺化資料採礦工具，協助利用現有的資料庫及各類資料來源，透過分析在大量資料中有效地找出隱藏其內的行為模式或彼此的關聯性，建立預測未來模型，以達到快速掌握趨勢，提升整體競爭力。另外，透過分析能讓大量的資料作更有效率的運用，淬煉出有用資訊和獲得其潛在的情報及知識，如此才能提供更充足的決策情報，課程案例針對業界需求設計實作，提昇企業競爭力。

工業技術研究院
Industrial Technology Research Institute

■ 課程大綱

第一天：

課程大綱	課程內容
前言	■ 大數據分析現況 ■ 大數據的分析優勢 ■ 何謂小數據分析? ■ 大數據與小數據差異
大數據之分析工具介紹	■ 大數據之分析工具簡介
大數據資料分析所遇到的困難	■ 資料前處理所面臨的問題： ◆ 異質資料來源合併與結合 ◆ 資料清理問題 ◆ 欄位擴編與新訂欄位處理 ◆ 資料型態轉換與設定問題 ◆ 資料分割與抽樣
機械學習演算法選定	■ 分類演算法應用 ■ 函數逼近演算法應用 ■ 分群演算法應用 ■ 關聯演算法運用 ■ 時間預測演算法運用

機械學習演算法介紹與案例操作	■ Cluster 分群演算法 ■ 羅吉斯分類演算法 ■ 決策樹分類規則演算法 ■ 類神經演算法應用 ■ 逐步回歸演算法 ■ 時間序列預測 ■ 購物籃演算法 ■ 關聯演算法 ■ SVM 支持向量機演算法介紹 ■ 貝式網路應用 ■ R 語言的結合

工業技術研究院
Industrial Technology Research Institute

第二天

中文文本分析應用介紹	■ 為何需要中文文本分析? ■ 文本分析的重要性? ■ 中文大數據文本分析應用於維護紀錄介紹
文本分析之範例操作	■ RSS 新聞資料載入功能 ■ PDF 資料載入 ■ CSV、Execl、Access、Map source 與 GIS 節點設定說明
新聞文本數據分析案例介紹	■ RSS 文字資料前處理與規則訂定 ■ 關鍵字篩選與詞雲分析 ■ 詞彙關聯分析 ■ 自動文字分類 ■ 客製化字典建立
文本數據分析案例實作 實際案例與 Q&A	■ 文本數據分析案例建模 ■ 文詞歸類規則建立 ■ 實際案例與 Q&A ■ 學員個別案例討論

■ **建議參訓對象：**

1. 具備大專以上等相關背景之相關工程師
2. 有興趣之學員

註:此為上機課程，請學員務必自備正常 size 的 NoteBook。

- **備配需求：RAM 至少 4G，4G 以上最佳，硬碟至少 40G，建議具有獨立顯卡。**
- **作業系統：Windows 7 以上。**

■ **講師簡介：**

黃老師	➢ 經 歷：科技業-軟體系統處技術部經理 ➢ 專 長：機器人模擬、布置規劃、系統模擬建模、系統分析、決策分析、資料採礦、系統模擬專案導入、類神經網路、人因工程、認知心理學、人機介面研究應用

工業技術研究院
Industrial Technology Research Institute

■ 舉辦日期：106/08/17(四)~08/18(五) 10:00 -17:00 （共 12hrs）

■ 費　　用：

加入工研院產業學院會員 (http://college.itri.org.tw/LoginMember.aspx) 可以保存您的學習紀錄、查詢及檢視您自己的學習歷程，未來有相關課程優先獲得通 知及更多優惠!

全系列 (12hr)	原價	開課 10 天前或同一公司二人(含)以上報名
非會員	每人 10,000 元	每人 9,500 元
會員	每人 9,000 元 勤學點數(500 點)折抵	每人 8,500 元 勤學點數(500 點)折抵