

臺灣財經情緒字典之編製與應用

Traditional Chinese Financial Dictionary with Sentiment Word Lists

黃裕烈*

國立清華大學計量財務與金融學系

國立臺灣大學計量理論與應用研究中心 (CRETA)

關鍵詞：GPT、大型語言模型、文字探勘、財經情緒字典

Keyword：Dictionary, GPT, large language model, text mining,

JEL 分類代號：C10, E30, O11

* 聯繫作者：黃裕烈，國立清華大學計量財務金融學系，新竹 30013 光復路二段 101 號，電話：(03) 516-2125，傳真：(03) 562-1823，E-mail: ylihuang@mx.nthu.edu.tw。作者感謝吳俊毅副研究員、徐之強以及徐士勛教授對本文初稿所提供的寶貴意見。此外，我們也感謝周玟妍，李佳儒，李冠宇，張庭維，簡丞威，林昕宏，戴唯倫，曾彥寧等多位研究生的協助。本研究承蒙教育部高教深耕計畫經費補助（計畫編號：113L900201）以及國家科學委員會計畫經費補助（計畫編號：NSTC-112-2410-H-007-090，NSTC-113-2410-H-007-068），在此一併致謝。文中若有任何錯誤，當屬作者之責。

摘要

本文旨在利用臺灣長期的財經新聞文本，結合大量人力、建立分類指標，並透過大型語言模型的協助，編製一部適合臺灣財經新聞用詞的情緒字典。藉由此系統性的編製過程，希望能長期維護此財經情緒字典，填補國內財經領域在文字資訊上的缺口。經過一系列的篩選過程，我們選出 9,890 個財經情緒詞彙，將其應用在中央銀行新聞稿的情緒分析，並與 GPT-4 模型的情緒評分進行比較。雖然直觀上 GPT-4 模型在文本情緒分析中表現會更為出色，但實證結果顯示，財經情緒字典在處理經濟發展以及金融穩定等議題時仍有一定的水準。此外，若研究者想從大量文本中快速獲得初步的情緒分析，或因資訊安全考量無法利用大型語言模型進行分析時，透情緒字典的評分方式能在效率與成本上具一定的優勢。最後，此情緒字典的編製過程除了可以用來維護或擴充財經情緒字典以外，也可適用在如 ESG (environmental social governance) 等相關議題上，並且透過字典所提供的相關資訊，也可提高相關套件在中文斷字 (segmentation) 的精確度。

Abstract

With the assistance of large language models, this study aims to develop a traditional Chinese financial sentiment dictionary using financial news reports from Taiwan. We first annotate the news reports and construct an indicator to select sentiment words. Then, we ask GPT-4 to provide additional sentiment words to enhance the dictionary. Through this systematic process, our goal is to fill the gap in Taiwan's financial sentiment analysis resources. After careful selection, we identified 9,890 financial sentiment words, which were applied in the sentiment analysis of the Central Bank's press releases. Comparing the results with those of the GPT-4 model, we find that the financial sentiment dictionary has achieved a certain level of accuracy in handling texts related to economic development and financial stability. In addition, for researchers looking to quickly obtain preliminary sentiment analysis results from large text datasets, the financial sentiment dictionary can offer advantages in terms of both efficiency and cost. Finally, this developing process can not only be used to maintain or expand the financial sentiment dictionary, but can also be applied to other related issues, such as ESG (environmental, social and governance). Moreover, the provided information in the dictionary can also be helpful for Chinese word segmentation.

1. 前言

透過文字資訊的輔佐來進行相關的實證分析是近期相當熱門的一個議題 (如 Baker et al., 2016 與 Thorsrud, 2020)，也是許多學者嘗試處理的一個問題。由於文字資訊為非結構性 (unstructured) 資料，因此研究者無法立即透過統計或計量方法進行分析，必須事先進行處理，轉換成數據資料 (numerical data) 後才能使用。經文字轉換後的數據資料若可以提供研究者額外的資訊並獲得顯著的實證結果，我們便認為此數據資料是一個重要的財經變數，其資訊內涵無法從既有的資訊中獲取。近期許多文獻 (如 Baker et al., 2016; Caldara and Iacoviello, 2022; Husted et al., 2022) 便依照上述的概念，從相關的文字資訊中萃取出重要的數據；如編製經濟不確定性 (economic uncertainty)、地域政治風險 (geopolitical risk) 與貨幣政策不確定 (monetary policy uncertainty) 等相關指標來進行實證分析。

為能讓非結構性的文字資料轉換成數據資料，研究者常透過分類模型 (或稱分類器, classifier) 來建構數據變數。最常見且應用範圍最廣的分類器便是以字典 (或關鍵字) 方式來建構數據變數；例如 Loughran and McDonald (2011)，Jiang et al. (2019) 與姜富偉等人 (2021) 透過文本資料編製英文與簡體中文的財經情緒字典來探討相關議題，而 Baker et al. (2016) 以及 Husted et al. (2022) 則是以關鍵字搜尋方法來探討不確定性的問題。另一常見的分類器則是利用類神經網路 (neural network, 簡稱 NN) 或機器學習分類模型來進行文本分類，進而建構數據變數；例如 Huang and Kuan (2021) 以及 Chen et al. (2024) 等便利用機器學習以及 NN 模型來建立數據變數，而近期的大型語言模型 (large language models, LLMs) 也可被歸類此。相較於字典方式，以機器學習模型進行文本分類雖然較新穎，但在文獻中卻常觀察到，不論採用哪一種分類模型，在多數的實證分析中亦會利用字典方法來建立情緒指標進行探討 (例如 Li, et al., 2024)。因此，利用字典法來建構情緒指標是一項重要且基礎的工作。

從計量的角度來看，建構一部好的字典除了需要有適當的分類方法

外，還需視資料來源進行調整，再加上人力、時間成本等各方面的配合才可能完成，是一項苦差事。例如，若分析的文本為財經相關新聞或是政府財經部門的文字紀要時，Loughran and McDonald (2011) 就強烈建議應避免利用大眾常用的情緒字眼來進行分析，反而應該編製一部財經情緒字典才能避免語意錯誤。例如 Ku et al. (2009) 基於社群媒體資訊所建立的台大情緒字典中，「財富」、「資源」、「失業人數」與「徵稅」等都被歸類為情緒字眼。但在財經相關的新聞中，這些字詞不見得都具有正、負面意涵，常常只是代表某一個名詞或數據變數而已。另外，文本橫跨的資料時間也要夠長，才能考量如「金融海嘯」、「量化寬鬆 (QE)」、與「COVID-19」等字詞的影響性。最後，也不宜直接將國外的字典直接翻譯（或簡轉繁體），因為不同的國情，用字遣詞一定會有差異。例如在 Jiang et al. (2019) 與姜富偉等人 (2021) 的簡體財經字典中，「樂顛顛」、「愛崗敬業」、「出籠」、「漏子」與「新股弱勢」等均為財經相關的情緒字詞，但這些字眼在國內並不常見，也不一定有相似的用法與意涵。

若從分類的角度來看，以 NN 模型來分類文本其正確率較高，但字典方法確可以為研究者帶來許多模型無法提供的效益。首先，由於字典中每一個字詞的意涵都相當精確，因此分類結果不容易有模糊不清的情況。例如「經濟復甦」、「成長優於預期」與「景氣增溫」等都是對景氣有正面的描述，而「寬鬆貨幣政策」與「調降政策利率」則是對貨幣政策有明確的指示；¹所以透過字典方式所分類出的詞義 (semantic) 相當精確，且不易產生混淆。但反觀 NN 模型 (如 BERT 模型)，其輸出的分類結果常以機率表示 (例如 54% 為正面，46% 為反面)，因此分類結果常會有含糊不清的情況，並且無法從中判斷此分類結果是受到那些情緒字眼所影響。另外，若是以詞嵌入向量 (word embedding) 來表示 NN 模型字詞時，每一個字詞 (例如「經濟復甦」) 都是由一串數字向量來代表，很難從中了解其經濟意涵。其次，透過情緒字典，研究者可以將此字典應用在許多不同

¹ 在此我們排除否定 (negation) 與諷刺 (irony) 的詞句情況。

的議題上，並快速了解文本對某一議題的正、負面看法，可視為非監督式 (unsupervised) 學習的一種分類器。然而若要採用 NN 等監督式學習的模型進行判斷，需要先花費大量人力或是金錢對不同的議題進行標記，因此常緩不濟急。最後，字典的維護與修改相當容易，並且可以針對不同時期以及議題進行增減，而 NN 模型則無法與其相提並論。由於字典法有這些優點，所以建構財經情緒字典相關的文章均發表在重要的財經期刊上 (如 Loughran and McDonald, 2011, *Journal of Finance*; Jiang et al., 2019, *Journal of Financial Economics*)，且其引用次數相當高，顯見其重要性。²

本文最主要的目的為利用文字探勘的技術，建構一套臺灣的財經情緒字典。本文的主要特色有：(1) 我們所採用的文本為臺灣網路財經相關新聞，據此所編製出的財經字典一定是臺灣新聞常用的字詞，不會夾雜其它無關的用字。(2) 我們採用的資料從 2003 年開始至 2023 年底，長達 20 年以上的網路財經新聞，因此橫跨的時間相當長。(3) 雖然國內有少數研究生論文會自行編製情緒字典，但多半是將國外的字典直接翻譯，或是依據非財經相關的文本內容編製而成，其資料期間短暫，無法進行長期維護。(4) 本文也嘗試透過 LLMs 的協助來編製字典 (容後說明)。故就我們所知，本文是第一篇利用長期的財經文本來編製臺灣財經情緒字典，希望透過相關分類模型之建立，可以簡單且長期維護此情緒字典，彌補國內財經領域在文字資訊的缺口。

我們考量以下兩種模型來建構財經情緒字典，分別是 PMI (pointwise mutual information) 指數以及 GPT-4 模型。以 PMI 指標編製字典是相當常見的一種方式，例如 Chen et al. (2018) 便依此方法，透過人工標記來編製財經社群媒體資料的英文情緒字典。利用 PMI 指標篩選出臺灣常用的情緒字詞 (以下簡稱基本情緒字詞) 後，我們還會下達適當的提問指令 (prompt)，要求 GPT-4 檢視這些情緒字詞是否合宜，並生成 (generate) 其它相關的情緒字詞來擴大我們的字詞數目 (以下簡稱擴充情緒字詞)。最

² 依據 google scholar 的資料顯示，Loughran and McDonald (2011) 至 2023 年底已有 5,000 筆以上的引用次數。

後，為避免 GPT-4 模型因幻覺 (hallucination) 而產生不恰當的字詞，或是所建議的情緒字詞並非臺灣常用的字眼，我們還會以人工方式檢視結果，並且計算這些情緒字詞是否曾出現在臺灣財經相關新聞中，刪除出現頻率過低的字詞，以編製成最後的財經情緒字典。這一系列的編製過程除了可以用來維護或擴充目前的財經情緒字典，也可應用在其它議題 (如 ESG, environmental social governance) 的情緒分析。此外，透過本文所提供的相關資訊，還可提高軟體套件在中文斷字 (segmentation) 的精確度。

為了解此財經情緒字典的優、劣勢，我們將歷年中央銀行 (簡稱央行) 在網站上公佈的會議新聞稿文字內容視為樣本外資料，透過 GPT-4 將該文本的句子依央行之四大經營目標 (mandates) 進行分類並給予情緒分數，還利用財經情緒字典中的正、負面情緒字詞給予評分，最後比較 GPT-4 與財經情緒字典的情緒評分結果。這樣的比較相當有趣。首先，GPT-4 因背後有大量資訊以及演算法的支持，所以直觀上此模型給予的情緒分數會比較合理。但 GPT-4 需要花一筆費用，並且因字符 (token) 數目的限制，所以需要等待一段時間才能完成所有任務。³但反觀財經情緒字典，雖然其評分的合理性與正確性理應低於 GPT-4，但它不需要花費任何費用，並且很快便能完成情緒分數的計算，因此在處理大量的文本或是機密文件時，財經情緒字典仍有一定的優勢。此外，我們也發現在處理央行經營目標「協助經濟之發展」與「促進金融穩定」相關文本內容時，財經情緒字典的評分表現與 GPT-4 相似，這表示我們所編製的財經情緒字典在這些方面的表現已有一定的水準。

本文的內容安排如下：第 2 節主要介紹資料的處理方式，第 3 節為模型簡介，包含 PMI 指標以及 GPT-4 的提問指令，第 4 節為實證結果，最後一節為結論。

³ 因為文本內容不多，所以在 GPT-4 的花費為 1.31 美元，但要花約 1 天的時間才能完成全部的任務。

2. 資料處理

為提高財經情緒字典的品質，在整理與收集資料時我們希望能考量財經相關的文本來源，並且涵蓋較長的資料期間。但因為人力的限制，我們不可能將每一天、每一則新聞均納入考量的範圍。為取得一個平衡，我們以央行新聞稿公佈的時間為基準，收集該日期前後 10 天（共 21 天）的網路新聞內容來當成編製基本情緒字詞的文本，以下為更詳細的資料處理方式。

2.1 央行新聞稿處理原則

我們先撰寫爬蟲程式 (web crawler) 從中央銀行全球資訊網站中下載歷年「中央銀行常務理事會議決議新聞稿」以及「中央銀行理監事聯席會議決議新聞稿」文字內容（以下簡稱央行新聞稿），⁴而網站中的其它資訊（如理監事聯席會議貨幣政策議事錄摘要、理監事會後記者會參考資料等）則不納入考量。為了配合各大報的網路新聞資訊期間，我們只收集 2003-06-26 至 2023-12-31 共 90 則央行新聞稿資訊。

我們將央行新聞稿視為政府經財部門的文本代表，主要是因為這些新聞稿除了會包含當前的財經時勢外，也涵蓋景氣、貨幣政策等不同議題，更是新聞媒體以及大眾所關注的焦點。並且央行新聞稿內容多從政府的角度來觀察經濟現狀，其用字遣詞必字字珠璣，情緒字眼也多偏保守或正面。但相反地，為引起讀者的興趣，網路新聞在報導或討論議題時，其使用的情緒字詞會更多元、更廣泛。由於央行新聞稿的資料特性，故在編製財經情緒字典時我們先將這些文本排除在外，以做為樣本外的文字集合。

2.2 網路新聞稿處理原則

為建立所需之文本，我們先撰寫爬蟲程式收集各大報（包含中國時報，聯合報，自由時報，蘋果日報，工商時報，經濟日報）的網路新聞內

⁴ 網址為 <https://www.cbc.gov.tw/tw/lp-357-1-1-20.htm>.

表 1: 各大網路報紙相關資訊與內容

報紙名稱	起始日期	新聞分類
中國時報	2009/9/28	焦點要聞, 財經焦點, 時論廣場
工商時報	2009/9/28	財經要聞, 企業經營, 投資理財, 金融·稅務, 產業·科技, 證券·權證
自由時報	2005/1/1	focus, business, 財經, 焦點
蘋果日報	2003/5/1	頭條要聞, 財經新聞, finance, 綜合報導, property
聯合報	2014/12/17	產經, 股市, 要聞, 房市, 證券, 金融
經濟日報	2014/12/23	產經, 股市, 要聞, 房市, 證券, 金融

容。⁵因網路新聞的起始日期不同, 並且新聞中也包含許多與財經無關的報導 (如各大報的娛樂新聞或是副刊新聞), 為方便後續分析, 我們依表 1 新聞類別進行處理, 只保留財經相關的新聞內容。因每日的新聞資訊量相當多, 不太可能對每一則新聞進行處理。因此, 我們以每次央行理監事會議的新聞稿日期為基準, 收集該日期前後 10 天的新聞內容來編製基本情緒字詞, 以做為折衷的處理方式。

透過 PMI 模型來篩選基本情緒字詞時需要先對文本進行標記, 所以我們將收集到的新聞內容, 以段落 (paragraph) 為模型的基本輸入單位, 透過隨機方式抽出大約 15,000 筆段落進行人工標記。⁶我們會以段落為基本文本資料, 主要是考量一篇新聞報導中可能會同時包含正、負面情緒字詞, 但為了讓讀者能清楚撰寫者想表達的意思, 常會在一個段落中只表達一種類別 (如正面) 的觀點, 而在另一段落中表達不同的觀點以平衡報導, 甚至也可能發生某些段落的內容與財經議題無關 (如背景相關報導) 的情況, 因此才會以段落為文本基本單位。若是某一段落中標記人員發現同時包含正、負面情緒語句時, 則以人工方式將其分成數個段落, 再分別標記結果。為減少人為標記的錯誤, 每一個段落至少會有 2 位研究人員進行標記, 最後再經作者檢查標記內容, 供後續使用。

⁵ 我們主要是收集各大報的新聞網址 (URL), 找出符合條件的網址後, 再透過程式前往該網址中對應的新聞內容進行後續的計算; 所有程式以 Python 與 R 進行撰寫。

⁶ 我們是以 HTML 標記 “\r\n” 或 “\n” 為段落區分符號。若該篇文章無法收集到此標記, 則視全文為一段落。另外, 為能有通盤的考量, 在隨機抽取時我們還要求在樣本期間內, 每天、每一份報紙的每一主題至少要有 2 個段落被選取到, 也避免挑選出太長或太短的段落; 相關細節此處就不多贅述。

雖然人工標記所需的人力相當繁雜，但若編製出一部專用的財經情緒字典，會對後續研究者帶來許多效益。例如，不需要再花時間與金錢標記內容，也不必考量資訊安全問題，便可透過字典搜尋方式快速地了解文本的情緒判斷。又例如後續學者能以此字典為基礎，與時俱進地增加字典的內容，如同 Loughran and McDonald 的 Master Dictionary with Sentiment Word Lists 字典一樣，不斷精進更新。

3. 模型設定

本節會依序介紹 PMI 指數以及 GPT-4 的提問指令。利用人工標記好的文本資訊，透過 PMI 指數來篩選出基本情緒字詞，再利用 GPT-4 提問指令，擴充這些基本情緒字詞。最後，我們會計算所有情緒字詞是否曾出現在臺灣財經相關新聞中，刪除出現頻率過低的字詞，以編製成最後的財經情緒字典；以下說明這些處理過程。

3.1 PMI 模型

PMI 指數主要是用來衡量兩個事件之間的相關性，其原理很簡單，但應用的範圍相當廣泛。例如 Pendzel et al. (2024) 透過 PMI 指數來了解美國社群媒體資料 (social media data) 中，各黨派發佈不文明言論 (political incivility) 的情況。又如 Chen et al. (2018) 曾利用此方法來編製財經社群媒體資料的英文情緒字典，他們發現透過 PMI 指數所找出的字典，其分類表現優於傳統的 χ^2 卡方檢定以及類神經網路 CNN (convolutional neural network) 模型。

若以編製中文情緒字典為例，PMI 指數主要計算以下式子：

$$\text{PMI}(\tau, s) = \log_2 \frac{P(\tau, s)}{P(\tau)P(s)}, \quad -\infty \leq \text{PMI}(\tau, s) \leq \min\{-\log_2 P(\tau), -\log_2 P(s)\},$$

其中 s 為某一段落的情緒類別 (如正面或負面)，而 τ 為該段落文字經

過中文斷詞後的某一字詞。⁷不難看出 PMI 主要是計算 s 與 τ 是否相關，若兩者相互獨立，則 PMI 值為 0；若文本中常利用字詞 τ 來描述 s 類別的情緒，則 PMI 值便會愈大，應該考慮將此字詞納入字典中。我們會利用標記好的段落，透過中央研究院詞庫小組所發展的中文斷詞暨實體辨識系統 (CKIP-Tagger) 就內容進行斷詞，再以詞頻 (term frequency) 為權重計算每一個字詞的加權 PMI 值：

$$\text{wPMI}(\tau, s) = \frac{\tau}{T} \text{PMI}(\tau, s),$$

其中 T 為字詞總數，依此排序，再以人工方式挑選出 wPMI 值較高且適當的字詞來編製財經情緒字典。至於 wPMI 的門檻值為何並無定論，只能以人工方式盡力篩選出適當的字詞。

3.2 GPT-4 提問指令

由於我們只利用隨機抽出的段落進行基本情緒字詞之篩選，並且文本來源也只限於央行理監事會議前後 10 天的網路新聞內容，恐多有遺珠之憾。因此我們也透過 GPT-4 這類大型語言模型的 API (application programming interface) 協助，嘗試擴充情緒字詞。為了讓 GPT-4 生成情緒字詞，我們參考 OpenAI Platform 使用手冊的建議，在提問中設定角色並提供詳細的資訊與例子供模型參考。以下是我們所使用的提問指令：

“你是一位熟悉臺灣總體經濟情況以及繁體中文的財經專家，每天都會閱讀大量的財經新聞報導。你的任務是編製臺灣的財經情緒字詞，如同 Loughran and McDonald 的 Master Dictionary with Sentiment Word Lists 字典一樣。請參考以下財經常用的情緒字詞，並額外提供與這些字詞相關的情緒字詞，不能重覆，不用解釋，但卻是財經新聞報導中

⁷ 處理中文斷詞時，我們也會以 n -Gram 方式適時地找出常見的字詞組合 (compound words)。例如將「景氣」與「循環」組合成「景氣循環」一詞，並以長詞優先方式來進行斷詞。

會採用的情緒字詞。例如參考字詞為‘猛進’，你會額外提供‘激增’等 20 個字詞。以下是參考的用語：”

其中參考用語為每一個基本情緒字詞。最後，我們會要求以列表方式輸出繁體中文字詞，方便後續整理。⁸例如若參考用語為‘敏捷’，則 GPT-4 會列表輸出‘機敏’、‘反應迅速’與‘決策果斷’等字詞，我們會取其聯集以形成擴充情緒字詞。

為避免 GPT-4 模型因幻覺而產生不恰當的字詞，或是所建議的情緒字詞為大陸用語，在臺灣並不常見，我們還會計算所有情緒字詞是否曾出現在臺灣財經相關新聞中。我們以 2003-05-01 至 2023-12-31 每日、每則財經新聞為文本基礎（共 1,560,766 則新聞），計算每一個情緒字詞出現的次數，刪除出現頻率過低（次數小於 30）的字詞，並以人工方式檢視最後結果。換言之，最後篩選出的情緒字詞都曾出現在臺灣財經相關新聞中。

4. 實證結果

4.1 財經情緒字典

透過上述的編製過程，我們篩選出歷年來臺灣常見的財經情緒字詞，包含正面情緒字詞 4,853 字，負面情緒字詞 5,037 字，總共 9,890 個字詞。表 2 列舉出 32 個正、負面情緒字詞，其中“次數”表示該字詞曾在財經新聞報導中出現過的次數，而“篇數”是指該字詞曾出現過的新聞篇數。因為一篇報導中相同的字詞可能出現 2 次以上，所以篇數值不會高於次數值。透過這些“次數”以及“篇數”的資訊，一方面可以了解到財經新聞常用的正、負面字詞，也可以利用這些訊息來當成情緒字詞的權重，或刪除“次數”較少的字詞以編製更精簡的情緒字典。而表中前 8

⁸ 其它 GPT-4 的參數設定為 temperature = 0.7，frequency penalty = 1，presence penalty = 1，其中 temperature 愈高表示生成的文字愈沒有“創意”，frequency (presence) penalty 愈高則代表給與重覆（已經）出現的字詞愈重程度的懲罰。然而依據我們的經驗，此處參數設定的影響不大。

表 2: 正、負面財經情緒字詞

正面字詞	次數	篇數	負面字詞	次數	篇數
獨一無二的	2891	2686	不確定性升高	687	642
更上一層樓	2683	2606	不確定性增加	627	602
值得信賴的	1600	1517	不可避免的	618	609
最具影響力	1176	1055	失業率上升	479	432
當紅炸子雞	1134	1096	一發不可收拾	461	455
得天獨厚的	881	861	越來越嚴重	406	388
具指標意義	789	780	將受到衝擊	385	378
有吸引力的	769	738	不切實際的	370	348
經濟成長	116973	64029	疫情影響	43875	35788
歷史新高	106602	77389	不如預期	28998	25386
大幅成長	42045	36536	金融海嘯	23563	18393
持續成長	39808	36071	疫情衝擊	21599	18480
永續發展	36142	21141	受到影響	14992	13941
優於預期	29221	24560	經濟衰退	13022	10032
值得注意	24212	23637	金融危機	11231	8322
經濟復甦	23100	19025	通貨膨脹	10482	8253
有機會	192181	148364	不確定	59191	49115
創新高	82646	65304	受影響	30832	26155
受惠於	67833	58122	不同意	16752	8170
有助於	50678	45483	負成長	14171	9575
如預期	36017	31565	高風險	11644	9001
全方位	29573	24866	嚴重的	11520	10645
穩定的	25195	23165	不得不	11126	10462
完整的	24777	22607	不景氣	10887	8949
投資	1626852	500164	風險	330633	179813
成長	1294869	495825	問題	321119	207579
獲利	609395	284442	減少	237532	156849
預期	542669	319615	無法	183048	143018
增加	522480	324672	衝擊	180704	134233
合作	480494	244281	降低	179851	141118
智慧	447075	172387	降低	179851	141118
新高	430309	235292	壓力	177794	127039

註: 文本計數來源為 2003-05-01~2023-12-31 各大網路財經新聞, 共 1,560,766 則新聞。

列 (row) 為字詞長度至少為 5 的正、負面情緒字詞，並且依“次數”大小排序。然後每 8 列為一組，依序列出字詞長度為 4 ~ 2 的正、負面情緒字詞。而字典中也包含 11 個單字，例如“漲”、“勝”以及“跌”、“差”等，在表中並未列出。為了方便後續維護以及使用，我們已將此字典內容放在網路上供大家下載，⁹並且依長詞優先以及次數頻率（由高至低）排序。值得一提的是，若依此排序方式加入中文斷字套件（如 Jieba）的字典中，可提高套件在斷字的精確度。此外，我們也參考 Loughran and McDonald 在網站上公佈的內容，增加“入選年度”的欄位，以表示該字詞是在那一年度被選入字典中。而在下載的檔案中，我們也撰寫一些 R 與 Python 函數，以利使用者透過此財經字典統計某一篇文本所出現過的情緒字詞次數。

4.2 樣本外實證分析

在編製財經情緒字典時，我們是依據網路報紙的新聞內容來篩選字詞，並沒有涉及到政府財經機關的文字內容。為了解此字典的表現，們利用央行新聞稿來當成我們樣本外的資訊，藉此分析央行新聞稿的情緒內容。依 Huang and Kuan (2021) 的實證發現，若將央行新聞稿內容依其經營目標先進行分類，其實證結果會比較有義意。據此，我們透過 GPT-4 將央行新聞稿的內容依其經營目標進行分類並給予情緒分數，同時，也透過我們所編製的財經情緒字典給予評分，再來比較其結果。我們所採用的提問指令大致如下：

“你是一位臺灣的中央銀行官員，熟悉中央銀行的貨幣政策目標，臺灣的總體經濟情況以及繁體中文。中央銀行的經營目標為促進金融穩定；健全銀行業務；維護對內及對外幣值之穩定；並於上列目標範圍內，協助經濟發展。這四項經營目標的說明如下。「金融穩定」指金融體系有

⁹ 下載網址為：<https://remi.ctmnthu.com/portfolio-archive/financial-dictionary/>

能力在不同經濟活動及不同期間有效率地分配資源，評估及管理金融風險以及承受不利衝擊。中央銀行特別重視總體金融穩定，持續評估金融體系風險可能來源，並檢視國內外總體經濟金融環境發展，適時採行總體審慎措施 (macroprudential measures)，例如調整特定不動產貸款的貸款成數 (Loan to Value Ratio) 限制，促使金融機構落實對不動產授信的風險控管，以避免系統危機 (systemic crisis) 的傷害。「健全銀行業務」可促使銀行體系安全暨效率運行，有助金融穩定；亦可強化金融中介功能，益於貨幣政策傳遞機制。中央銀行會適時採行妥適貨幣政策措施，並採總體審慎措施，致力於健全銀行業務。「維護對內及對外幣值之穩定」：一般而言，所謂對內幣值的穩定係指國內物價穩定，而對外幣值的穩定係指新台幣匯率的動態穩定，以維持新台幣對內及對外的購買力。「協助經濟發展」：在景氣過熱時，採取貨幣緊縮性政策；在景氣不振時，採取貨幣寬鬆性政策，有助於穩定經濟，並協助經濟成長。請以句子為基本並且將句子放在 sentiment 欄位，再將句子進行分類並且給予情緒分數。...

我們要求 GPT-4 先依「促進金融穩定」、「健全銀行業務」、「維護對內及對外幣值之穩定」、「協助經濟發展」與「其它」對文本進行分類，每一個句子可以有多個標記結果，但每一個句子都要標記清楚。接著，我們再要求 GPT-4 對分類後的句子給予文字情緒分數，正面為 1，負面為 -1 而其它則為 0。最後，我們將 GPT-4 評分過的句子依央行開會的日期以及四大經營目標，計算有多少句子 GPT-4 認為是正面，多少句子認為是負面，再依此計算每一開會日期的 PN 指標：

$$PN \text{ 指標}_t = \text{正面句子數目}_t - \text{負面句子數目}_t$$

其中下標 t 為央行會議日期。同樣地，利用相同的文本，我們也透過財經情緒字典計算其 PN 指標。圖 1~圖 4 為不同經營目標下，依據 GPT-4 以及財經情緒字典所計算出的 PN 指標，其中藍色虛線為 GPT-4 所計算的結果，而紅色實線則是依據財經情緒字典所計算的 PN 指標。

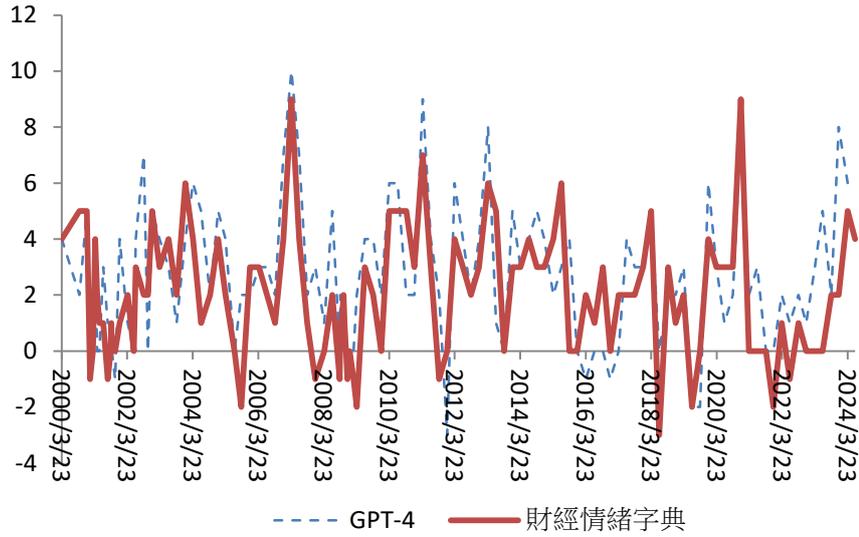


圖 1: 「促進金融穩定」之 PN 指標

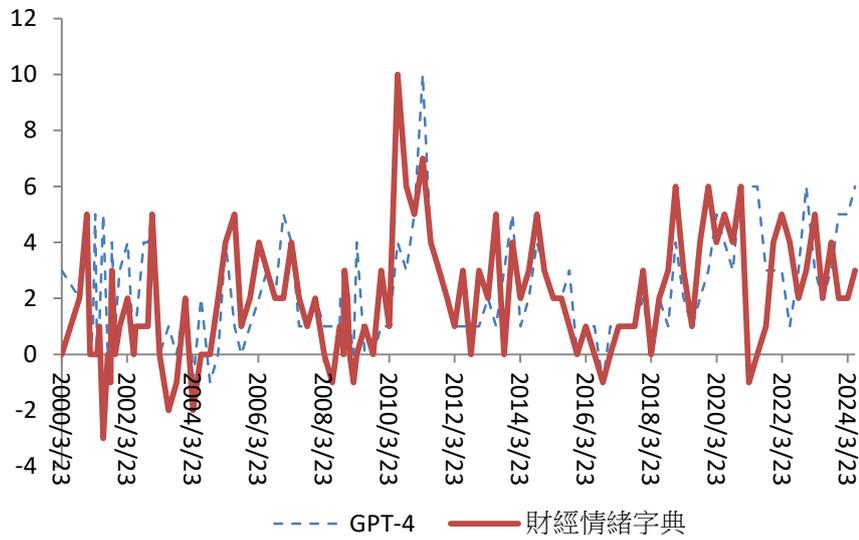


圖 2: 「健全銀行業務」之 PN 指標

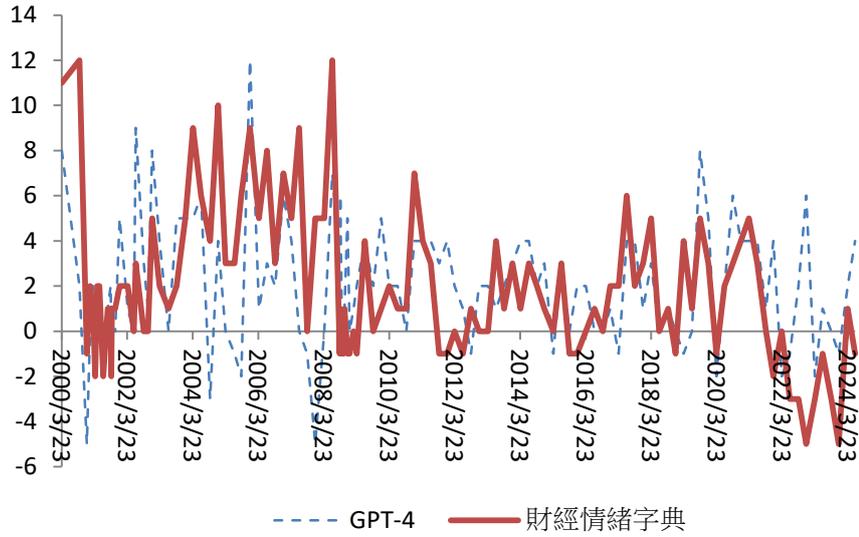


圖 3: 「維護對內及對外幣值之穩定」之 PN 指標

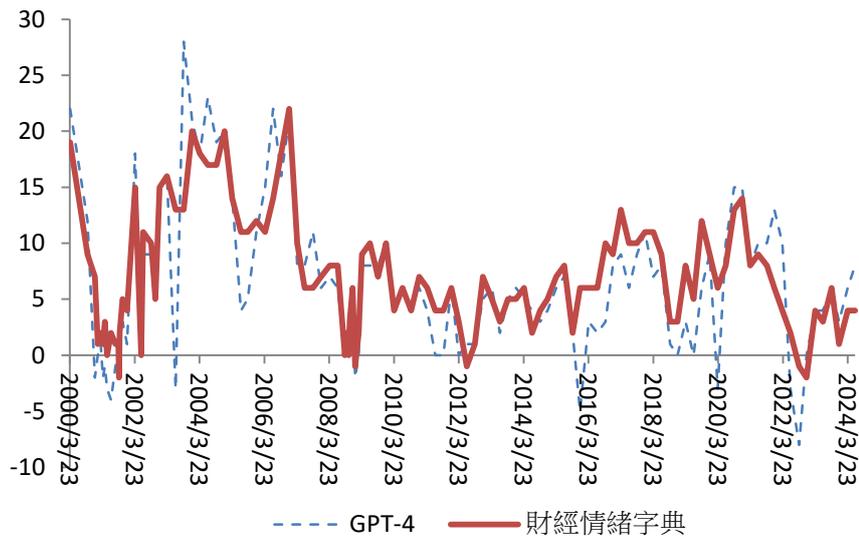


圖 4: 「協助經濟發展」之 PN 指標

從圖中不難發現，透過 GPT-4 與財經情緒字典所計算出的 PN 指標對某些經營目標的文字情緒刻畫有相似的走勢。例如圖 4 中對於「協助經濟發展」類別的文字內容，兩種方法所計算出的 PN 指標變化的趨勢一致；經計算這兩條數列的相關係數高達 0.84。類似的情況也可以在圖 1 「促進金融穩定」PN 指標走勢被觀察到，圖中兩條數列的相關係數值為 0.70。但若觀察圖 2 「健全銀行業務」以及圖 3 「維護對內及對外幣值之穩定」的 PN 走勢會發現，兩種方法所得到的結果其相關性並不太高，其相關係數值分別為 0.49 以及 0.41。¹⁰

直觀上，GPT-4 背後有大量資料以及演算法的支持，所以在很多分類任務（包含情緒評分）都有相當優異的表現，因此不難想像 GPT-4 的情緒評分數結果會比較正確。但獲得到這些結果也並非完全沒有代價。例如，我們需花費一些經費才能透過 API 方式進行大量文本的情緒評分，文本內容愈多，所花費的成本就愈高。此外，因為字符數目的限制，GPT-4 在執行任務時會暫時性終止，需等待一段期間才可以接續執行任務，因此完成任務的時間會因文本增加而變長。雖然這可以透過會員升級或是更換模型的方式克服，但仍需一定的費用。

反觀透過財經情緒字典來進行評分，不需花費太多時間就可以對大量的文本進行情緒評分，也不需額外費用，所以相較之下此方式仍有一定的優勢。特別是研究者想要快速對大量文本有一初步了解，或是分析的文本有資訊安全的考量時，透過情緒字典的協助可以很快地完成任務需求。當然，其代價可能是情緒的評分不夠精確。然而，從上述的實證結果會發現，目前所編製的情緒字典對於經濟發展以及金融穩定類別的央行新聞稿評分已有一定的水平，但對其它類別的情緒計分則有待加強。解決或克服此情況也不是難事，只要將這兩類的文本，透過 PMI 指標以及 GPT 的協助，將相關的情緒字詞加入目前的字典中，就可以改善目前的情況。換言之，維護或是增加字典的內容並不是一件困難的事。

¹⁰因實證結果會受到文本（如政府文稿或一般新聞）的差異而不同，故此處的重點不在於分類是否正確或評分是否合理，而是想突顯出透過字典方式來判斷文本情緒仍有一定的水準與價值。

5. 結論

透過文字探勘技術從文本中提取關鍵資訊，已成為近年來熱門的探討議題。此類資訊能補充或擴展數據資料的內涵，進而使後續的實證研究更為嚴謹。而文本的情緒分析則是這項技術常見的應用範例。由於財經新聞在用詞遣字上具有獨特性，若僅依賴一般情緒字眼進行情緒分析，往往會產生詞義上的偏誤。因此，Jiang et al. (2019)、姜富偉等人 (2021) 以及 Loughran and McDonald (2011) 特別編纂了中、英文財經情緒字典，供學術研究使用。從這些文章的引用次數以及發表於重要期刊的情況來看，編製字典有其重要的學術貢獻。

在臺灣，或因財經新聞報導取得不易，或因人力與時間上的限制，許多研究者常透過國外的字典直接翻譯（或簡轉繁體）方式來進行財經新聞的情緒分析，這樣的做法有相當大的改進空間。為此，本文嘗試利用長期的財經新聞文本，結合大量人力以及 PMI 與 LLMs 分類模型，編製一部適合臺灣財經新聞的情緒字典，希望透過相關分類模型的建立，能簡化並長期維護此情緒字典，從而填補國內財經領域在文字資訊上的缺口。經過一系列的編製過程，本文最終篩選出 9,890 個財經情緒字詞，並將其應用在央行新聞稿的情緒分析。同時，我們也利用 GPT-4 對央行新聞稿進行分類與情緒分析，從中比較 GPT-4 與財經情緒字典的評分結果。直觀上，GPT-4 受益於大量資料以及強大演算法的支持，在情緒分析的精確度上應優於字典式分析。然而，實證結果顯示，我們所編製的財經情緒字典在經濟發展以及金融穩定等議題上已有一定的水準。此外，若需從大量文本中快速獲得初步的情緒分析結果，利用情緒字典不僅能節省時間，還能降低成本，並且無資訊安全的考量。

如同 Loughran and McDonald 的情緒字典作法，一部好的字典需要長期不斷地精進修正，才能產生更大的效益。在未來，我們除了定期擴充新聞內容，重新計算每一個字典的“次數”與“篇數”資料外，也會透過本文的編製過程，新增重要的情緒字詞，並且希望能如 Loughran and

McDonald 的情緒字典一樣，提供更多面向（如不確定或是法律）的字詞。此外，我們也期望有更多的研究者能對此字典提出建議，進一步擴充詞彙，增加應用範圍，持續優化其功能。

參考文獻

1. 姜富偉，孟令超，唐國豪 (2021). 媒體文本情緒與股票回報預測. *經濟學 (季刊)*, **21 (4)**, 1323–1344.
2. Baker, Scott R, Bloom Nicholas and Steven J. Davis (2016), “Measuring Economic Policy Uncertainty,” *The Quarterly Journal of Economics*, **131 (4)**, 1593-1636.
3. Caldara, D., and Iacoviello, M. (2022). Measuring Geopolitical Risk. *American Economic Review*, **112 (4)**, 1194-1225.
4. Chen, C. C., Huang, H. H., & Chen, H. H. (2018). NTUSD-Fin: A Market Sentiment Dictionary for Financial Social Media Data Applications. *In Proceedings of the 1st Financial Narrative Processing Workshop*. 37-43.
5. Chen, Chung-Chi., Huang, Yu-Lieh, and Yang, Fang. (2024), “Semantics Matter: An Empirical Study on Economic Policy Uncertainty Index,” *International Review of Economics & Finance*, **89**, 1286-1302.
6. Huang Y.L. and C.M. Kuan (2021). Economic Prediction with the FOMC Minutes: An Application of Text Mining, *International Review of Economics and Finance*, **71**, 751-761.
7. Husted, L., Rogers, J., and Sun, B. (2020). Monetary Policy Uncertainty. *Journal of Monetary Economics*, **115**, 20-36.
8. Jiang, Fuwei, Joshua Lee, Xiumin Martin, and Guofu Zhou (2019). Manager Sentiment and Stock Returns, *Journal of Financial Economics*, **132**, 126-149.
9. Ku, Lun-Wei, Ho, Xiu-Wei and Chen, Hsin-His (2009), Opinion Mining and Relationship Discovery Using CopeOpi Opinion Analysis System, *Journal of American Society for Information Science and Technology*, **60**, 1486-1503.

10. Li, Q., Shan, H., Tang, Y., & Yao, V. (2024), Corporate Climate Risk: Measurements and Responses. *The Review of Financial Studies*, **37(6)**, 1778-1830.
11. Loughran, Tim and Bill McDonald (2011), When Is a Liability not a Liability? Textual Analysis, Dictionaries, and 10-Ks, *Journal of Finance*, **66**, 35-65.
12. Pendzel Sagi, Nir Lotan, Alon Zoizner and Einat Minkov (2024). Detecting Multidimensional Political Incivility on Social Media, arXiv:2305.14964
13. Thorsrud, L.A. (2020). Words are the New Numbers: A Newsy Coincident Index of the Business Cycle, *Journal of Business & Economic Statistics*, **38**, 393–409.