

張詠淳助理教授學術分享：社群媒體文字探勘

張詠淳助理教授於 2008 年至 2017 年先後於中央研究院物理研究所及資訊科學研究所進行近 10 年的研究工作，並於 2016 年 6 月取得國立臺灣大學資訊管理博士學位。於 2017 年 8 月至北醫大任教，現為管理學院大數據科技與管理研究所助理教授。



張詠淳老師專精於文字探勘與自然語言處理、對談式問答機器人，及相關的機器學習與語言辨識技術。近期研究重點之一為「結合句構知識表達與機器學習」，打造一個具深度理解的智慧型對談機器人。另一個研究重點則是基於過去在文字分類上相關的研究成果，研發一個全自動的文字分類架構，此應用非常廣泛，例如：情感分析、主題偵測、甚至是問答系統等。

（編按）【左圖：張詠淳助理教授】

本研究團隊近期有別於以往的研究，從社群媒體中分析短文本（short text）主題對讀者產生的情緒，進而彙整成群眾之輿論與觀感。我們提出了群眾意見關鍵詞向量（Public Opinion Keyword Embeddings, POKE）用以表達每則來自社群媒體的短文本，從實驗結果顯示，本研究方法能有效地表達短文本群眾意見的意涵，並結合視覺化分析方法進而更深入瞭解社群多媒體中的群眾意見。其相關結果也用於分析 TripAdvisor 上，旅客意見中的情感分析對於飯店營運策略上之建議，相關研究成果也獲得國際期刊的肯定（Social media analytics: Extracting and visualizing Hilton hotel ratings and reviews from TripAdvisor. 2017, IF 3.872 [SCI, SSCI], categories 7% in INFORMATION SCIENCE & LIBRARY SCIENCE）。

此外，有鑑於國內缺乏一完善的中文前處理工具，使得在進行中文自然語言處理時困難重重。本實驗室與中央研究院智慧型代理人實驗室（IASL）共同開發一多目標中文前處理系統（Multi-Objective NER POS Annotator, Monpa），**Monpa** 是一套基於雙向遞歸神經網絡（bidirectional recurrent neural networks）的多任務標註方法，能夠同時進行斷詞（segmentation）、詞性（part-of-speech）標註以及專有名詞辨識（named entity recognition），透過 Monpa 的處理，除了能夠獲取基礎的詞彙資訊

外，所提供的專有名詞資訊（人名、地名以及組織名）有助於後續關鍵詞彙的抽取。後續，我們將朝向將此成果開源釋出，期許整夠嘉惠國內學子以及解決產業技術之缺口。（文/張詠淳，管理學院大數據科技與管理研究所助理教授）

【右圖：Monpa 操作介面】

