

開發機器學習之情資分析技術與應用



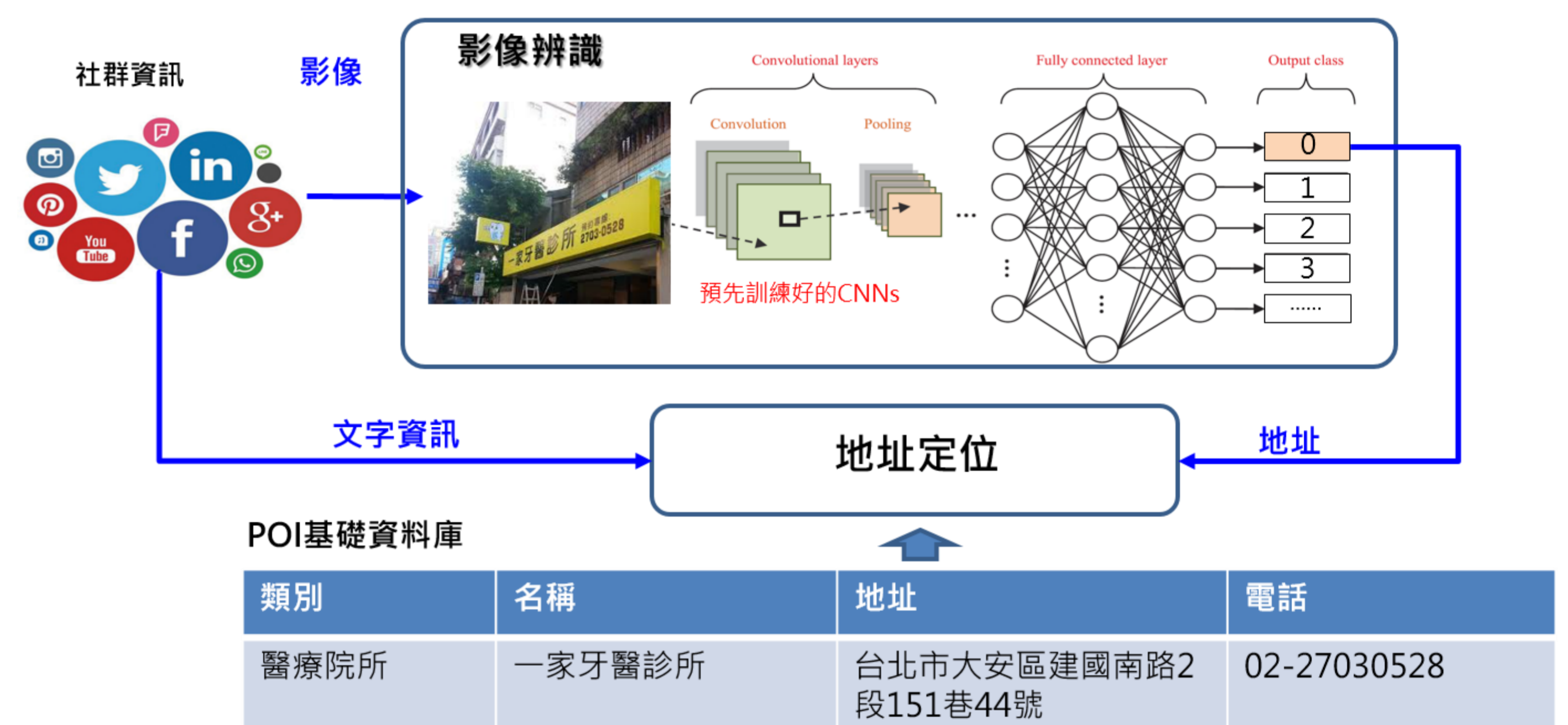
主管單位：國家災害防救科技中心
張子瑩 劉致灝 郭玫君 張智昌 陳俊元

研究背景

災防科技中心整合多元災防資料，包括基礎、監觀測、模式、以及災情資料等，超過20單位、近300項圖資，在多來源的災防資料中，透過機器學習的方式，能協助人員偵測重要訊息，或進行異常值偵測，以建立「智慧化」的第一步。本年度在機器學習上，著重在影像重要訊息的偵測，透過卷積神經網路法，從災情照片的特殊地標，再進一步判斷招牌上的電話資訊，得以準確進行電話與地址座標配對，以協助災情點位確認。另外也透過關聯規則探索技術，分析災防示警資料間的關聯性，以協助偵測可能因傳送系統設定、人為操作錯誤等，而發生短時間重複傳送、發布相同內容、疑似感測器(Sensor)故障造成短時間不斷發布/解除訊息，或未發布的示警即發布解除訊號等狀況，以輔助人員判斷資料的正確性，提高傳遞示警訊息的品質。

機器學習自動化分類技術

運用卷積神經網路(Convolutional Neural Networks, CNN)進行災情照片訓練及判釋，透過1000張招牌影像作為訓練樣本及模型建置，對於重要地標招牌的特徵，可有效辨識，採用隨機選取的200張影像進行連鎖商店招牌(統一超商、全家便利商店、全聯福利中心以及大潤發量販店)特徵辨識的驗證，正確率可達95%。因此再以相同的學習模式，針對招牌上的數字進行特徵值學習，以電話號碼辨識為主，透過電話號碼資訊，可以容易搜尋到正確地地址位置，以獲得更準確地位置定位座標。



機器學習自動資料品質告警

整理2015年至2018年的示警資料，透過關聯規則探索 (Association Rule Discovery) 方式，也就是機器學習中非監督式學習的一個演算法，找出資料共同出現的情況，分析資料間的相似性與相關性，透過分析出的每一個規則，以支持度(Support)、可信度(C Confidence)、增益度(Lift)三種規則特徵來找示警共同出現的機率。透過Apriori 演算法，在可信度(confidence)為0.2與支持度(support)為0.05的條件下，先找出112條規則，取出以降雨示警為首的示警的規則共52條，再把重複的規則去除後，最後獲得21條規則如表1。其中21條規格的三種特徵的分布情形，雖然支持度的機率都不高，僅第18條規則高於0.5，但可信度大於0.5且增益度皆有較高的趨勢，可為後續當相似的預警發布時的參考。以第2條規則說明，當降雨示警、水庫放流示警與雷雨警報發生時，同時會伴隨發布淹水示警的可信度(條件機率)為0.735，其增益度為2.047，代表當看到降雨示警、水庫放流示警與雷雨示警一起發布時，有很大的可能性會伴隨發布淹水示警，且遠大於降雨示警後發布淹水示警。

