書報討論書面報告

11363117 戴子璿

日期:2025/04/01

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:

這次的題目透過社群網路的貼文及貼文地理位置來進行事件預測，在這次的演講中，提到了很多問題解決方案，如貼文的取得方式及資料處裡，在關鍵字的策略就說到，透過Seed Queue的初始關鍵字清單中，挑選出1-2個關鍵字當作種子，進行文章搜尋，取得高頻關鍵字，在進行迭代，到最後收斂出事件及地點，透過動態的關鍵字組合，最佳化搜尋結果。

在效能的指標中，透過報導時間與預測時間來進行評估出lead-time來評分，還有Date score、Population score、Event Type及Concatenation Score來評估出分數，讓AI不斷訓練，增加準確度，配合最佳化搜尋結果，讓結果更加正確。

在這次的演講主題中，我了解了在AI訓練中可以做到的方法設計，以及問題解決方案，在未來的實驗中，也可以參考這些做法對遇到的問題來做解決方案的設計應用。

參考文獻:

該論文研究預測活動型社交平台(EBSN)中的活動參與情形，針對實際參與出席率進行預測，本研究特別重視出席率，因為它更能反映用戶參與意圖與主辦單位行銷策略的轉換成效。

該論文透過VENZI的社交APP平台來獲得以下四個數據，活動遊覽數、申請與通過人數、聊天訊息數及過往參加經驗等，比起一般的社群貼文更有預測價值。

實驗第一個步驟將資料來源進行前處理，移除機器人資料與重複紀錄、排除參與人數過少的活動、建立新特徵及資料正規化以便建模。該論文研究使用皮爾森（Pearson）與斯皮爾曼（Spearman）相關係數，分析各個特徵與目標變數的關聯性。特徵透過訊息標準差 × 聊天人數來捕捉互動強度。

使用三個模型來預測出席人數與出席數，線性回歸、Kappa加權距離模型及XGBoot。

最後XGBoost 表現最差，顯示此資料特徵更偏向線性或區域性加權分佈。線性回歸模型在預測總出席人數上表現最佳，Kappa 加權距離模型在預測出席率上略勝一籌。

該研究可以強化活動曝光、培養活躍用戶群，針對歷史參與活躍者進行定向行銷，將有效提高活動成功率。