

台灣財經情緒字典與議題分類字 詞之編製與應用：大型語言模型之 協作與分類評估

黃裕烈*

國立清華大學計量財務與金融學系
國立台灣大學計量理論與應用研究中心 (CRETA)

關鍵詞：大型語言模型、文字探勘、財經情緒字典

Keyword：Dictionary, 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。作者感謝陳重吉研究員、吳俊毅副研究員、管中閔、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

本文旨在利用台灣的財經新聞文本，建立分類模型以篩選出台灣慣用的新聞字詞，並透過大型語言模型 (LLMs) 的協助，對這些新聞詞彙進行檢視、擴充以及情緒分類的確認。不同於既有文獻，本文還針對「景氣」、「貨幣政策」、「利率」等特定財經議題建構分類字詞，以填補這方面的不足。經過一系列的篩選，我們選出 7,814 個財經情緒詞彙；經過比對，有 1/4 以上的財經慣用語彙並未收錄於現有的情緒字典中。若以中央銀行公佈的會議新聞稿為樣本外的文本，以 LLMs 的分類結果為標準，比較不同字典的情緒分類精確度可以發現，本文的財經情緒字典在運算時間與成本上較 LLMs 更具優勢，而在分類的精確度雖不及 LLMs，確優於現有的情緒字典。此外，我們也發現，依本文所編製的情緒字典來建構情緒指標，該指標對許多台灣總體經濟變數有顯著的解釋能力與預測力，其表現均優於現有字典所建構的指標內容。最後，我們也觀察到，依據不同議題所建構的分類詞彙與情緒指標，能更細緻捕捉文本中的語意差異，進一步提升文字資料轉化為數據的內涵。

Abstract

This study aims to develop a classification framework based on Taiwanese financial news texts to identify commonly used financial terms and, with the assistance of large language models (LLMs), verify, expand, and assign sentiment labels to these terms. Unlike existing literature, this paper also constructs topic-specific lexicons for key economic themes such as business cycles, monetary policy, and interest rates to address gaps in conventional approaches. Through a multi-stage filtering process, we compile a sentiment dictionary containing 7,814 finance-related terms, over one-quarter of which are not covered in existing sentiment dictionaries. Using press releases from Taiwan's central bank as out-of-sample data and treating LLM-based sentiment classifications as the benchmark, we find that the proposed dictionary offers significant advantages in computational efficiency and cost. Although LLMs yield higher classification accuracy, our dictionary outperforms existing alternatives in both accuracy and practical applicability. Furthermore, sentiment indices constructed with our dictionary demonstrate superior explanatory and predictive power for various Taiwanese macroeconomic indicators. Finally, topic-specific sentiment lexicons further capture nuanced contextual meanings in financial texts, enhancing the transformation of unstructured text into structured economic insights.

1. 前言

透過文字資訊來進行相關的實證分析是近期相當熱門的一個議題，也是許多學者致力克服的問題（如 Baker et al., 2016 與 Thorsrud, 2020）。由於文字資訊具備非結構（unstructured）特性，因此在使用上需事先將文字轉換成數據資料（numerical data），才能配合計量模型與其它數據進行分析。通常研究者會利用分類模型（或稱分類器）來進行資訊的轉換，而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政策不確定性的問題。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；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；近期的大型語言模型（large language models, LLMs）也可被歸類此。從分類器的角度來看，這兩種方式各有優、缺點。例如透過 LLMs 來分類文本時其正確率較高，但處理大規模資料時（如歷年新聞報導），所花的時間與金錢成本也較可觀。相反地，採用字典方法進行大量文本分類，雖然犧牲了部分精確度，卻能以快速且低廉的特性取得成本效益上的優勢。換言之，這兩種方法相輔相成，能夠互補彼此的短處，為不同情境下的文本分類任務提供多樣化的解決路徑。

儘管傳統的字典方法具有上述效益，但其成敗關鍵仍繫於字典的品質。建構一部優良的字典極其不易，除了需有妥善的分類方法，還要依文本來源進行調整，並且需要不斷地更新字詞與長期維護，是一項繁瑣且艱辛的工作。¹例如，若分析的文本來源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 Harvard Psychosociological Dictionary（簡稱 Harvard 字典）所提供的情緒字詞來進行分析，應該特別編製一部財經情緒字典才能避免語意錯誤。主

¹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*）。

要是因為 Harvard 字典所提供的情緒分析比較適用於日常生活，因此像「財富」、「資源」、「失業人數」與「徵稅」等字詞都被歸類於某種正、負面情緒用字，但這些字詞在財經相關新聞中多半只是代表某一個名詞或數據變數，不必然帶有任何正、負面情緒意涵。同樣地，基於社群媒體資訊所編製的台大情緒字典 (Ku et al., 2009) 也跟 Harvard 字典類似，若是直接透過此部字典對財經相關新聞進行情緒判斷，也會因為網路社群語境的差異而產生語意偏差。此外，若針對台灣財經新聞進行分析時，也不宜直接採用 (但可參考) 國外財經字典的翻譯版本 (或簡轉繁體)。主要是因為國情不同，用字遣詞也會有所差異。例如在 Jiang et al. (2019) 與姜富偉等人 (2021) 的「簡體版財經字典」中，「愛崗敬業」、「學雷峰」、「樂顛顛」、「出籠」、「漏子」、「套期保值」與「新股弱勢」等均為財經相關的情緒字詞，但這些詞彙在國內並不常見，也不一定有相似的用法與意涵。

本文主要目的為透過文字探勘技術，以台灣歷年網路財經相關新聞為文本，除了編製一套適合台灣的財經情緒字典，還依據「景氣」、「貨幣政策」、「利率」、「匯率」、「物價」等議題進行字詞的分類。由於中文博大精深，相同的字詞在不同的議題中常有不同的意涵，因此依議題來分類字詞有其必要性。以「放緩」一詞為例，若是指經濟活動「放緩」，這在「景氣」的議題上屬偏負面或中性信號；但若是指通膨「放緩」，則對「物價」議題而言應是正面消息。又如主管機關「上調」存款準備率以及「上調」經濟成長率，兩者對整體經濟的解讀便不盡相同。事實上，這樣類似的字詞在財經新聞中相當多見，如「攀升」、「上漲」、「修正」、「下調」等字詞在描述「景氣」、「物價」、「利率」、「匯率」議題中都可能代表不同的意涵。因此，在解讀財經新聞時，需要考量議題與字詞之間的關聯性。然而，目前既有的中、英文字典在處理這類詞彙時，往往只是簡單地將其劃分成正面以及負面兩種情況，這樣硬性地區分在實際應用中極易造成誤判。

綜合而言，本文的主要特色在於：(1) 充分考量中文的語言特性，不僅編製一部財經情緒字典，還依不同議題進行相關字詞的分類，這是過往中、英文字典相關的文獻中並不多見。(2) 我們所採用的文本為台灣網路財經相關新聞，這能確保所建構的財經詞彙貼合並反映台灣財經領域常用的表達習慣。更重要的是，本研究的資料從 2003 年開始至 2024 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，能納入如「金融海嘯」、「量化寬鬆 (QE)」、「COVID-19」與「貿易戰」等字詞的影響性。(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。相對的，本文則嘗試透過 LLMs 的協助對字典進行維護 (容後說明)，因此後續的更新與長期維護相當容易。(4) 儘管 LLMs 的技術近年來發展快速，但在面對海量文本資料的處理需求時，傳統的字典分類方法在運算效率與成本效益上仍有比較優勢。因此，在諸多相關研究中 (如 Baker et al., 2016; Caldara and Iacoviello, 2022; Husted et al., 2022) 此方法依然受到青睞。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，並與 LLMs 相輔相成，發揮互補效益，彌補國內財經領域在文字資訊方面的缺口。

我們考量以下兩種模型來建構財經情緒字典內容，分別是利用 PMI (pointwise mutual information) 指數來篩選出常用詞彙，以及透過 LLMs 模型以及人工方式來檢視與維護字典內容。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。值得注意的是，我們不直接翻譯或轉換國外財經字典，也不利用 LLMs 模型直接生成 (generate) 字詞，最主要的原因是希望藉此可以找出國外字典並未收錄，但卻是台灣慣用的財金詞彙，以及避免 LLMs 因幻覺 (hallucination) 而產生不恰當的字詞。待篩選出常用詞彙後，我們再下達適當的提問指令 (prompt)，要求 LLMs 檢視這些情緒字詞是否合宜並舉例說明，以利後續的維護。

透過上述方式，我們最終篩選出 7,814 個財經情緒字詞以及數千個跟議題相關的分類字詞 (視不同議題而異)。相較於簡體版財經字典 (包含 9,228 個字詞) 以及台大情緒字典 (包含 23,729 個字詞)，我們所篩選出的財經情緒字詞較為精簡。與其它字典相比對後，我們發現有 4,332 個台灣慣用的財金詞彙 (如「除權息行情」、「景氣復甦」、「颱風尾」等) 並未收錄於簡體版財經字典中，而有 1,931 個常見財金詞彙並未納入台大情緒字典 (如「資金不足」、「低檔」、「停電」等)。由此可見，透過 PMI 模型與 LLMs 的配合確實可以篩選出台灣慣用的財金詞彙。我們也發現，從財經新聞的

角度來看，現有的字典中存在不少標記錯誤的情況；例如台大情緒字典中至少有 202 個字詞正、負字詞標記相反（例如將「漂白」記為正面，而「利益」標成負面），而簡體版財經字典中至少有 14 個字詞標記錯誤。另外也有許多字詞語意模稜兩可，需要視上、下文或額外訊息才能判斷，但確硬是被歸類成某一情況（如「高峰」一詞）；²類似這樣的字詞在台大情緒（簡體版財經）字典中至少有 475 (158) 個。最後，我們亦針對特定議題編製了相應的分類字詞，這些詞彙在不同議題下可能帶有相異的意涵，但均未曾在文獻中被系統性地整理或探討。

為了解財經情緒字典的優、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。首先，我們透過 LLMs 所提供的 API (application programming interface) 功能，以經濟學家的角色將文本中的句子進行情緒分類，再與本文所編製的財經情緒字典評分結果相比較。從運算時間的角度來看，利用 LLMs 進行分類所花的時間成本是字典方法的數百（千）倍。以 2000-10-05 的央行新聞稿內容（共 35 個句子）為例，LLMs 花了 8 分 34 秒（共 514 秒），但字典方式只花約 0.1 秒即可判斷完成，並且免費。³但就分類的精確度來說，無庸置疑，LLMs 的表現一定優於字典方式。如果以 LLMs 的分類結果為標準答案，計算各字典的評分結果會發現，本文所編製的財經情緒字典所得到的準度 (accuracy) 以及平均 F1-score 分別為 0.67 與 0.65，其表現均優於台大情緒字典的 0.48 與 0.42 以及簡體版財經字典的 0.52 與 0.48。換言之，相較於 LLMs，本文的財經情緒字典在運算時間與成本上具比較優勢，而在分類的精確度雖不及 LLMs，確優於現有的情緒字典。最後，我們還探討情緒字典的相關應用，並發現以本文所編製的情緒字典來建構情緒指標，該指標對許多台灣總體經濟變數有顯著的解釋能力與預測力，其表現均優於現有字典所建構的指標。

本文的內容安排如下：第 2 節介紹資料的處理方式，第 3 節為模型

² 「高峰」可能是指經濟活動達到高峰，也可能指失業率達到高峰，但該字詞確被分類成正面詞彙。當然，也有很多詞彙可能因國情不同、用法不同，或比較適用於日常生活中描述個人情緒，不應收錄於財經字典中。這些在前文中已有討論，此處便不多贅述。

³ LLMs 所花費的時間會受到模型大小、輸出結果的精準度、每日請求數目 (requests per day) 以及網路速度...等所影響。本文是利用 Gemini 2.5 pro-preview-05-06 為基礎進行情緒分類。而字典方法所花費的時間會受到電腦規格...等所影響，我們所用的 CPU 為 i7-14700。

簡介，包含 PMI 指標與提問指令，第 4 節為實證結果，最後一節為結論。

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，不可能將每天、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，以下為資料的處理方式。

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），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 2024-12-31 共 93 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

為建立所需的文本，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）的網路新聞內容。⁴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

⁴ 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。

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

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。因為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。

由於 PMI 模型在估算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。⁵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。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,⁶最後再經作者檢查標記內容, 供後續使用。

雖然人工標記所需的人力相當繁雜, 但若能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

⁵ 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, 另外我們也控制段落的長度以及挑選原則; 相關細節此處就不多贅述。

⁶ 人工標記資料的工作歷時相當長, 從 2020 年便開始執行。在此我們特別感謝黃偉德, 曾彥寧, 戴唯倫, 林昕宏, 張庭維, 李佳儒, 周玟妍, 李冠宇等研究人員的付出與貢獻。

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（簡稱 Master 財經字典）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 LLMs 的提問指令。首先，我們會利用人工標記的文本資訊，透過 PMI 指數來篩選出基本情緒字詞，再利用提問指令，擴充這些基本情緒字詞。接著，計算所有情緒字詞是否曾出現在台灣財經相關新聞中，刪除出現頻率過低的字詞。最後，透過 LLMs 模型進行檢測，以編製成財經情緒字典，以下是處理過程的說明。

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

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。

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：

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

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 擴充情緒字詞

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 或 Gemini 2.5 Pro 這類大型語言模型的協助，嘗試擴充情緒字詞。為了讓 LLMs 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：

“你是一位熟悉台灣總體經濟情況以及繁體中文的經濟學家，每天都會閱讀大量的財經新聞報導。你的任務是編製台灣的財經情緒字詞，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。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，在台灣並不常見，我們還會計算所有情緒字詞是否曾出現在台灣財經相關新聞中。我們以 2003-05-01 至 2024-12-31 每日、每則財經新聞為文本基礎（共 2,281,382 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）的詞彙共 9,890 個。換言之，篩選出的情緒字詞都是台灣慣用且曾出現在財經新聞中。

3.3 情緒字詞的檢測

基於模型的快速發展，LLMs 不僅可以完成分類任務，還可以利用其推理 (reasoning) 與思考 (thinking) 能力，協助研究者檢測或判斷分類結果。據此，我們透過 Gemini 2.5 Pro 的協助，分析 3.2 節所選出的情緒字詞，並且依據 Wang et al. (2020) 的建議，要求模型舉例說明情緒字詞的應用場景，藉以自我判斷 (self-consistency) 分類的結果是否合宜。若認為某一字詞語意模稜兩可（例如「高峰」），需要視上、下文或額外訊息才能判斷其正、負面情緒，此時，我們也會要求模型，需各舉一例來協助理解。以下是所使用的提問指令：

“你是一位熟悉台灣總體經濟情況以及繁體中文的經濟學者，正在編製一套專屬於台灣的「財務與經濟新聞情感字典」。附上 CSV 檔，檔案中第一欄位是詞彙，第二欄位是情緒分類 (True 為正面，False 為負面，Ambiguous 代表某些情境下為正面，某些情境為負面)。以下是任務需求：1. 請依據第二欄位的內容，依據以下步驟思考情緒分類的結果：(1) 詞彙是否常見於台灣的財經新聞中？(2) 詞彙是否與下列主題相關：財務、經濟、股市、匯率、債市、利率、物價、產業、市場、總體經濟以及其它跟財經相關的新聞等？(3) 該詞彙在語境中通常表現正面或負面情緒？2. 分析後，再舉一些例子說明它是正面，負面（或其它）情況。3. 舉例後再一步一步思考與分析詞彙，並做出最後的答案，回覆的答案可以跟之前不同，可以是 True, False, Ambiguous。4. 不論答案是正面或負面，都一定要舉例子說明。若答案是 Ambiguous，則需要正面或負面各舉一個例子。”

而在檢測各議題相關的情緒字詞時，我們也會採用類似的提問指令詢問 Gemini 2.5 Pro，此處便不多贅述。

透過提問指令，Gemini 2.5 Pro 便依序分析 9,890 個情緒字詞，過程中我們也會記錄其分析內容。以「備受肯定」為例，其分析內容摘錄如下：

備受肯定 (*Highly recognized/affirmed*):

Frequency/Domain: Common. Used for policies, companies, individuals, products. Relevance: Relevant, often relates to reputation, trust, or success in the market/industry. Sentiment: Positive. Being affirmed implies success, good reputation, or effectiveness. Suitable: True.

JSON: {"word": "備受肯定", "analysis": "指某個對象（如政策、企業、產品、策略）因其成效、品質或重要性而獲得廣泛的認可或讚揚。在財經語境中，通常與成功、有效或有價值的事物相關聯。例句：公司所進行的投資案備受肯定。", "suitable": true}

不難發現 Gemini 2.5 Pro 會嘗試將字詞翻譯成英文，再撰寫英文的分析內容，思考後生成例句，產生中文的判斷結果，並直接利用這些結果來當成字詞是否適宜的標準，以節省人力。經過 LLMs 的檢測，最終我們篩選出 7,814 個財經情緒字詞以及數千個跟議題相關的詞彙。

4. 實證結果

4.1 財經情緒字典

透過上述的編製過程，我們篩選出歷年來台灣常見的財經情緒字詞，包含正（負）面 3,467 (4,347) 個，共 7,814 個字詞。表 2 列舉出 20 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。透過這些“次數”以及“篇數”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。為了方便後續維護以及使用，我們將

表 2: 正負面財經情緒字詞摘錄

正面字詞	次數	篇數	負面字詞	次數	篇數
最具影響力	1,430	1,273	不確定性增加	751	725
除權息行情	2,161	1,408	麻煩製造者	693	556
印象深刻的	1,068	1,049	低經濟成長	237	223
贏得勝利	619	595	難以彌補	349	337
膾炙人口	804	772	避重就輕	2,205	1,994
獨一無二	4,331	3,947	緩不濟急	1,944	1,771
歷史新高	119,171	86,738	緊張局勢	3,752	3,264
非凡的	706	681	懲罰性	2,368	1,836
盈餘	176,818	87,771	擔心	120,329	99,825

註: 文本計數來源為 2003-05-01~2024-12-31 各大網路財經新聞, 共 2,281,382 則新聞。

此字典內容放在網路上供大家下載,⁸並且依長詞優先以及次數頻率由高至低排序。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文本中所出現過的情緒字詞次數。

由於我們是基於台灣財經相關新聞為文本來源,透過 PMI 指標篩選出台灣慣用的財經情緒字詞,因此這些字詞不一定會被收錄於簡體版財經字典以及台大情緒字典中。經過比對,共有 4,332 (1,931) 個台灣慣用財經詞彙並未收錄於簡體版財經(台大情緒)字典中,占有字詞數的 55% (25%),比例相當高。表 3 列舉出 30 個具代表性,但確未被既有字典所收錄的情緒字詞。表中可以發現,由於簡體版財經字典並不會用「景氣」一詞來描述經濟成長,而台大情緒字典中也不會特別關注財經相關字詞,因此第一行(column)與經濟景氣相關的詞彙便未被收錄。同理,台灣財經新聞中常見的公司財報字詞(如表中第 2 行)、描述股市的慣用語彙(如表中第 3 行)以及近期被受關注的事件(如「COVID-19」、「地緣政治」、「貿易戰」等)均被忽略。由此可見,透過 PMI 模型以及第 3 節所介紹的字詞篩選過程,確實可以彌補目前字典的不足,挑選出台灣財經

⁸ 下載網址為: <https://remi.ctmnhu.com/portfolio-archive/financial-dictionary/>

表 3: 未收錄於既有字典的代表性字詞摘錄

景氣升溫	價值投資	價量齊揚	地緣政治	貿易戰	獨佔
緩步復甦	財測樂觀	漲停鎖死	經濟疲弱	灰犀牛	填權
景氣復甦	庫存回補	高殖利率	利率倒掛	黑天鵝	填息
強勁擴張	轉虧為盈	軋空行情	金融海嘯	景氣差	作多
數據好轉	數位轉型	慶祝行情	COVID-19	正成長	買超

註: 共有 4,332 (1,931) 個台灣慣用財金詞彙並未收錄於簡體版財經 (台大情緒) 字典中。

新聞中常用的詞彙，並且可以避免因直接翻譯或轉換國外財經字典所造成的語意偏差，也減少直接利用 LLMs 生成字詞而產生不恰當的內容。⁹另外，我們也比對出既有字典中，標記錯誤以及語意模稜兩可的字詞，在前言中已有一些討論，故此處便不多贅述。

在編製財經情緒字典時，我們主要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所公佈的文字資訊。為了解各部字典的表現，我們以央行新聞稿中 4,110 個句子為樣本外資訊，利用 Gemini 2.5 Pro API 以及字典方式對每一個句子進行情緒分類。由於 Gemini 2.5 Pro 在文字分類任務的表現相當優異，因此我們以該模型所判斷的結果為標準答案，藉以計算不同字典的分類結果。表 4 列出分類結果的混淆矩陣 (confusion matrix)，其中“實際”代表 Gemini 2.5 Pro 所分類的情況，而“預測”則為各字典的分類結果。舉例來說，就表中上半部而言，Gemini 2.5 Pro 判斷為正面的句子共有 1,379 句 (1,091+103+184)，給定這 1,379 個句子，本文的情緒字典判定為正面的情況有 1,091 句，負面有 103 句，而其它 (即無情緒分類或中性) 則有 184 句。從表中不難發現，本文的情緒字典除了“實際”為正面的情況略遜於台大情緒字典外，其它各種情況均優於其它字典。若計算各評分指標，本文的情緒字典所估算出的指標分數為準度 0.67、精度 (precision) 0.66、召回度 (recall) 0.66 與 F1-score 0.65，而簡體版財經 (台大情緒) 字典所估算出的指標分別為 0.52 (0.48)、0.50 (0.53)、0.49 (0.46) 與 0.48 (0.42)。換言之，從各項

⁹ 我們曾嘗試以 RAG (retrieval-augmented generation) 方式直接要求 LLMs 生成 500 個字詞，但成效並不理想。一開始 LLMs 所生成的字典符合預期，但後來常會因「幻覺」或為了硬湊足 500 個字詞而產生不恰當的內容。

表 4: 各字典分類結果的混淆矩陣

本文編製財經情緒字典				
實際 \ 預測	正面	負面	其它	
正面	1091	103	184	
負面	158	481	191	
其它	483	246	1173	

簡體版字典				台大情緒字典				
實際 \ 預測	正面	負面	其它	實際 \ 預測	正面	負面	其它	
正面	954	110	314	正面	1184	92	102	
負面	313	237	280	負面	524	165	141	
其它	655	269	978	其它	1184	99	619	

註: “實際”代表 Gemini 2.5 Pro 所分類的結果，而“預測”為各字典的分類情況。

評分指標來看，本文的情緒字典的表現均優於既有的字典。

雖然上述的評比中我們以 Gemini 2.5 Pro 的答案為基準，認定字典的分類表現一定比 LLMs 差，但獲得這些標準答案所花的時間成本確遠遠高於字典分類方式。由於 LLMs 對每日的請求 (requests) 數目會設定上限，超過額度時就必需等待，所以我們花了 2.5 個工作天才能完成所有句子的分類，共花費約 7.5 美元。即便不論受額度限制所造成的影響，為求答案精準，LLMs 平均也會花約 10 秒以上來分析一個句子。但相反地，不論是那一部字典，透過此方式來分類“所有”的句子也不會超過 2 秒，並且完全免費。換言之，LLMs 與字典這兩種方法互有優劣，應相輔相成，才能夠互補彼此的短處。

4.2 以議題為基準的分類字詞

除了上一小節所介紹的財經情緒字典之外，本文還針對不同的議題編製相關的分類字詞，以考量同一字詞在描述不同議題時可能出現相異的敘述情況。表 5 摘錄一些分類字詞內容，其中上半部 (下半部) 字詞標記為 TRUE (FALSE)，僅代表在這些情況下，常常有利於經濟的成長。舉例來說，當貨幣政策為寬鬆、調低利率、台幣走軟、CPI 年增率趨緩、物價下

表 5: 議題相關的字詞摘錄

「景氣」	「貨幣政策」	「利率」	「匯率」	「物價」
優於預期	政策利率調降	利率維持低檔	拖累台幣匯率	CPI 年增率趨緩
經濟復甦	上調通膨目標	調低重貼現率	台幣失守	調降 CPI
持續增溫	量化寬鬆	負利率政策	美元走升	抑制通膨
景氣回升	降息循環	調低利率	台幣續貶	通膨下調
經濟擴張	寬鬆銀根	連續降息	台幣走軟	控制通膨
上調	擴張	利率下滑	美金飆漲	形成通縮
走強	鴿派	降息循環	股匯雙殺	陷入通縮
復甦	寬鬆	相繼降息	阻貶	抗通膨
反彈	減息	鴿派	降息	低通膨
增溫	降準	降息	雙跌	反通膨
不確定性增加	上調基準利率	調升重貼現率	外資大舉匯入	上調 CPI 年增率
經濟衰退	實施附賣回	預防性升息	台幣走強	CPI 持續回升
下滑風險	緊縮貨幣	貼現率走升	拉抬台幣	CPI 升至
景氣不振	縮減購債	利率攀升	股匯雙升	推升通膨
疫情衝擊	利率上調	加碼升息	熱錢流入	通縮降溫
降溫	鷹派	升息政策	美金下殺	擺脫通縮
減緩	縮表	利率上調	美金攪破	上調物價
低迷	緊縮	高利率	拋匯	助長通膨
衰退	升息	鷹派	阻升	高通膨
疲軟	升準	升息	雙升	大通膨

降時常有利於經濟發展，因此標記為 TRUE。如同前言，「上調」經濟成長以及「上調」基準利率所描述的經濟狀況不盡相同，不可一視同仁。因此，依不同議題來分類字詞有其必要性，也是既有中、英文字典中尚需考量的部份，而本文所提供的分類字詞內容恰可補充不足之處。

為了解這些字詞的分類表現，我們仍以央行新聞為樣本外資料，依議題的不同下達不同的提問指令，並以 Gemini 2.5 Pro 的回答為標準，計算在不同議題下，以字典方式進行分類的準度、精度、召回度與 F1-score，表 6 列出這些結果。從表中各項指標可以發現，除「匯率」以外，其它議題的平均表現差異不大，F1-score 約落在 0.60~0.67 之間。雖然目前無從比較這些結果的優異程度，但我們認為各議題的分類字詞仍有改善的空間，有待未來的改進。

表 6: 各議題之分類結果

	「景氣」	「貨幣政策」	「利率」	「匯率」	「物價」
準度	0.64	0.67	0.67	0.70	0.69
精度	0.68	0.75	0.68	0.53	0.70
召回度	0.63	0.58	0.66	0.57	0.66
F1-score	0.64	0.60	0.67	0.53	0.67

註: 準度指 accuracy, 精度指 precision 而召回度指 recall 指標。

4.3 情緒字典的相關應用

情緒字典的應用相當廣泛，我們從一些案例中便可看出端倪。以 Loughran and McDonald (2011) 所建構的 Master 新聞字典為例，它是一部引用次數相當高的財經情緒字典。依據 google scholar 的資料顯示，該文至 2024 年底已有 6,500 筆以上的引用次數。除此之外，在許多與文字探勘相關的財經文獻中都可以發現，不論一開始所探討的主題為何，所用的技術多麼複雜，最後的實證分析都會透過字典方法來建立情緒指標，將文字資訊轉換成數據進行分析 (例如 Li, et al., 2024)。若該指標可以提供研究者額外的訊息並獲得顯著的實證結果，我們便認為此文字資訊與情緒指標是一個重要的財經變數，其資訊含量無法從既有的數據資料中獲取。除了情緒指標的編製以外，中文的財經情緒字典還有一些應用值得分享。首先，因文字資料的特性，中文的文本常需要經過適當的斷詞 (segmentation) 才容易進行後續的分析。而財經情緒字典中的字詞恰可提供這方面的資訊，提高斷詞套件 (如 Jieba) 的精確度。另外，也可以透過字詞嵌入 (word embedding) 或是 RAG 的方式，將這些情緒字詞的訊息提供給 LLMs 參考，以了解財經字詞的慣用方式，這對後續的輸出有很大的助益。

接著，我們僅以情緒指標的建構來說明其可能的應用情況。首先，我們以每季央行所發佈的新聞稿為文本，利用字典方式對文本中每一個句子進行情緒分類 (即 4.1 與 4.2 節的結果)，計算以下的情緒指標：

$$Tone_t = \frac{\text{正面句子數目}_t - \text{負面句子數目}_t}{\text{正面句子數目}_t + \text{負面句子數目}_t}, \quad (1)$$

其中 t 為央行新聞稿公佈的季度，而正面句子數目, (負面句子數目,) 為當季新聞稿中利用字典方式標記為正面 TRUE (負面 FALSE) 的句子個數。接著，我們參考黃裕烈等人 (2021) 的做法，以線性迴歸模型：

$$y_t = \alpha_0 + \sum_{i=1}^n \alpha_i y_{t-i} + \beta \times \text{Tone}_{t-j} + \varepsilon_t, \quad n \leq 4, \quad (2)$$

來了解 Tone 指標對台灣總體經濟變數的解釋能力 ($j=0$) 以及預測能力 ($j=1$)，其中，參數 n 是以 SIC (Schwarz information criterion) 選取準則來決定，而 y_t 為台灣重要的總體變數，包含實質 GPD 成長率 (經季節調整)，非農就業人數成長率，失業人數成長率，失業率 (經季節調整)，出口額成長率，進口額成長率，新增出口訂單成長率，CPI 成長率 (經季節調整)，美元兌新台幣匯率 (簡稱美元匯率)，M1A 成長率，M1B 成長率以及 M2 成長率等 12 個變數，資料來源為台灣經濟新報。若總體變數為月資料，我們會計算每季的平均以符合迴歸模型的要求。若模型中 β 顯著，代表除了透過 AR (auto-regression) 模型來補捉總體時間數列的持續特性 (persistence) 外，還有其它資訊可能受到 Tone 指標所影響。除 Tone 指標外，我們也計算 PN (positive - negative) 指標，其定義為式 (1) 的分子部份。因這兩個指標的實證結果相似，故後續只呈現式 (2) 的結果。

4.3.1 情緒指標的解釋能力

表 7 列舉出式 (2) 的迴歸結果。從表中不難看到，依據本文 4.1 節所編製的字典來建構 Tone 指標，其對實質 GPD 成長率、非農就業人數成長率等 8 個變數有顯著的解釋能力，並且模型的參數估計值 β 也符合經濟直覺。例如當央行對當期的經濟有比較正面的看法時，新聞稿中便會撰寫比較多的正面情緒語句，Tone 指標的數值會增加，並與當期的實質 GPD 成長率與非農就業人數成長率有正向關係，而跟失業率等變數有負向關聯性，並且台幣在當期可能有升值 (美元匯率下降) 的情況。

為比較不同字典與 LLMs 方法的分類情況，我們依據 4.1 節中各種字典與 Gemini Pro 2.5 的分類結果來建構 Tone 指標，並估算式 (2) 的迴

表 7: Tone 情緒指標的解釋能力

總體變數	β	Std.	t -stat	p -value	n
實質 GPD 成長率	8.9215	2.5494	3.4995	0.0007 ***	4
非農就業成長率	0.3212	0.1506	2.1328	0.0353 *	4
失業人數成長率	-0.1612	0.0713	-2.2609	0.0257 *	1
失業率	-0.1433	0.0641	-2.2361	0.0281 *	1
出口額成長率	0.0071	0.0127	0.5614	0.5757	1
進口額成長率	0.0016	0.0006	2.7055	0.0080 **	1
出口訂單成長率	0.0593	0.0284	2.0890	0.0420 *	1
CPI 成長率	0.0009	0.0006	1.4754	0.1430	1
美元匯率	-0.7111	0.2018	-3.5230	0.0006 ***	2
M1A 成長率	0.0088	0.0040	2.1824	0.0312 *	1
M1B 成長率	-0.0005	0.0027	-0.1827	0.8554	1
M2 成長率	0.0004	0.0011	0.3941	0.6943	1

註: Std. 為參數標準差, t -stat 為 t -統計量, n 為式 (1) 中的 AR 參數。符號 *, **, *** 分別代表 5%, 1% 以及 0.1% 以下的顯著水準。

歸模型。為精簡篇幅，我們省略各個模型所估算出的標準差， t -統計量等結果，僅將 β 估計值列於表 8，並以符號 †, *, **, *** 代表 10%, 5%, 1% 以及 0.1% 以下的顯著水準。例如表 8 第一行與表 7 第一行的數值一樣，都代表依據本文所編製的字典來進行分類後所估算的參數 β 值，而表中其它各行數值則代表利用不同的字典與 Gemini Pro 2.5 所估算出的 β 值。從表 8 的結果可以觀察到，以 Gemini Pro 2.5 進行分類所建構的 Tone 指標，對實質 GPD 成長率等 9 個變數有顯著的解釋能力，略優於表 7 所呈現的結果。而依據簡體版（台大情緒）字典所建構的指標則只對 4 (2) 個總體變數有顯著的解釋力，並且一些 β 參數估計值正、負符號也有違經濟直覺。故整體而言，他們的表現不及本文所編製的字典，

最後，我們也根據 4.2 節的分類結果來建構 Tone 指標，以了解不同議題分類字詞的應用表現。同樣地，為求精簡，表 9 僅列出 β 估計值以及標示其顯著情況。表中不難發現，依「景氣」議題所建構的 Tone 指標對實質 GPD 成長率等 9 個變數有顯著的解釋力，表現最優。而「匯率」議題下的 Tone 指標只對 CPI 成長率有顯著的解釋力，表現較差。此

表 8: 依不同方式所建構的情緒指標之解釋能力

總體變數	本文字典	台大情緒字典	簡體版字典	Gemini Pro
實質 GPD 成長率	8.9215 ***	-2.9824	6.6708 †	0.2283 †
非農就業成長率	0.3212 *	0.1530	0.5478 *	0.0203 **
失業人口成長率	-0.1612 *	-0.0522	-0.1199	-0.0059 *
失業率	-0.1433 *	0.0091	-0.2655 **	-0.0023
出口額成長率	0.0071	-0.0110	-0.0022	-0.0208 †
進口額成長率	0.0016 **	-0.0213	0.0120	0.0041 **
出口訂單成長率	0.0593 *	0.0277	0.0458	0.0005 *
CPI 成長率	0.0009	0.0006	0.0002	0.0001
美元匯率	-0.7111 ***	-0.0624	-0.6320 *	0.0001 *
M1A 成長率	0.0088 *	-0.0116 *	0.0005	0.0000
M1B 成長率	-0.0005	-0.0085 *	-0.0055	0.0010 †
M2 成長率	0.0004	-0.0016	-0.0011	0.0020 **

註: †, *, **, *** 分別代表 10%, 5%, 1% 與 0.1% 以下的顯著水準, 表中數值為式 (2) 中的 β 估計值。

外, 當央行新聞稿內容對於「景氣」議題有較多正面的描述時, 該議題的 Tone 指標數值會增加, 因此對於當期的實質 GPD 成長率與非農就業人數成長率有正向關聯性, 而跟失業率等變數有負向關係。同理, 當新聞稿中對「貨幣政策」議題有較多寬鬆字句, 或是對「利率」議題有較多調降的描述情況時, 這些議題的 Tone 指標數值會增加, 並跟當期的 M1B 貨幣供給成長率有正向關係。另外, 若對「物價」議題有較多下降或是趨緩的敘述時, 該議題的 Tone 指標也會增加, 並與當期的 CPI 成長率有負向的關聯性; 上述這些結果均符合經濟直觀。當然, 表中也會有一些不易解釋的結果, 例如「貨幣政策」或「利率」議題的 Tone 指標對失業率之間的負相關, 以及「匯率」議題對 CPI 成長率的負向關聯等。

綜合而言, 上述的實證結果告訴我們: (1) 央行新聞稿的文字內容對於了解當前台灣的總體經濟情況是一份重要的資訊。(2) 透過適當的字典分類方式來建構情緒指標, 可以將文字資訊轉換成數據, 而這些指標的資訊含量相當重要, 無法從既有的數據資料庫中獲取。(3) 即便控制時間數列的持續特性, 這些情緒指標仍對許多台灣總體經濟變數有顯著的解釋能力。換言之, 本文所建構的情緒指標是一個重要的解釋變數, 若忽略這些重要的控制變數, 在進行迴歸分析時恐會得到偏誤的實證結果。

表 9: 不同議題所建構的情緒指標之解釋能力

總體變數	「景氣」	「貨幣政策」	「利率」	「匯率」	「物價」
實質 GPD 成長率	6.8638 ***	0.1575	0.1179	1.2791	-1.1489
非農就業成長率	0.2894 **	-0.0737	-0.1547 *	-0.2731	-0.0234
失業人口成長率	-0.1938 ***	0.0810 **	0.1221 ***	0.1152	0.0474 †
失業率	-0.1579 ***	0.0533 *	0.0977 ***	0.0511	0.0091
出口額成長率	0.0158 *	0.0007	-0.0008	-0.0017	-0.0068
進口額成長率	0.0247 **	0.0004	-0.0009	0.0150	-0.0120 *
出口訂單成長率	0.0159	0.0086	0.0199 †	-0.0235	-0.0013
CPI 成長率	0.0015 ***	-0.0007 *	-0.0006 *	-0.0018 *	-0.0008 **
美元匯率	-0.4384 ***	0.0720	0.1509 †	0.2201	0.2219 **
M1A 成長率	0.0025	0.0020	0.0002	-0.0014	-0.0009
M1B 成長率	-0.0030 *	0.0026 *	0.0020 *	0.0031	0.0008
M2 成長率	-0.0006	0.0005	0.0004	0.0017	0.0000

註: †, *, **, *** 代表 10%, 5%, 1% 與 0.1% 以下的顯著水準, 表中數值為式 (2) 中的 β 估計值。

4.3.2 情緒指標的預測能力

除了上一小節的解釋力外, 我們也計算落後一期的 Tone 指標 (即 $Tone_{t-1}$), 分析該指標是否對台灣總體經濟變數有預測能力, 表 10 列舉出式 (2) 的迴歸結果。從表中可以觀察到, 依據本文 4.1 節所編製的字典來建構 Tone 指標, 其對非農就業成長率、失業人口成長率等 5 變數有顯著的預測能力。若是依據簡體版 (台大情緒) 字典所建構的指標, 則只對失業率與 M1B 成長率 (M1A 與 M1B 成長率) 等 2 個總體變數有顯著的預測力, 其預測表現不如表 10。為求精簡, 我們便不列舉出簡體版 (台大情緒) 字典的迴歸分析結果。

表 11 則是列出各種議題下 Tone 指標的預測結果; 同樣地, 表中僅列出 β 估計值以及標示其顯著情況。從表中可以看到, 依「景氣」議題所建構的 Tone 指標對失業人口成長率等 4 個變數有顯著的預測力, 「匯率」議題下的 Tone 指標只對 CPI 成長率有顯著的預測力, 而其它議題下的 Tone 指標則分別對 3 個總體變數有預測能力。上述這些結果都可以說明, 透過情緒字典來編製指標確實有其實用性以及應用性, 而建構一部合適的財經情緒字典是一項基礎且重要的工作。

表 10: Tone 情緒指標的預測能力

總體變數		Std.	<i>t</i> -stat	<i>p</i> -value	<i>n</i>
實質 GPD 成長率	0.0490	0.1132	0.4331	0.6658	4
非農就業成長率	0.0209	0.0063	3.3170	0.0013 **	4
失業人口成長率	-0.0055	0.0025	-2.1402	0.0345 *	1
失業率	-0.0043	0.0026	-1.6301	0.1069	1
出口額成長率	0.0004	0.0005	0.8106	0.4194	1
進口額成長率	0.0008	0.0006	1.2499	0.2141	1
出口訂單成長率	-0.0011	0.0017	-0.6729	0.5042	1
CPI 成長率	0.0000	0.0000	1.7004	0.0919 †	1
美元匯率	-0.0155	0.0078	-1.9871	0.0486 *	2
M1A 成長率	0.0000	0.0002	-0.0538	0.9572	1
M1B 成長率	-0.0002	0.0001	-2.4591	0.0155 *	1
M2 成長率	0.0000	0.0000	-1.0438	0.2989	1

註: †, *, **, *** 代表 10%, 5%, 1% 與 0.1% 以下的顯著水準, 其他符號同表 7。

5. 結論

透過文字探勘技術從財經新聞或央行新聞稿中提取關鍵資訊, 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。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。鑑於現有經情緒字典在分析台灣財經文本時, 常因語境、文化及用字遣詞的差異而有所不足, 故本文致力於編製一套專屬於台灣、並能反映在地慣用語彙的財經情緒字典。據此, 我們嘗試利用長期的台灣財經新聞文本, 結合人力以及 PMI 與 LLMs 等分類模型, 編製一部適合台灣財經新聞用詞的情緒字典。不同於既有文獻, 我們還針對「景氣」、「貨幣政策」、「利率」等特定財經議題建構分類字詞, 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。

表 11: 不同議題所建構的情緒指標之預測能力

總體變數	「景氣」	「貨幣政策」	「利率」	「匯率」	「物價」
實質 GPD 成長率	1.9585	0.3297	1.5313	1.2791	0.4988
非農就業成長率	0.1556	0.0078	0.0193	-0.2731	-0.0916
失業人口成長率	-0.1327 ***	0.0105	0.0961 ***	0.1152	0.0098
失業率	-0.0975 *	0.0163 *	0.0857 **	0.0511	-0.0231
出口額成長率	0.0061	0.0017	-0.0030	-0.0017	-0.0129
進口額成長率	0.0109	0.0014	-0.0068	0.0150	-0.0152
出口訂單成長率	-0.0779 ***	-0.0017	0.0039	-0.0235	0.0621 †
CPI 成長率	0.0009 *	-0.0001	-0.0008 **	-0.0018 *	-0.0012 *
美元匯率	-0.1223	-0.0110	0.0280	0.2201	0.3350 *
M1A 成長率	0.0006	0.0012 **	0.0020	-0.0014	0.0023
M1B 成長率	-0.0024	0.0008 **	0.0019 †	0.0031	0.0010
M2 成長率	-0.0007	0.0000	0.0001	0.0017	0.0004

註: †, *, **, *** 代表 10%, 5%, 1% 與 0.1% 以下的顯著水準, 表中數值為式 (2) 中的 β 估計值。

經過一系列的實證分析後發現, 本文所編製的字典不僅在分類準確度上優於現有字典, 透過此字典所建構的情緒指標對台灣多項總體經濟變數 (包含 GDP 成長率、就業、失業與匯率等核心變數) 具有顯著的解釋與預測能力, 其表現均優於現有字典。此外, 我們也發現, 依不同議題所建構的分類詞彙與情緒指標, 能更細緻捕捉文本中的語意差異, 進一步提升文字資料轉化為數據的內涵。整體而言, 本文不僅提供一套兼具在地性與可維護性的財經情緒字典, 也展示文字探勘與人工智慧技術在財經實證研究中的應用潛力。而在後續的字典維護工作上, 我們會利用 LLMs 進行文字段落的標記, 以取代人力。接著, 透過 PMI 模型的字詞篩選以及 LLMs 的擴充與檢測確認, 形成一套工作流程。未來, 我們還會嘗試利用既有資料來訓練 NN (如 BERT) 模型, 以提升字典方法在分類文本的精準度 (但仍無法替代 PMI 篩選出慣用字詞)。另外, 若還能結合更多語料來源、進一步細緻化分類方法, 並與最新語言模型技術整合, 將有助於強化財經文字資訊的應用深度, 並為相關研究提供更具彈性與解釋力的工具。

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. 黃裕烈，葉錦徽與陳重吉 (2021), “臺灣經濟政策不確定性指標之建構與分析,” *經濟論文叢刊*, **49 (2)**, 307-334.
3. Baker, Scott R, Bloom Nicholas and Steven J. Davis (2016), “Measuring Economic Policy Uncertainty,” *The Quarterly Journal of Economics*, **131 (4)**, 1593-1636.
4. Caldara, D., and Iacoviello, M. (2022). Measuring Geopolitical Risk. *American Economic Review*, **112 (4)**, 1194-1225.
5. 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.
6. 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.
7. 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.
8. Husted, L., Rogers, J., and Sun, B. (2020). Monetary Policy Uncertainty. *Journal of Monetary Economics*, **115**, 20-36.
9. Jiang, Fuwei, Joshua Lee, Xiumin Martin, and Guofu Zhou (2019). Manager Sentiment and Stock Returns, *Journal of Financial Economics*, **132**, 126-149.
10. 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.

11. Li, Q., Shan, H., Tang, Y., & Yao, V. (2024), Corporate Climate Risk: Measurements and Responses. *The Review of Financial Studies*, **37(6)**, 1778-1830.
12. 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.
13. Pendzel Sagi, Nir Lotan, Alon Zoizner and Einat Minkov (2024). Detecting Multidimensional Political Incivility on Social Media, arXiv:2305.14964
14. 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.
15. Wang, Xuezhi, Jason Wei, Dale Schuurmans, Quoc Le, Ed Chi, Sharan Narang, Aakansha Chowdhery and Denny Zhou, “Self-Consistency Improves Chain of Thought Reasoning in Language Models,” *The Eleventh International Conference on Learning Representations*.