

日期：2025/06/03

講者：呂昌田教授

題目：Future Insights_Harnessing AI and Social Media for Advanced Event and Epidemic Forecasting

心得報告

現代人或多或少都有在用社群媒體，也越來越多人的資訊是從社群上得知而非電視新聞，從 Facebook、Twitter 到 Instagram、Threads 等社群媒體漸漸地取代掉傳統電視，這種大家都能發送訊息、互動性高的大型聊天現場縮短了全球的距離。現在經過的每分每秒都有數以萬計的新貼文在社群媒體上產生，有些可能只是使用者隨手發出的無意義貼文，也有新聞媒體的時事報導或是政府的宣傳等，隨著人工智慧的發展與普及，社群媒體也引入了人工智慧讓使用者得到更個人化的內容，相輔相成讓使用者更常使用社群媒體，而企業再使用貼文來訓練模型。每一篇使用者貼文都能拿來當作即時數據來預測某地區甚至是全球正在發生什麼事情。例如：特定關鍵字如「發燒」、「咳嗽」、「確診」在這幾天內突然上升，則可能代表疫情又逐漸升溫；若出現某月某號某某地點則有可能是集會活動等。

講師講的這篇研究就是利用社群媒體的海量貼文來訓練預測事件的模型，其中包含疫情、暴動及集會，評估指標有 Quality（品質）、Lead time（提前時間）：預測事件比實際發生相差多少天、Recall（召回率）：預測到多少事件 Precision（準確度）以及 Probability（事件發生的可能性），這裡我有點遺漏掉簡報的幾個點但大概就是那些，講師的節奏有點快而且切簡報的速度也蠻快的我抄個筆記他就跳了 8 頁了。

模型的資料來源為 Twitter、Google Trends、Heatmap data 等資料互相搭配。資料型態通常包含特定關鍵詞像是疫情類的「發燒」、「咳嗽」，謠言類的「zombie」、暴力事件的「terrorists（恐怖分子）」等種類。資料的多樣性包含不同的語言、地區以及資料形式。接著是資料的可視化例如 Word Cloud、地圖等。

接著講師講了一套預測模型的應用：模型會先透過社群媒體上的關鍵詞數量變化多寡來監測疫情，看「生病」相關的詞彙是不是突然增加之類的，若增加且可能性高於特定值後使用 SEIR 模型來模擬病毒的蔓延過程，在這個時候就可以讓政府來超前部屬像是準備疫苗或是部屬防疫政策等。但疾病檢測的部分的挑戰有：範圍太廣無法得知準確地點（從何處開始擴散）、發文的人可能正在旅行造成地點不準確，以及一個模型要訓練的時間會拉很長，發現模型表現不好要重新訓練的話資訊就會落後很多……等問題。講師提出的一個應對方法是把目標再切割成一個一個很小的區塊，並且一個模型只專門負責一項特定任務，這樣訓練很多個針對細項的小模型後再丟到分類器（Classifier）裡，再輸出最後的結果。

我個人來看感覺就是：有資源果然才能幹大事。以現在社群媒體的貼文數量，光是收集資料就要一定的運算資源，更何況還要跑模型訓練。雖然整體的流程都差不多，收集資料、定義資料種類、調整模型訓練之類的，只是規模再放大個幾百倍的感覺。但後面講師的解法也給了我一點啟發：可以把一個大目標分成好幾個小目標，讓每個模型針對特定目標就好，這樣後續在優化感覺也比較快，像是疫情檢測的部分需要調整可能就不用整個模型重新訓練那樣。我剛好也有整個模型跑出來的結果不盡理想的狀況，或許可以嘗試看看再把目標切成更小的區塊，看看是否能改善成果。

關鍵字：社群媒體、人工智慧、疫情檢測

參考文獻

1. 傳統媒體 vs 社群媒體 <https://feja.org.tw/38614/>
2. 什麼是社群媒體人工智慧？ <https://www.kaspersky.com.tw/resource-center/preemptive-safety/social-media-ai>