

探索社群媒體使用者面對 ChatGPT 之情緒意向

吳家瑋、李皓鈞、王玗澄、郭達穎、吳怡潔

國立政治大學



1 簡介

2 相關研究

3 研究架構

4 結論分析與討論

5 結論與未來展望

簡介

- 大型語言模型 ChatGPT，因其強大的自然語言處理和對話能力，至今仍造成全球轟動。再者，相較其他社群平台，ChatGPT 是有史以來成長最快的應用程式之一。
- 然而，社會大眾對於 ChatGPT 的看法相當兩極化。
 - 1 ChatGPT 提供的內容可能有明顯的錯誤。
 - 2 公司擔心資料外洩，而明文規定使用規範。
 - 3 協助使用者免費完成文章或程式碼之改寫。
 - 4 幫助使用者處理簡單的翻譯或是文章摘要。
- 因此，我們好奇臺灣使用者對「ChatGPT」看法為何。

相關研究

本研究主要關切社群媒體上台灣使用者對 **ChatGPT** 的看法，因此將以兩大面向進行文獻探討。

- 台灣官方語言—繁體中文現有的自然語言處理工具。

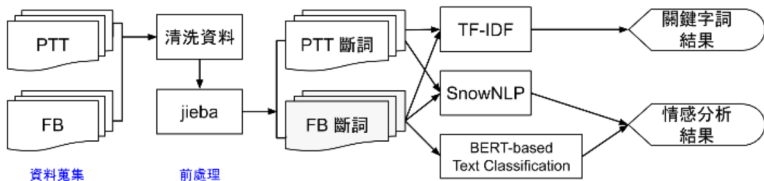
- 1 Jieba 是基於機器學習演算法的斷詞工具，它在斷詞的速度和精確度方面表現較佳。
- 2 SnowNLP 是一個基於機器學習演算法的工具，其主要專注於情緒分析。

- 網路社群意向蒐集與分析之相關方法。

- 1 情感分析模型 CSentiPackage(基於 CNN)，結合不同面向決定文章情感向度。
- 2 BERT 繁體中文預訓練模型，獲取嵌入向量 (embedding) 來建立後續情感分析模型。

研究架構

- 本研究想要探討的社群媒體主要在批踢踢實業坊（PTT）以及 Facebook。
- 根據本研究的架構依序流程為資料蒐集、資料前處理、關鍵資訊分析和情感分析。



研究架構-資料收集 (PTT)

- 受限於 PTT 尚未有 ChatGPT 討論專板，故本研究收集 2023 年 1 月到 4 月標題及內文有 ChatGPT 之文章並加以計算聲量 (social volume):

$$\text{聲量} = \frac{\text{提起 ChatGPT 文章數}}{\text{看板文章總數}} \quad (1)$$

研究架構-資料收集 (PTT)

- 雖然八卦板和閒聊板文章數較多，但聲量較股票板和科技工作板低，因此我們選擇「股票板」及「科技工作板」為分析對象。

| 看板 | Jan_GPT | Feb_GPT | Mar_GPT | Apr_GPT | Total_GPT | total_num | social_volume |
|-----------------|----------|-----------|----------|----------|-----------|-------------|--------------------|
| Gossiping | 19 | 293 | 171 | 118 | 601 | 180000 | 0.003338889 |
| C_Chat | 3 | 53 | 27 | 17 | 100 | 64200 | 0.001557632 |
| NBA | 0 | 1 | 0 | 0 | 1 | | |
| Stock | 6 | 14 | 6 | 5 | 31 | 5544 | 0.005591631 |
| Baseball | 0 | 2 | 0 | 0 | 2 | | |
| Military | 0 | 0 | 0 | 0 | 0 | | |
| LOL | 0 | 0 | 0 | 0 | 0 | | |
| Lifeismoney | 0 | 0 | 0 | 0 | 0 | | |
| Nswitch | 0 | 0 | 0 | 0 | 0 | | |
| HatePolitics | 0 | 52 | 19 | 7 | 78 | 42440 | 0.001837889 |
| car | 0 | 1 | 0 | 0 | 1 | | |
| PC_Shopping | 0 | 4 | 2 | 0 | 6 | | |
| home-sale | 0 | 0 | 0 | 0 | 0 | | |
| sex | 0 | 0 | 3 | 0 | 3 | | |
| KoreaStar | 0 | 0 | 0 | 0 | 0 | | |
| Beauty | 0 | 0 | 0 | 0 | 0 | | |
| movie | 0 | 0 | 0 | 0 | 0 | | |
| basketballTw | 0 | 0 | 0 | 1 | 1 | | |
| MobileComm | 0 | 0 | 0 | 4 | 4 | | |
| Tech_Job | 2 | 13 | 2 | 4 | 21 | 1743 | 0.012048193 |
| Japan_Travel | 0 | 0 | 0 | 0 | 0 | | |
| AllTogether | 1 | 0 | 0 | 0 | 1 | | |
| Elephants | 0 | 0 | 0 | 0 | 0 | | |
| japanavgirls | 0 | 0 | 1 | 0 | 1 | | |
| SportLottery | 0 | 0 | 0 | 0 | 0 | | |

研究架構-資料收集 (Facebook)

- Facebook 可以讓使用者成立社團討論主題，故本研究蒐集目前已建立的三個公開社團作為分析對象。
- 本研究固定資料蒐集的時間範圍從 2023 年二月到 2023 年四月。

| 公開社團名稱 | 建立日期 | 參與人數 | 蒐集貼文量 |
|---------------------------|------------|---------|-------|
| ChatGPT Taiwan | 2022/12/08 | 34,687 | 251 |
| ChatGPT & OpenAI 中文討論區 | 2023/02/07 | 137,999 | 586 |
| ChatGPT 生活運用 | 2023/02/17 | 242,394 | 603 |

研究架構-資料前處理

此步驟的主要任務為清洗資料。

- 去除網址。
- 去除停止詞。(如：大家好、嗨等頻繁出現的功能性詞語)
- 使用 Jieba 進行斷詞。

研究架構-關鍵資訊分析

計算關鍵詞用 Term Frequency-Inverse Document Frequency:

- Term Frequency (TF):

$$TF(t, d) = \frac{\text{Number of times term } t \text{ appears in document } d}{\text{Total number of terms in document } d}$$

- Inverse Document Frequency (IDF):

$$IDF(t) = \log \left(\frac{\text{Total number of documents}}{\text{Number of documents with term } t \text{ in it}} \right)$$

- TF-IDF (Term Frequency-Inverse Document Frequency):

$$TF\text{-}IDF(t, d) = TF(t, d) \times IDF(t)$$

研究架構-情感分析 (SnowNLP)

- SnowNLP 可判斷文章的意象，越接近 0 代表負向而越接近 1 則代表正向。
- 本研究認為須考慮中立立場，故定義 threshold 如下：

$$y = \begin{cases} 0 \leq y < 0.45 & , y = \textit{Negative} \\ 0.45 \leq y \leq 0.55 & , y = \textit{Netural} \\ 0.55 < y \leq 1 & , y = \textit{Postive} \end{cases}$$

研究架構-情感分析 (BERT 嵌入向量訓練模型)

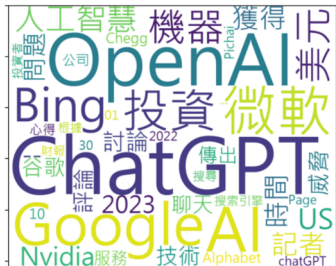
- 本研究根據 NTUSD 的 11,088 個語意詞，再加上 FB 較常出現之情緒詞，總共 12339 個來訓練模型。
- 立場定義如下：

$$x = \begin{cases} |P_{c1}(x) - P_{c2}(x)| \leq 0.1, x = \text{Netural} \\ |P_{c1}(x) - P_{c2}(x)| > 0.1, P_{c1}(x) > P_{c2}(x), x = \text{Positive} \\ |P_{c1}(x) - P_{c2}(x)| > 0.1, P_{c1}(x) < P_{c2}(x), x = \text{Negative} \end{cases}$$

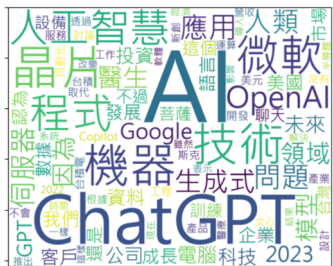
其中 $P_{c1}(x)$ 為模型判斷詞語 x 為正面機率； $P_{c2}(x)$ 為負面機率。

- 最後加總文章中所有 x 的機率，判別文章的情緒意象。

- 股票板使用者較注重於 ChatGPT 乃至 AI 產業投資前景。
- 科技工作板使用者關心 AI 硬體產業外，也涵蓋生成式技術之功能，但同時也質疑此技術是否真能取代現有人類工作或是否會造成職場衝擊。



股票版關鍵字分析



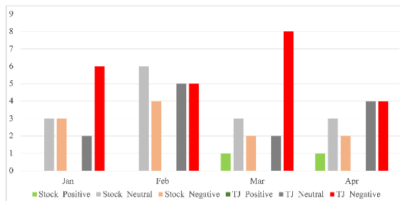
科技工作關鍵字分析

- 本研究從月份和全局之角度對 FB 三個社團和整體進行分析。
- 不同社團間仍存在小部分差異，例如正負向字詞比例大小，而「簡單」、「方便」、「厲害」是所有討論區一致的看法。



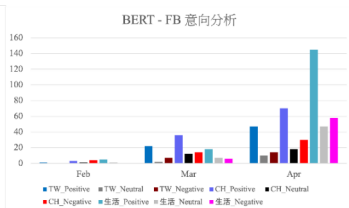
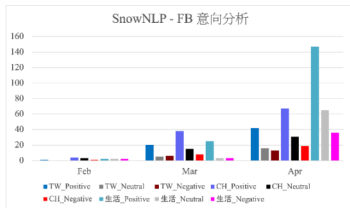
情感分析—PTT

- 由於 PTT 較 FB 的貼文數少，故選用 SnowNLP 分析情感。
- 科技工作板無正向文章，股票板在三月後才有正向文章。
- 股票板中立文章大致多於負向文章，三月後分類文章數雙雙下降，不過中立負向的比例變動不大。
- 科技工作板於一月和三月皆出現負向文章數多於中立的情況，從內容研判原因包括 OpenAI 出現競爭對手、擔憂 AI 取代工作以及看壞台灣 AI 產業發展。



情感分析—FB

- Facebook 中，本研究同時考量 SnowNLP 和 BERT 嵌入向量訓練的模型，綜合結果來判斷文章的意見向度。
- ChatGPT Taiwan 的正向貼文比例最高，其他兩個社團則在四月之後，正向貼文比例才有增加。
- ChatGPT 生活運用在四月則出現最多負向的貼文。



結論與未來展望

■ 結論

- 1 PTT 上的股票板使用者較關注和 AI 相關的投資及發展前景。
- 2 科技工作板使用者討論的範圍較深較廣，包括 AI 硬體產業鏈、生成式 AI 技術及背景以及可能造成的危害與威脅。
- 3 PTT 中兩版的情感立場以中立及負面為主。
- 4 ChatGPT Taiwan 的正向貼文比例最高，而 ChatGPT 生活運用在四月則出現最多負向的貼文。

■ 未來展望

- 1 加強文字處理和情感分析工具的效能品質。
- 2 擴展社群媒體範圍，如：整站 PTT。
- 3 整合國際間對 ChatGPT 的意向分析。

感謝大家聆聽