

國立臺灣大學管理學院會計學研究所



碩士論文

Graduate Institute of Accounting

College of Management

National Taiwan University

Master Thesis

應用遷移學習與文字探勘分析致股東報告書

Application of Transfer learning and Text Mining on Reports to  
Shareholders

陳予得

Yu-Te Chen

指導教授：林嬋娟 博士

Advisor: Chan-Jane Lin, Ph.D.

中華民國 110 年 4 月

April 2021



國立臺灣大學 (碩) 博士學位論文  
口試委員會審定書

應用遷移學習與文字探勘分析致股東報告書  
Application of Transfer learning and Text Mining on  
Reports to Shareholders

本論文係陳予得君 (R07722033) 在國立臺灣大學會計學系、所完成之碩 (博) 士學位論文，於民國 110 年 1 月 27 日承下列考試委員審查通過及口試及格，特此證明

口試委員：

林 瑋 娟 (簽名)  
(指導教授)

周 濟 平

盧 信 鈞

系主任、所長

劉 麗 仁 (簽名)  
(是否須簽章依各院系所規定)

## 誌謝

本身自己對資料分析或資料科學非常感興趣，因此當聽到嬋娟老師很鼓勵學生朝這些議題研究時，我是非常興奮的。尤其很感謝嬋娟老師的栽培以及討論研究上不同的想法和方向。同時感謝濟群老師和信銘老師在口試當天的鼓勵以及寶貴的建議，豐富了我論文的內容及嚴謹度。同時也非常感謝碩君會與我討論生活或研究的大小事。感謝昱斌、承芸、昱邦、柏翰於生活及論文上的協助，讓我更貼近生活。感謝鈺翔叮囑我的進度，以及在懷疑自己的時候，也能找回目標。

最後，感謝嬋娟老師不厭其煩地教導，總是說：做研究其實就是要花時間，可以在過程中，反覆地思考，試圖解決自己提出的問題。讓我體會到做研究的美妙之處。

## 中文摘要

本研究先以自然語言處理方法中的 BERT (Bidirectional Encoder Representation from Transformers) 建立文字探勘模型，並利用致股東報告書對 BERT 進行微調 (fine-tuning)。接著探討 BERT 是否解決過往文字探勘方法的缺點，最後以情緒分析 (Sentiment Analysis) 剖析致股東報告書的語調，研究致股東報告書語調對於公司未來績效的影響。

實證結果顯示，致股東報告書須針對中英夾雜問題做前處理，而經過驗證資料集表現篩選超參數 (hyperparameter) 後，BERT 模型分類準確率高達 86%。經過視覺化 BERT 模型的運作，發現其能捕捉否定詞修飾的詞彙，且同樣能成功捕捉形容詞所修飾的名詞。語境測試結果顯示，將文字順序隨機打亂後，BERT 表現大幅下滑，因此可知 BERT 確實有學習到語言結構。

然而關於語調對公司未來績效的影響，從實證結果發現，當年(t)的致股東報告書情緒對隔年(t+1)的盈餘並無顯著影響，推論原因可能是樣本篩選不夠具代表性，或是台灣致股東報書本身與美國的 MD&A 資訊含量有差異，導致台灣的致股東報告書與未來盈餘並無呈顯著關聯。

**關鍵詞：**深度學習、BERT、文字探勘、情緒分析、盈餘預測、致股東報告書

# ABSTRACT



First, this study applies BERT (Bidirectional Encoder Representation from Transformers) to construct a text mining model, and uses Report to Shareholders to fine-tune BERT. Next, we will discuss whether BERT can overcome some weaknesses of traditional text mining techniques. Finally, this study tries to assess the impact of the tone in Report to Shareholders on company's future performance by using Sentiment Analysis.

The empirical result shows that the problem of mixing Chinese and English in Report to Shareholders must be tackled, and after choosing the best hyperparameter based on validation performance, the classification accuracy reaches up to 86%. By visualizing the operation of BERT, we find that BERT can not only capture the relation between the word and its negation, but also capture the relation between the adjective and the noun successfully. The result from the context test also shows that the performance of BERT drop significantly after the text sequence is randomly shuffled, so it is considered that the language structure of Chinese is learned by BERT.

However, regarding to the impact of tone in Report to Shareholders on the company's future performance, the empirical result shows that the sentiment in Report to Shareholders has no significant impact on the next year's earnings. The results suggest that the sample may not be representative enough or Taiwan's Report to Shareholders has less information values than the US's MD&A information content so that there is no significant relation between the tone and the next year's earnings.

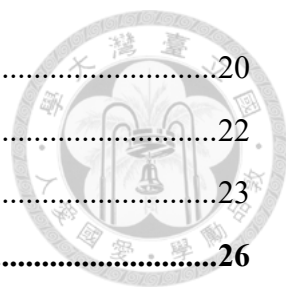
**Keywords** : Deep Learning, BERT, Text Mining, Sentiment Analysis, Earnings Prediction, Report to Shareholders

# 目錄



口試委員會審定書 .....	i
誌謝 .....	ii
中文摘要 .....	iii
ABSTRACT .....	iv
目錄 .....	v
圖目錄 .....	vii
表目錄 .....	viii
<b>第一章 緒論.....</b>	<b>1</b>
第一節 研究動機與目的.....	1
<b>第二章 文獻探討.....</b>	<b>4</b>
第一節 文字探勘於會計與財金領域上之應用.....	4
(一) 文字探勘於會計財金領域之應用.....	4
(二) 文字探勘方法之比較分析.....	7
第二節 自然語言處理方法-BERT .....	9
(一) 遷移學習 .....	9
(二) BERT .....	11
第三節 致股東報告書資訊價值.....	12
<b>第三章 研究設計.....</b>	<b>15</b>
第一節 研究流程.....	15
第二節 監督式學習.....	16
第三節 模型選擇與視覺化.....	17
(一) 模型選擇 .....	17
(二) 交叉驗證.....	18
(三) BERT 視覺化.....	19
(四) 語境測試.....	19
(五) 預測盈餘能力 .....	19
第四節 樣本選取.....	20

(一)	樣本期間.....	20
(二)	樣本篩選.....	22
(三)	樣本標記.....	23
<b>第四章</b>	<b>實證結果.....</b>	<b>26</b>
<b>第一節</b>	<b>模型分類結果.....</b>	<b>26</b>
(一)	模型選擇.....	27
(二)	交叉驗證.....	28
<b>第二節</b>	<b>視覺化與語境測試.....</b>	<b>32</b>
(一)	BERT 視覺化.....	32
(二)	語境測試.....	35
<b>第三節</b>	<b>預測盈餘能力.....</b>	<b>37</b>
(一)	敘述性統計.....	37
(二)	回歸結果.....	38
<b>第四節</b>	<b>額外測試.....</b>	<b>41</b>
(一)	字典法.....	41
(二)	TF-IDF.....	43
<b>第五章</b>	<b>研究結論與限制.....</b>	<b>46</b>
參考文獻	.....	48
<b>附錄</b>	.....	<b>51</b>
附錄一	BERTViz.....	51



## 圖目錄

圖一、BERT 訓練結構.....	11
圖二、研究流程圖.....	15
圖三、模型流程圖.....	17
圖四、台灣 IC 產業產值及成長率.....	22
圖五、BERT 視覺化(1).....	32
圖六、BERT 視覺化(2).....	33
圖七、BERT 視覺化(3).....	34





## 表目錄



表一、文字探勘於會計與財金領域之應用 .....	6
表二、文字探勘方法分析 .....	8
表三、半導體產業家數、產值及成長率 .....	21
表四、致股東報告書句子拆分為訓練、驗證和測試資料集 .....	25
表五、混淆矩陣 .....	26
表六、致股東報告書原始文字於驗證資料集的結果 .....	29
表七、經前處理的致股東報告書於驗證資料集的結果 .....	30
表八、經前處理的致股東報告書於 5-Fold CV 平均結果 .....	31
表九、隨機打亂的致股東報告書於 5-Fold CV 平均結果 .....	36
表十、敘述性統計 .....	38
表十一、下一年淨利的線性回歸結果(BERT).....	39
表十二、下一年淨利的線性回歸結果(人工) .....	40
表十三、情緒分數敘述性統計-字典法 .....	41
表十四、字典法情緒分析結果 .....	42
表十五、字典法與 BERT 之比較 .....	42
表十六、情緒分數敘述性統計-TF-IDF .....	44
表十七、TF-IDF 情緒分析結果.....	44
表十八、三個方法之比較 .....	45

# 第一章 緒論

## 第一節 研究動機與目的



美國 Securities and Exchange Commission (SEC) 要求公開發行公司於年報中揭露 Management's Discussion & Analysis (MD&A)，因為 SEC 認為財務報表及其附註，無法提供充分資訊使財務報表使用者判斷公司盈餘品質以及預測公司未來績效 (SEC 1987)。因此 SEC 要求公司提供 MD&A，其內容應使讀者瞭解對於公司現在或未來營運、流動性或資本有重大影響之市場趨勢、需求、事件或不確定性 (SEC 2003)，根據 SEC Financial Reporting Manual<sup>1</sup> 於 MD&A 有以下要求及目的：

- (1) 提供關於公司財務報告之文字敘述，使投資者能從管理階層角度分析公司。
- (2) 加強整體財務資訊之揭露，並提供應分析之財務資訊之背景。
- (3) 提供關於公司盈餘及現金流量之品質和潛在變化之資訊，以便投資者能確信過去績效有代表未來績效之可能性。
- (4) 要求揭露營業結果、未來資本支出以及已知市場趨勢或不確定性可能重大地影響收入或費用…等等。

過去有許多文獻顯示在財務分析中，MD&A 在討論未來營運方向或資本支出未來規劃的確會幫助財務報表使用者預測公司未來表現。如：Bryan (1997) 調查 MD&A 是否能提供其他額外的資訊，他將 MD&A 揭露之訊息分類為有利、中立的及不利的，並且研究 MD&A 語調與未來銷售、資本支出、營運現金流量以及盈餘變化是否有關聯，發現 MD&A 的確有與未來財務數據相關。Cole and Jones (2004) 發現在零售業裡，有關 MD&A 揭露收入變化之原因，如既有店面銷售成長或擴增新店面，以及揭露未來資本支出有助於使用者預測未來銷貨收入以及盈餘。Sun (2010) 發現關於 MD&A 存貨揭露的確有幫助使用者解讀這些存貨大量地增加的現象以及預測未來公司的表現。Bochkay and Levine (2019) 透過字典法與文字頻率矩陣建立文字資訊相關變數，發現財務數據結合 MD&A 之文字資訊相較於只有財務數據能夠更準確地預測未來一年的股東權益報酬率 (ROE)。然而前述的文獻須依賴人工進行分類，導致樣本數量太少的問題，可能會造成推論無法普及的問題。

從文本分析相關文獻中，發現有許多不同的文本分析方法，試圖解決上段所述

---

<sup>1</sup> 資料來源：<https://www.sec.gov/corpfin/cf-manual>


之問題，例如：Li (2010a) 隨機抽取三萬行句子訓練樸素貝氏分類器，並對一千多萬行句子進行語調之分類，並發現 Forward-Looking Statement (FLS) 與未來盈餘呈現顯著正相關。Loughran and McDonald (2011) 利用字典法對文本進行分類，將文本語調與公司股價、交易量、舞弊等進行連結。Siano and Wysocki (2020) 利用 BERT (Bidirectional Encoder Representation from Transformers)、字典法及樸素貝氏分類進行語調預測準確度比較，發現 BERT 表現得更為準確；並且提到過去會計研究者受限於較少的資訊背景以及硬體設備，因此無法將機器學習廣泛地應用於文本分析上，但這些限制會隨著愈來愈方便使用的遷移學習 (transfer learning) 工具 (例如：BERT) 而解除。本研究之目的在參考 Siano and Wysocki (2020) 的作法，先將 BERT 應用於中文文本語調分類，且嘗試使用 BERT 視覺化，解釋 BERT 與過往文字探勘方法的差異，最後參考 Li (2010) 的作法，觀察 MD&A 的文字內容是否能提供未來盈餘額外的資訊價值。

台灣公司年報並不包含上述之 MD&A，惟年報裡的致股東報告書內容與上述 MD&A 要求及目的十分相近。根據公開發行公司年報應行記載事項準則<sup>2</sup>規定，致股東報告書應包含前一年度之營業結果、本年度營業計畫概要、未來公司發展策略以及受到外部競爭環境、法規環境及總體經濟環境之影響。因此本文選擇致股東報告書作為本研究文本分析之標的。基於半導體業的景氣經常隨著產品需求和趨勢變化而產生興衰，若能從管理階層的角度瞭解發展趨勢或是需求變化，則可以預測未來公司的表現，因此本研究將以半導體產業為研究對象。綜合上述，本研究旨在介紹與應用最新的自然語言處理工具 BERT，探討 BERT 模型應用於致股東報告書之情緒分類表現、BERT 視覺化和語境測試，以及說明 BERT 與過去傳統文字探勘方法的不同之處和其如何解決傳統文字探勘方法的缺點。而本研究另外透過已經訓練好的 BERT，幫助我們辨別致股東報告書的情緒，並進一步結合財務數據去預測未來盈餘，作為將 BERT 應用於會計主題上的範例。

本研究利用 BERT 模型對致股東報告書進行情緒分析，實證結果顯示若能適當處理致股東報告書的中英夾雜的狀況，能提高 BERT 模型情緒分析的準確率。在交叉驗證下，BERT 模型準確率高達 86%，且模型在預測正向語調和負向語調的表現結果相差無幾，與 Loughran and McDonald (2016)和 Siano and Wysocki (2020)的研

---

<sup>2</sup> 資料來源：<https://law.moj.gov.tw/LawClass/LawAll.aspx?pcode=G0400022>



究結果表現不一致。推測原因有四點，一是 BERT 有考慮語言結構，在模型視覺化和語境測試的結果顯示，BERT 不但有考慮文字的排列的順序性，更能成功地捕捉形容詞和否定詞與其各自修飾的詞彙之關係；二是類別數量的不同，觀察近千份的致股東報告書，發現多數句子屬於中立的語調，故情緒的標記不適合使用二分法；三是標記類別的方式，若依照 Siano and Wysocki (2020)的作法同一篇致股東報告書標記成一個情緒，可能會導致模型在學習時，學習到錯誤的類別；四是關於致股東報告書的本質，部分公司在前後年出具的致股東報告書，內容相差不遠，甚至出現一樣的句子，因此容易讓模型預測更加準確。本研究最終使用文字資訊結合財務數據以預測公司未來盈餘作為應用 BERT 模型範例。該作法目的是驗證致股東報告書的資訊價值。研究結果顯示，當年致股東報告書情緒與隔年的盈餘並沒有呈現顯著關係。推測原因有兩點，一是選擇致股東報告書前十行過於單薄，理想的作法是隨機選取整篇致股東報告書的其中十行句子；二是台灣的致股東報告書與美國的 MD&A 本身資訊含量差異，導致中文致股東報告書與隔年盈餘較無關聯。

本研究貢獻有以下兩點，第一，應用 BERT 於致股東報告書進行情緒分析，因為過去會計領域分析股東報告書方法，常是字典法、樸素貝氏演算法或是其他傳統機器學習模型，而目前遷移學習被大量應用於自然語言處理，並且取得相當好的成果。本文實證結果顯示 BERT 同樣能成功應用在中文致股東報告書情緒分析上，且獲得不錯的準確率，因此可以增加未來會計領域研究方法的多樣性。第二，分析 BERT 與傳統文字探勘方法的差異，以視覺化和語境測試解釋差異，能夠更直觀地瞭解 BERT 模型相較於其他傳統文字探勘方法的優勢。第三，將 BERT 應用於盈餘預測作為會計研究的範例。

本研究共有五章，除本章說明研究動機與目的外，其餘內容安排如下：第二章為文獻探討，第三章說明本研究的設計與流程，第四章說明實證結果，最後一章為結論以及研究限制。

## 第二章 文獻探討

### 第一節 文字探勘於會計與財金領域上之應用

文字探勘應用文獻採用之方法包括計算語言學 (computational linguistics)、自然語言處理 (natural language processing)、資訊檢索 (information retrieval)和內容分析 (content analysis)<sup>3</sup>，而其中自然語言處理是指讓機器擁有理解人類語言之能力。由於本研究主要目的之一為透過自然語言處理判斷致股東報告書文字情緒(正向、負向或中立)，因此本研究主題屬於上述文字探勘的一種。以下文獻將先介紹文字探勘應用於會計的哪些主題上；其次介紹文字探勘方法之比較分析。

#### (一) 文字探勘於會計財金領域之應用

有關會計與財金領域之文字資訊，一般會聯想到由公司自己發佈在年度財務報表上之文字資訊，但其實屬於會計與財金領域的文本還囊括了分析師研究報告、法人說明會、新聞媒體之輿論等不同類型的資料。因此以下介紹將包含上述不同類型的文本，而有關於會計與財金領域之文字探勘應用文獻大致可分為 (a)情緒分析、(b)可讀性、(c)風險和(d)競爭程度等不同的主題進行分析 (Li, 2010b; Loughran and McDonald, 2016)，將其彙整成表一。

(a) 情緒分析 (sentiment analysis)：目的是探討文章想表達的觀點及態度，可能是樂觀、悲觀或者是確定、不確定性組合而成，而中立語調可以視為第三種語調，因為大部分詞句既不樂觀也不悲觀。Li (2010a) 的研究中，作者從前瞻性陳述內容隨機選取三萬行句子訓練樸素貝氏演算法，訓練目標有二，一是讓模型理解句子的語調是正向、負向、中立或不確定；二是讓模型理解內容的主題，分成有關營業收入、成本、淨利等等十二種類別，經過交叉驗證 (cross validation) 後，語調預測與主題預測分別獲得約 67%和 63%的準確率。作者進一步透過已經訓練好的樸素貝氏演算法去預測其他樣本的語調，並結合其他有關未來績效的控制變數，不但發現影響前瞻性陳述內容語調的因子，還發現前瞻性陳述內容語調與未來盈餘呈正向關聯。Loughran and McDonald (2011)利用 Harvard Dictionary (Harvard-IV-4 TagNeg)分析 10-K 文本，卻沒有發現 Harvard 負向詞彙

<sup>3</sup> 資料來源：<http://terms.naer.edu.tw/detail/1683205/>，根據定義內容分析係透過量化和質性的分析，客觀和有系統地對文件內容進行研究，解析傳播內容中各種語言和特性。

與年報發佈後的股價超額報酬率呈現關聯，作者認為 Harvard-IV-4 TagNeg 有將近四分之三的詞彙在財務相關文章不應該被分類為負向。而作者依照 10-K 的內容建立了一個專門的字典，並再重新量化分析一遍，結果發現負向詞彙與年報發佈後的股價超額報酬率、異常交易量和報酬波動性呈現顯著關聯。Price, Doran, Peterson, and Bliss (2012) 除了利用 Harvard-IV-4 字典，同時結合自己建立有關於盈餘詞彙的字典，以分析法說會資訊，研究結果顯示，法說會語調不但會引起市場價格的短期變化，還會引起股票交易量的變化。

- (b) 可讀性 (readability): 目的是探討讀者是否能良好理解文章表達之觀點或內容。Li (2008) 利用 Fog Index<sup>4</sup> 和字數兩項指標衡量年報可讀性，並發現公司盈餘越低，其年報可讀性越低，反之亦然。Biddle, Hilary, and Verdi (2009) 同樣透過 Fog Index 及其他兩項變數衡量年報品質，研究結果顯示公司年報品質與資本投資效率 (capital investment efficiency) 呈現正相關。Miller (2010) 透過分析 10-K 文章長度及可讀性，發現公司出具愈複雜的年報，其公司股票交易量愈低。作者進一步探討影響交易量是來自於小額投資人還是大額投資人，結果顯示 10-K 可讀性對小額投資人判斷產生顯著影響。Loughran and McDonald (2014) 發現文件大小與字數、詞彙數和 Fog Index 呈正相關，反而在 10-K 文件中，每個句子平均單字數與合成詞比率呈現負相關，因此認為合成詞比率作為 Fog Index 要素是有偏誤的，故作者建議使用 10-K 文件大小作為衡量可讀性的指標。
- (c) 風險：目的是探討公司面臨不同類型的風險。Rogers, Van Buskirk, and Zechman (2011) 在探討盈餘公告 (earnings announcements) 語調與被股東提起訴訟的風險，結果顯示當公司在盈餘公告時，態度愈樂觀，公司被股東提起訴訟的機率愈高。Campbell, Chen, Dhaliwal, Lu, and Steele (2014) 探討 10-K 強制揭露風險因子帶來的影響，並發現公司面臨愈高風險，愈願意揭露更多的風險因子及其性質。
- (d) 競爭程度：目的是探討公司競爭環境。Li, Lundholm, and Minnis (2011) 研究公司管理階層對於公司競爭環境的態度與盈餘品質的關聯，個體經濟學表示當公司競爭中，所有競爭公司的投資報酬趨於平等，企業將離開無利可圖的行業，進入有利可圖的行業，因此最終邊際收益會平均回歸 (mean reversion)。作者透

---

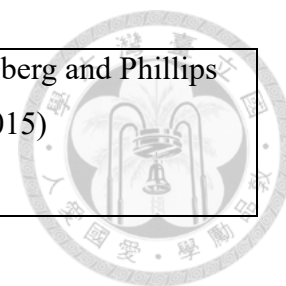
<sup>4</sup> Fog Index = 0.4(每個句子平均單字數+合成詞比率(complex words))

過分析 10-K 去評估其競爭環境，並發現當 10-K 表示公司競爭愈激烈，其淨營運資產報酬率回歸平均的速度和邊際收益遞減的速率愈快，因此可以推斷 10-K 有關競爭環境的資訊能幫助衡量邊際收益遞減速率。Hoberg and Phillips (2015) 透過單詞向量化衡量 10-K 中關於產品描述，以判斷不同公司之間的产品是否相似，根據該產品相似度，每年可以建立屬於公司自己的競爭者。該作法解決傳統使用 Standard Industrial Classification (SIC) 作為競爭衡量的問題，因為 SIC 不會因為公司產品市場變化做出重大改變，也不會因為新市場誕生而及時創造一個新產業的代碼以及 SIC 沒有提供產業內和產業間相似度分析。

表一、文字探勘於會計與財金領域之應用

主題	定義	研究主題	代表文獻
情緒分析	分析文字不確定性、樂觀、悲觀或中立	<ol style="list-style-type: none"> <li>1. 公司文字資訊語調與公司未來財務績效關係；</li> <li>2. 公司文本語調對公司股價之影響；</li> <li>3. 法人說明會與股價之影響</li> </ol>	Li (2010a); Loughran and McDonald (2011); Price et al. (2012)
可讀性	公司揭露文字資訊是否便於理解或者晦澀難懂的程度	<ol style="list-style-type: none"> <li>1. 衡量可讀性方法；</li> <li>2. 年報可讀性與公司績效關係；</li> <li>3. 年報可讀性與公司投資效率之關係；</li> <li>4. 年報可讀性對投資人之影響</li> </ol>	Li (2008); Biddle et al. (2009); Miller (2010); Loughran and McDonald (2014)
風險	公司面臨之不同種類風險	<ol style="list-style-type: none"> <li>1. 盈餘宣告與公司訴訟風險之關聯；</li> <li>2. 強制揭露風險確實促進公司對風險的揭露</li> </ol>	Rogers et al. (2011); Loughran and McDonald (2011); Campbell et al. (2014)
競爭程度	公司面臨之	<ol style="list-style-type: none"> <li>1. 競爭程度對盈餘品質之</li> </ol>	Li et al. (2011);

	競爭程度	影響； 2. 測量競爭程度之方法	Hoberg and Phillips (2015)
--	------	---------------------	-------------------------------



## (二) 文字探勘方法之比較分析

文字探勘最基本的問題是如何將非結構化的文件萃取出與研究主題相關的特徵，早期文獻採用人工閱讀的方式，而現今文字探勘方法以電腦計算為基礎，能將隱含於字裡行間的資訊轉換為數值型態，不僅能夠降低人工成本，還能夠大量增加研究效率。因此在此將介紹不同文字探勘方法以及其各自的優劣勢，並彙整成表二。

- (a) 人工判讀：早期有關於公司文本分析文獻使用大量人工閱讀以理解文章的情緒或觀點，像是 Bryan (1997) 透過人工閱讀 250 篇 MD&A 和 Cole and Jones (2004) 閱讀 568 篇零售業的 MD&A。這樣的方法有其優點：(i) 方法簡單；(ii) 準確率高；(iii) 適用於大部分的分析，如情緒分析、文本內容分析或是命名實體識別 (Named Entity Recognition)。而相反的人工閱讀有其他致命的缺點：(i) 需要大量人力；(ii) 樣本數量少，限制實證結果的可推論性 (generalizability)；(iii) 因為判斷涉及主觀性，因此難以複製。
- (b) 字典法：字典法是基於預設的字典和規則，將研究文本的詞彙逐一映射至預設的不同種類。字典可以分為通用性及專用，如 Harvard-IV-4 是常用的英文通用辭典，而像是 Henry (2008) 和 Loughran and McDonald (2011) 建立的辭典屬於會計專用的辭典，具有更高的分類準確度。Hoberg and Phillips (2015) 透過字典法將 10-K 中關於產品描述映射於向量空間中，並分析產品描述的文字相似度，藉此衡量公司之間的競爭程度。字典法的優點有 (i) 方法簡單；(ii) 容易應用於不同主題。而其缺點是 (i) 忽略研究主題須具備之先前知識，如在 MD&A 中，多數句子為中立語調，若單用正負情緒的字典容易造成分類不準確的問題；(ii) 詞彙存在多種意義，在不同情境有不同的語意，而字典法沒辦法捕捉這樣語境問題，會造成判斷不準確；(iii) 若沒有透過人工標籤，沒有辦法驗證分類準確率；(iv) 對語言依賴性大；(v) 不具學習能力。
- (c) 機器學習：機器學習本質上是統計算法，具有類似人工智慧的學習能力。而被應用於會計文本分析上的方法，有 N-grams、支撐向量機法、及最廣為人知的



樸素貝氏演算法。樸素貝氏演算法是假設辭彙之間獨立下，運用貝氏定理而做成的分類器，其概念是假設一個句子中，每個詞彙都與其他詞彙無關，而由於獨立的性質，故將各個詞彙出現在特定類別的機率相乘，求出該句子出現在特定類別的機率。Antweiler and Frank (2004) 蒐集 Yahoo! 財經頻道及 Raging Bull 上 150 萬則股票貼文，隨機選取 1,000 則貼文訓練樸素貝氏演算法。儘管貼文對股票報酬影響有限，作者仍發現貼文的則數與股票報酬波動有關。Li (2010a) 隨機從前瞻性陳述內容(FLS)抽取 30,000 筆句子，透過樸素貝氏演算法訓練，發現 FLS 與後續盈餘呈正相關。Huang, Zang, and Zheng (2014) 利用樸素貝氏演算法分析 363,952 份分析師報告，並發現分析師報告中的正向語句對公司發布財報五年後盈餘增長率有很大的影響。Li (2010b) 指出樸素貝氏演算法相較字典法擁有下列優點：(i) 研究主題不需要有相關的字典；(ii) 字典法未考慮辭彙所處之語境，如「增加」本來應是正向辭彙，但若與「成本」一起出現，則應被分類負向詞彙；(iii) 字典法忽略研究主題須具備之先前知識；(iv) 能衡量模型準確率。相較字典法擁有的缺點則是(i) 難度較字典法大；(ii) 詞彙之間不一定獨立；(iii) 監督式學習需要預分類。

表二、文字探勘方法分析

方法類型	優點	缺點	相關文獻
人工判讀	<ol style="list-style-type: none"> <li>1. 方法簡單</li> <li>2. 準確率高</li> <li>3. 適用於大部分的分析</li> </ol>	<ol style="list-style-type: none"> <li>1. 耗費大量人力與時間</li> <li>2. 樣本數量少</li> <li>3. 涉及主觀性，難以複製</li> </ol>	Bryan (1997); Cole and Jones (2004)
字典法	<ol style="list-style-type: none"> <li>1. 方法簡單</li> <li>2. 容易應用於不同主題</li> </ol>	<ol style="list-style-type: none"> <li>1. 忽略研究主題須具備之先前知識</li> <li>2. 未考慮詞彙存在多種意思</li> <li>3. 若無人工標籤，無法驗證分類準確率</li> <li>4. 對語言依賴性大</li> </ol>	Henry (2008); Loughran and McDonald (2011); Hoberg and Phillips (2015)

		5. 不具有學習能力	
機器學習	1. 研究主題不需要有相關的字典 2. 考慮辭彙所處的語境 3. 能衡量模型準確率	1. 難度較大 2. 詞彙之間不一定獨立 3. 監督式學習需要預分類	Antweiler and Frank (2004); Li (2010a); Huang, Zang, and Zheng (2014)

## 第二節 自然語言處理方法-BERT

因本文將採用 BERT，故此節將先介紹遷移學習，接著探討什麼是 BERT 和 BERT 是如何進行預訓練的。

### (一) 遷移學習

Pan (2009) 指出不論是機器學習或是深度學習在分類或回歸的問題中，會有一個常見的假設是訓練資料和測試資料服從相同的分佈及特徵空間(feature space)。當分佈改變時，大部分模型需要重新蒐集訓練資料去重建一個模型，然而在實際應用中，蒐集資料的動作既耗時又複雜，要建置一個大量、高質量被標註的數據庫極度困難，故不充足的訓練資料是無可避免的問題。因此若能減少蒐集資料的需求，將其他領域中的知識遷移至使用者研究之領域，以提高使用者研究領域之分類效果，將是一個不錯的選擇，故遷移學習的概念由此誕生；簡而言之，遷移學習提供會計研究一個具「大數據」能力的「預訓練」模型，只需要一小部分的會計專業領域訓練資料，便能成功地將模型「微調」至會計專業領域並解決相關的研究問題。

近期關於自然語言處理領域中，許多預訓練模型相繼提出，如 ELMo (Peter, Neumann, Iyyer, Gardner, Clark, Lee, and Zettlemoyer,2018)、GPT (Radford, Narasimhan, Salimans, and Sutskever,2018)、BERT (Delvin, Chang, Lee, and Toutanova,2019)，成功地打破先前其他模型在通用語言理解評估基準 (General Language Understanding Evaluation, 簡稱 GLUE)<sup>5</sup>、SQuAD (Stanford Question

<sup>5</sup> GLUE 是由紐約大學、華盛頓大學和 DeepMind 的團隊提出的數據集，其中包含 11 項常見的自然語言處理的任務，例如：CoLA (The Corpus of Linguistic Acceptability) 是辨認每個句子語法是否正確，或是 SST-2 (The Stanford Sentiment Treebank) 是關於句子的情緒分析。

Answering Dataset)<sup>6</sup> 及 SWAG (Situations With Adversarial Generations)<sup>7</sup>的紀錄。這些不同的模型都有以下的特點：(1)模型結構非常大且深度非常深，如 BERT 參數有差不多 1.1 億個。(2)用大量的資料預訓練，如 BERT-Chinese 是用中文維基百科預訓練而成的。而這些特點能夠讓模型捕捉到更良好的詞向量(word embeddings)<sup>8</sup>，例如：同個詞彙其詞向量會隨著語境變化而變化、具有捕捉否定詞以及理解相隔較遠的詞彙意義之特色，可解決字典法、樸素貝氏演算法或其他傳統文字探勘方法的缺點。

上述提到許多其他不同的預訓練模型，為什麼本研究將鎖定於 BERT 呢？原因如下：

- (1)BERT 是 Google 開發的免費開源軟體，因此任何人都可以使用。
- (2)Delvin et al. (2019) 指出 BERT 在不同自然語言處理任務中，表現得更為出色。
- (3) BERT 的操作簡單，且網路有非常多資源，容易上手。

根據上述之特點，可以預期 BERT 模型在會計專業領域中的情緒分析能夠表現良好，有關 BERT 在財金、會計及社會科學領域的應用，例如 Hiew, Huang, Mou, Li, Wu, and Xu (2019)鎖定三家在香港交易所 (Hong Kong Stock Exchange) 上市且股票交易熱門的公司，透過蒐集有關這三家公司的微博 (Weibo) 貼文，訓練五種不同的情緒分析模型，發現 BERT 從中脫穎而出。Elwany, Moore, and Oberoi (2019) 目的係區分法律文件中協議條款是屬於自動更新的類別，還是屬於經過固定期限後會到期的類別，因此利用大量特定領域法律文件微調 BERT 以提高模型分類準確性，最後得到 Precision、Recall 和 F1-score 皆高達九成。Siano and Wysocki (2020)探討 BERT 應用於季別盈餘報告中之情緒分析表現，透過比較 Loughran and McDonald 的字典法、樸素貝氏演算法、隨機分類及 BERT 的情緒分類模型，發現 BERT 較其他三個方法表現得更為出色，並且進一步測試 BERT 是否只依高頻率的單字標記句子的情緒及是否考慮語境問題，發現儘管刪除高頻率單字會使 BERT 模型分類準確率些微下降，但仍表現得較其他三個模型好，且當句子被打亂時，BERT

---

<sup>6</sup> SQuAD 是由史丹佛大學的數據集，其目標為給定一篇文章，準備相關的問題，需要模型給出問題的答案。

<sup>7</sup> SWAG 任務是給定一個陳述句子和其他四個句子，模型判斷哪個句子最具有邏輯性，相當於閱讀理解的問題。

<sup>8</sup> 詞向量係將抽象的辭彙之間的語意關係量化成向量形式。

表現顯著下降，因此佐證 BERT 考慮到上下文問題，並非只依靠單詞而做判斷。Li, Li, Wang, Jia, and Rui (2020) 蒐集投資者於 Eastern Stock Exchange 的股票線上評論，將其分類為正向、負向及無，並且細分為八種類別，發現 BERT 分類表現較支撐向量機及 LSTM (Long short-term memory)<sup>9</sup> 模型來得更為準確。



## (二) BERT

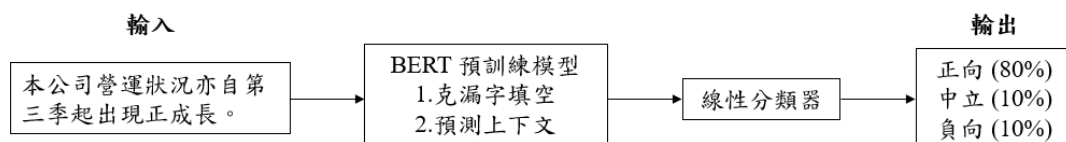
BERT 是 Google 以無監督式學習訓練大量文本而建置的模型，其本身的概念是語言模型(Language Model, LM)的一種變形，語言模型是給定  $m$  個詞彙組成的句子，去預估下一個詞彙  $w_{m+1}$  出現的機率分佈，其數學形式如公式(1)

$$P(W) = P(w_{m+1} | w_1, w_2, w_3, \dots, w_m) \quad (1)$$

BERT 不同於以往訓練語言模型作法只從句子左邊訓練到右邊或是從句子右邊訓練到左邊，Delvin et al. (2019) 證明 BERT 的雙向訓練語言模型比上述單向訓練的語言模型能更好地捕捉詞彙順序及上下文關係。那麼 BERT 是如何進行預訓練的以達到所謂的雙向訓練呢？其訓練的任務有二：

- (1)克漏字填空：隨機選取 15%的辭彙將其遮蔽，模型將透過未遮蔽的詞彙去學習預測最有可能出現的辭彙。
- (2)預測上下文：給定兩個句子，模型將學習預測這兩個句子是否為上下文或是沒有任何關係。

這兩項任務最大好處是其不需要透過標註過的資料訓練(無監督式學習)，因此機器能夠自動學習大規模的資料以及詞向量。當預訓練的模型建置好後，研究者可以依不同任務需求，利用其專業領域的數據在 BERT 最後一層微調。以情緒分析為例，將在 BERT 模型最後一層加上線性分類器進行微調，因此，從頭開始訓練的參數只有屬於線性分類器的參數而已，圖一、BERT 訓練結構顯示其流程。



圖一、BERT 訓練結構

<sup>9</sup> LSTM 是目前遞歸神經網路 (Recurrent neural network) 最常使用的模型，適合處理時間序列的問題，由於文字結構包含先後順序的問題，因此自然語言處理常常使用該模型。

在此定義和說明本研究在訓練 BERT 模型使用的專有名詞。首先將先介紹何謂過擬合 (overfitting) 以及欠擬合 (underfitting)。過擬合是指當模型過於精確地訓練特定資料集，導致其無法良好地調適其他資料或預測未來的結果，因此模型會在未學習過的資料集上表現不佳；欠擬合是指模型學習能力不足，無法學習到數據集中的“一般規律”，導致其表現不論在已經學習的資料集或未學習得資料集皆會非常差勁。上述兩種情形是在機器學習領域中應當避免的情況。

而在涉及類神經網路訓練過程中，模型至少要包含調下列的超參數：


- (1)學習率 (learning rate)：係指控制當模型基於損失函數產生之梯度調整神經網路的權重的速度，較高的學習率會導致在訓練過程中無法收斂於局部極小值，較低的學習率可避免錯過局部極小值，但其所須要收斂的時間較長，且可能有欠擬合情形；
- (2)Epochs：係指神經網路遍歷一次完整的訓練資料並且返回了一次，太多 Epochs 會花費大量時間且造成模型過擬合，太少 Epochs 則會讓模型有欠擬合的情形；
- (3)Batch Size：係指一次訓練的樣本數目，太大的 Batch Size 可能導致過擬合情形，太小的 Batch Size 可能導致欠擬合情形；
- (4)有代表性的驗證資料集：驗證資料集扮演的角色，是檢視模型在新數據上的表現，同時透過調整超參數，去選擇在驗證資料集表現最好的模型，最後再透過該模型在測試資料集的表現作為衡量標準。

其中學習率、Epochs 及 Batch Size 調整是為了達到最低的損失和最高的準確率，同時避免過擬合或欠擬合的情況發生。而將驗證資料集特別排除於訓練及測試資料集之外的原因是避免高估模型的準確率，因為若將驗證資料集與測試資料集混合在一起，並透過該資料集的表現去選擇超參數的話，會造成模型結果嚴重高估而失去參考依據。

### 第三節 致股東報告書資訊價值

由於本研究將進行預測盈餘的測試，目的在透過 BERT 分析非結構化的致股東報告書文字內容後，探討該文本是否能提供財務數據以外的資訊，以幫助預測未來一年的盈餘，因此以下將整理關於 MD&A 文字資訊如何增強財務數據的資訊價值之相關文獻。

SEC 認為單就財務報表上面的數字以及附註不足以讓投資者判斷該盈餘品質和過去績效有代表未來的可能性，因此要求公司揭露 MD&A 以提供足夠的資訊幫



助投資者評估公司未來營運與長期發展，而大部分文獻便是蒐集 MD&A 揭露的訊息，並探討這些訊息實際上是否能幫助投資者更準確地預測公司未來走向。例如：Bryan (1997) 透過人工閱讀 1990 年 250 篇 MD&A，依照 MD&A 揭露的訊息，區分為銷售價格、銷售量、營收、成本、資本支出和未來趨勢七大變數，並分別標記其情緒是正向、中立、負向或沒有揭露，作者發現綜合上述七大變數，出現最多次的情緒類別是中立語調。為了研究 MD&A 和公司未來財務績效的關聯，作者使用羅吉斯回歸 (Logistic regression)，依變數分別是未來銷貨收入變化方向、未來 EPS 變化方向、未來營運現金流量變化方向及未來資本支出變化方向，而應變數包含上述七個變數和其他財務控制變數。研究結果顯示 MD&A 有關銷售量變化和未來趨勢之揭露與未來一年的銷貨收入變化方向有顯著正相關；有關未來趨勢之揭露與未來一年的 EPS 變化方向同樣是顯著正相關，因此發現 MD&A 揭露的訊息的確能夠幫助投資者預測未來銷貨收入變化和 EPS 變化。然而七大變數中沒有任何一個與未來現金流量變化方向有關，而這個結果與 SEC 預期相關揭露能幫助財報使用者預測營運現金流量的假設相互違背；有關公司財務流動性、已規劃資本支出及未來趨勢之揭露則與公司未來一年的資本支出變化方向有顯著相關。

由於 SEC 認為公司於 MD&A 揭露銷貨收入變化的原因比只有揭露其變化來得更具有價值，因此 Cole and Jones (2004) 探討零售業於 MD&A 揭露不同的銷貨收入變化原因，是否幫助投資者預測公司未來營運和資本支出。零售業銷貨收入變化原因主要有二，一是來自於同店銷售 (comparable store sales) 的成長；二是來自於營運規模的變化，例如：新店面開幕、舊店面倒閉或購置其他店面。作者預期同店銷售的成長與新店面的開幕應與未來銷貨收入成長呈現正相關，且同店銷售的成長可能與未來盈餘和股價呈現正相關，因為同店銷售的成長意味著比新店面開幕更少的成本賺取更多的收入。作者蒐集 160 家零售業的 MD&A，依變數是未來收入變化、未來盈餘變化和同期股價報酬，並依照這三個不同的依變數分別做四種回歸，第一種是只包含財務控制變數；第二種是財務控制變數加上有關該年同店銷售成長和新店面開幕或關閉的數量；第三種是第二種回歸變數加上公司預計在下一年開幕的店面數量；第四種是第二種回歸變數加上公司預計下一年的資本支出。研究結果如同作者的預期，同店銷售的成長與新店面的開張和未來一年的銷貨收入變化呈顯著正相關，且下一年開幕的店面數量同樣與未來一年的銷貨收入變化呈

現顯著正相關；同店銷售的成長與未來盈餘和股價亦呈現顯著正相關，上述結果皆證明 MD&A 揭露不同銷貨收入變化原因具有不同的資訊價值。

同樣地，存貨的變化原因對製造業影響甚鉅，一般來說，存貨數量大量增加代表公司未來財務表現不佳，但其實有可能是因為公司預期產品未來需求遽升，而開始做準備。故 Sun (2010) 探討 MD&A 有關存貨變化有利或不利的解釋與公司未來財務績效是否有關。作者閱讀了 568 個 MD&A，並將存貨變化的解釋分類呈有利的、不利的及中立的。為了探討存貨解釋預測公司未來績效的能力，依變數為未來的 ROA 和未來的銷貨收入成長，而自變數是有利的解釋、不利的解釋以及其他控制變數。研究結果顯示對於存貨變化有利的解釋與未來的 ROA 和未來的銷貨收入成長皆呈現顯著正相關，而存貨變化不利的解釋則與下一年的 ROA 呈現顯著負相關，可以得知公司對存貨變化做出有利的解釋，其未來 ROA 和銷售成長會高於那些對存貨變化做出中立解釋的公司。Bochkay and Levine (2019) 蒐集 1994 年至 2012 年 10-K 資料，並利用 Loughran and McDonald (2011) 辭典，計算詞彙出現的頻率。接著使用 Ridge Regression 篩選前五十個與預測未來盈餘顯著相關的詞彙，發現超過一半以上的詞彙是負向詞彙，而正向詞彙只佔四分之一，同時作者還發現若減少伴隨著成本出現，模型會將減少視為正向。研究結果顯示，結合財務資訊與 MD&A 文字資訊的模型預測未來一年的 ROE 較只有包含財務資訊的模型更為準確。同時發現 MD&A 的資訊價值會隨著公司特性而不同，那些未來績效起伏較大、未來績效差、嚴格投資人審查及高機率出現財務困境等等的公司，其 MD&A 具有更高的資訊含量。

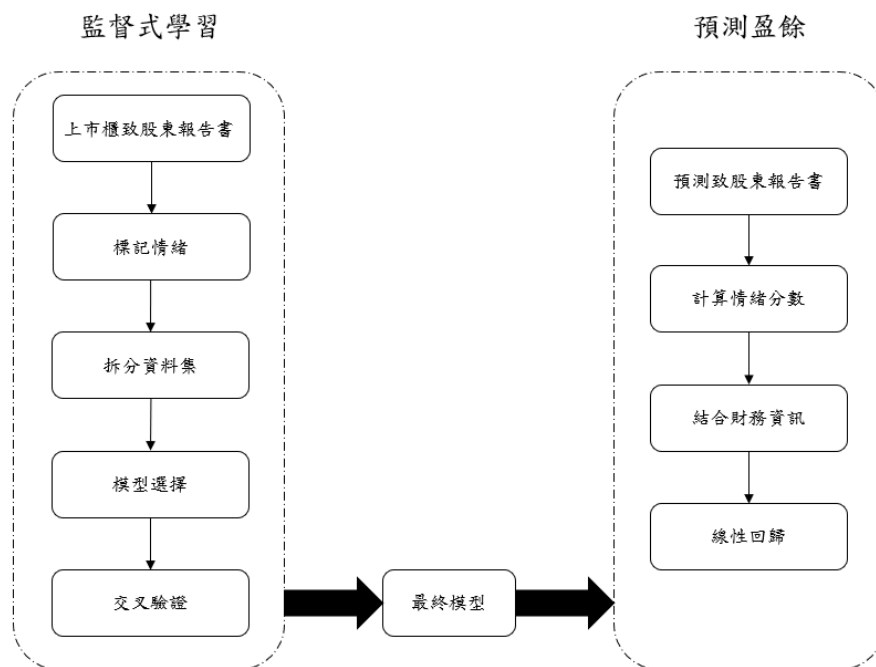
# 第三章 研究設計

## 第一節 研究流程



研究流程如圖二。首先，本研究將樣本限縮於半導體產業的上市(櫃)公司，並選取致股東報告書前十行句子進行人工標記。接著，依照 8:2 將資料集拆分為訓練資料集以及驗證資料集，訓練資料集再依 8:2 拆分為子訓練資料集和測試資料集，最後透過子訓練資料集和驗證資料集的結果，篩選最好的模型，並檢視其於測試資料集的表現。為了避免拆分資料集的偏誤，在此對樣本進行 5 折交叉驗證 (5-fold Cross Validation)，每一次會產生一個預測結果，最後再將 5 次的預測結果平均，作為該模型最後的表現。最後，將最終模型應用於 2011 年至 2018 年的致股東報告書情緒預測，並結合各年度的財務資訊對隔年的盈餘 (t+1) 進行線性回歸。

本章將依研究流程的順序進行介紹，首先介紹何謂監督式學習，說明為什麼需要手動標記致股東報告書的情緒。再來將介紹模型選擇，解釋本研究如何選擇最終模型，最後介紹解讀 BERT 的方法，並強調 BERT 相較於過往文字探勘方法的優點，以及將 BERT 應用於盈餘預測作為 BERT 應用於會計研究的範例。



圖二、研究流程圖



## 第二節 監督式學習

監督式學習定義<sup>10</sup>為透過模型於訓練資料中，學到或建立一個假設，並依照該假設去預測新的資料，而訓練資料是由輸入物件和標籤化輸出所組成。由於本研究第一部分主要係透過自然語言處理預訓練的模型，對致股東報告書進行文字情緒分析，目的是使機器能夠像人一般理解致股東報告書，以解決過往文字探勘方法的缺點；而這樣類型分析屬於監督式學習的範疇，因此以下將先介紹監督式學習的基本概念。

監督式學習普遍包含四個主要部分，分別為標記資料、特徵、模型結構以及可區分的訓練及測試資料，以下以本研究資料作為例子：

- (1) 標記資料：致股東報告書句子正向、負向和中立情緒；
- (2) 特徵：致股東報告書的句子；
- (3) 模型結構：類神經網路；
- (4) 可區分訓練及測試資料；

其中，標記資料是耗費最多時間以及人力的環節，因為其需要人工判讀文章，以正確地標記句子的情緒。本研究雖與 Siano and Wysocki (2020)同樣是用 BERT 模型訓練樣本，但與 Siano and Wysocki (2020)不同的地方是標記資料的方法，本研究係透過人工蒐集及閱讀方式標記句子的情緒，但 Siano and Wysocki (2020)係依據每年同季別的銷售總額變化比率標記，若該年與去年同季的銷售總額變化比率大於該年與去年同季銷售總額變化比率的中位數，則標記文章為正向，而其他樣本(季度銷售總額變化小於或等於該季度樣本銷售總額中位數)則標記為負向。上述做法，使他們擁以下之優點：

- (1) 大量地減少資料標記時間；
- (2) 取得平衡資料；
- (3) 仍達到非常高之準確率；

而本研究改變他們樣本標記的方法，原因是致股東報告書不只有總結去年的營運績效，還有揭露公司對公司未來發展的看法，因此即便得知該年與去年同季的銷售總額變化比率為正，而將樣本標註為正向，但公司管理階層可能對未來抱持悲觀

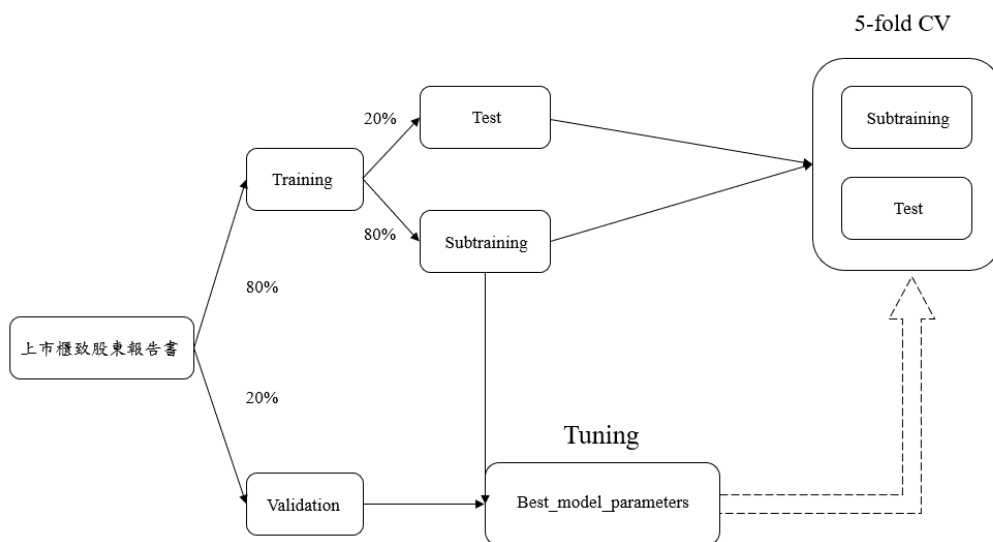
---

<sup>10</sup> 資料來源：[https://en.wikipedia.org/wiki/Supervised\\_learning](https://en.wikipedia.org/wiki/Supervised_learning)

的想法，故在致股東報告書揭露一些負面的消息；因此，本研究選擇透過人工標註的方式，仔細地閱讀每段句子，依專業判斷標註情緒，目的是為了讓機器學習時，能像人類般理解句子，進而讓準確率提高。本研究優點有人工標記資料較為準確，能應付管理階層預測明年營運走向與該年營運狀況相互違背的情況，例如致股東報告書對於未來表示樂觀，然而去年營運實際上卻表現慘澹，則該狀況於 Siano and Wysocki (2020)的作法會造成將實際上語調為樂觀的句子卻被標記成悲觀的句子，會造成嚴重失真的問題。

### 第三節 模型選擇與視覺化

本研究所採用之類神經網路係 BERT，該模型利用文字當作特徵以模擬語言結構，透過微調已經被訓練好的 BERT，讓模型能用更少的時間、更少的資料學習特定領域特性，以達到更好的效能。本研究參考 Siano and Wysocki (2020)的作法，並將此作法流程用圖三表示，隨機選取佔總樣本數 20% 句子當作驗證資料集 (validation)，並透過該驗證資料集的表現去選擇適當的超參數 (hyperparameter)，而剩下句子會再依 80% / 20% 比例分為子訓練資料集 (sub-training) 和測試資料集 (test) 進行交叉驗證，旨在避免拆分訓練資料集及測試資料集產生之偏誤，導致結果不可靠。



圖三、模型流程圖

#### (一) 模型選擇

為了決定將使用哪些學習率、Epochs 數量及 Batch Size 大小，本研究訓練許多

不同的模型，並評估其各自在驗證資料集的表現。本研究驗證資料集佔總樣本數的 20%，而剩下句子按 80%/20% 比例拆分為子訓練資料集及測試資料集，以平衡模型選擇所需之驗證資料集及評估結果的測試資料集數量。

本文在閱讀致股東報告書過程中，發現內容包含許多關於產品代號、專有名詞英文縮寫或是日常用語的英文，由於本研究所使用的是 BERT 中文預訓練模型，因此不適合處理太多英文及無意義的產品代碼。故將一部分英文字建立字典進行轉換，而長度大於 4 且不在字典中的辭彙將其轉換為『產品』，藉此幫助模型能夠更容易理解句子，預期該作法能有效地降低模型評估損失並提升各項衡量指標。因此於測試時，將在輸入的部分嘗試原始文字版本以及經過英文字轉換過後的版本。關於 Epochs 數量，由於本研究使用 Early Stopping<sup>11</sup> 的技巧，故先選擇 5 個 Epochs，較 Delvin et al. (2019) 推薦的 Epochs 數量再多一個，接著嘗試 Delvin et al. (2019) 推薦不同的學習率 2e-5、3e-5、5e-5，和其他不同的學習率 2e-4、1e-5、4e-5 和 2e-6；經過測試後，在 2e-5 的學習率下，嘗試 2~5 個的 Epochs；而 Batch Size 則嘗試了 8 個和 16 個。根據驗證集結果顯示，最好的輸入是經過英文字轉換過後的文本，而學習率是 2e-5、Epochs 為 2~3 個、Batch Size 為 16 個，並發現當 Epochs 大於三個或學習率大於 2e-5，模型便開始容易產生過擬合的情形，因此最後選擇讓驗證資料集損失最低的模型超參數。

## (二) 交叉驗證

當透過驗證資料集選擇出表現最好的超參數後，需要評估該模型在樣本外的測試。為避免拆分訓練集和測試集產生之偏誤，會做 5 折交叉驗證 (見圖三)。為什麼使用 5 折而不是更多折的原因是模型的訓練速度，當使用 5 折交叉驗證時，需要訓練五次模型，然而基於 BERT 模型非常龐大，導致訓練一次模型費時 25~40 分鐘，若使用更多折數的交叉驗證會使效率大打折扣，因此最後本研究決定使用 5 折交叉驗證。首先將驗證集資料外的樣本數隨機拆分成五組資料，當拿其中一組作為測試資料集時，其他四組作為訓練資料集，因此最終將產生五組模型評估損失和衡量指標，如準確率等，最後再將 5 次的模型評估損失及衡量指標平均，作為該模

---

<sup>11</sup> 在機器學習中，Early Stopping 是一種在使用梯度下降的疊代優化方法時，可對抗過擬合的正則化方法，在某個節點之前，訓練集使得模型在驗證集的數據上表現得更好；但在該節點之後，更多地訓練會增大誤差。在本研究中是以 validation loss 作為衡量基準。



型表現分數。

### (三) BERT 視覺化

眾所周知，深度學習模型就如同黑盒子一般，使用者將輸入及參數餵入模型，最後經過複雜的運算，得出一個結果，不免讓人懷疑或好奇模型究竟是關注到了什麼又或是怎麼去學習這些資料的。由於 BERT 的模型非常複雜，其所學習到的權重很難像一般的線性回歸 (Linear regression) 或羅吉斯回歸去解讀在不同特徵下其權重的意義。因此，資料科學家開發視覺化工具來幫助理解深度學習的模型。附錄一將介紹 BERTViz 的視覺化工具及其背後原理。最後將於實證結果說明 BERT 視覺化的結果。本研究預期 BERTViz 能夠透過視覺化方式，捕捉到否定詞與其修飾的詞彙，以解決字典法的否定詞解讀問題，例如：「營收未如預期成長」能夠捕捉到「未」是在形容成長，而導致該句話為負向語調；「原物料成本上漲」能夠捕捉到上漲是在形容原物料成本或是「虧損金額減少」能知道虧損減少是正向，而非因為「虧損」、「減少」這兩個負向詞彙，而判斷為負向。同時，同樣預期 BERT 能夠處理「指代消解」 (Coreference Resolution) 的問題，能夠找回原先被代名詞替換過的主詞，以幫助其理解上下文的關係。

### (四) 語境測試

為了證明 BERT 有考慮語言結構及語境的問題，並非像是字典法般只考慮相關情緒詞彙的出現頻率或是樸素貝氏演算法假設詞彙間獨立，而不考慮詞彙排列順序，本文參考 Siano and Wypocki (2020) 的作法，將測試資料集文字順序打亂。若 BERT 有考慮語言結構及語境的問題，這樣的作法預期會大幅降低模型的準確度。

### (五) 預測盈餘能力

本文參考 Li (2010a) 預測未來盈餘的作法及控制變數，因為 Li (2010a) 是利用樸素貝氏演算法預測前瞻性內容 (FLS) 的語調並結合財務資訊，預測未來公司績效，以判斷 FLS 是否有提供額外的資訊價值。本研究與 Li (2010a) 差別有三，一是分類語調的模型，本研究使用的 BERT 模型，而 Li (2010a) 是使用樸素貝氏演算法。二是在 Li (2010a) 使用的文本是包含於 MD&A 的 FLS，該文本有出現於 10-Q 和 10-K 的財報，因此其預測未來盈餘是以季度為單位。但在台灣，致股東報告書只有出現於年度財務報告中，故在本研究預測未來盈餘是以年度為單位。三是控制變數的數量，某些控制變數在 TEJ 資料庫並未找到，如 Fog 指數、在 Compustat 資料庫

非遺失項目數量 (non-missing items)、地理區隔 (geographic segment)、是否合併和是否二次發售的虛擬變數。本研究依變數為下一年的盈餘 (t+1 年)，語調作為本研究感興趣的自變數且包含其他控制變數，如：當年盈餘 (t 年)、股價年報酬率、應計項目、公司市值、股價淨值比、股價報酬率波動、盈餘波動、營運項目數量和成立年數。本研究預期語調與下一年的盈餘呈正相關，代表當管理階層對於公司營運保持樂觀，下一年盈餘應該會上升。

## 第四節 樣本選取

根據第一節所述之監督式學習所需具備的四個要素，第二節已經介紹完本研究所使用的模型結構以及如何拆分訓練資料集和測試資料集，本節將介紹如何蒐集監督式學習所需之特徵跟標記資料，這些文本皆來自於公開資訊觀測站的公司年度財務報告。

### (一) 樣本期間

本研究將樣本限縮在台灣半導體產業，原因是半導體產業多年來一直由政府有計劃地輔佐、推動加上業界自身穩健地經營，產業鏈可謂相當完整。半導體產業鏈包含上游的晶圓和 IC 設計業，中游的 IC 製造業及下游的 IC 封裝業和 IC 測試業，其中 IC 製造業主要以晶圓代工與 DRAM 製造為主。半導體產業的總產值佔我國 GDP 14%且出口比重高達 30%，為台灣經濟的支柱。半導體業的景氣時常隨著產品需求變化而產生興衰，因此若能從管理階層的角度瞭解目前的公司發展趨勢或是未來的需求變化，則可以預測未來公司的表現。本研究目的在透過 BERT 模型分析半導體產業的致股東報告書，有效率地瞭解管理階層所傳達的情緒，並結合其他財務資訊，以驗證致股東報告書與未來的盈餘表現的關聯，因此鎖定半導體業上市(櫃)之公司為研究對象。

表三顯示，2011 年至 2019 年之家數逐年下降，且近五年下降幅度更高達 10%；IC 產業產值則是從 2011 年開始年年成長，但成長率有起伏變化(參考圖四)，可以發現半導體業有受到國際情勢以及消費性電子產品發展產生景氣波動，根據工業技術研究院產業經濟與趨勢研究中心編寫的半導體產業與應用年鑑<sup>12</sup>，得知半導體景氣波動背後之原因，如 2010 年隨著金融風暴的平息，智慧型手機持續帶動手機

<sup>12</sup> <https://ieknet.iek.org.tw/book/BookListFree.aspx>

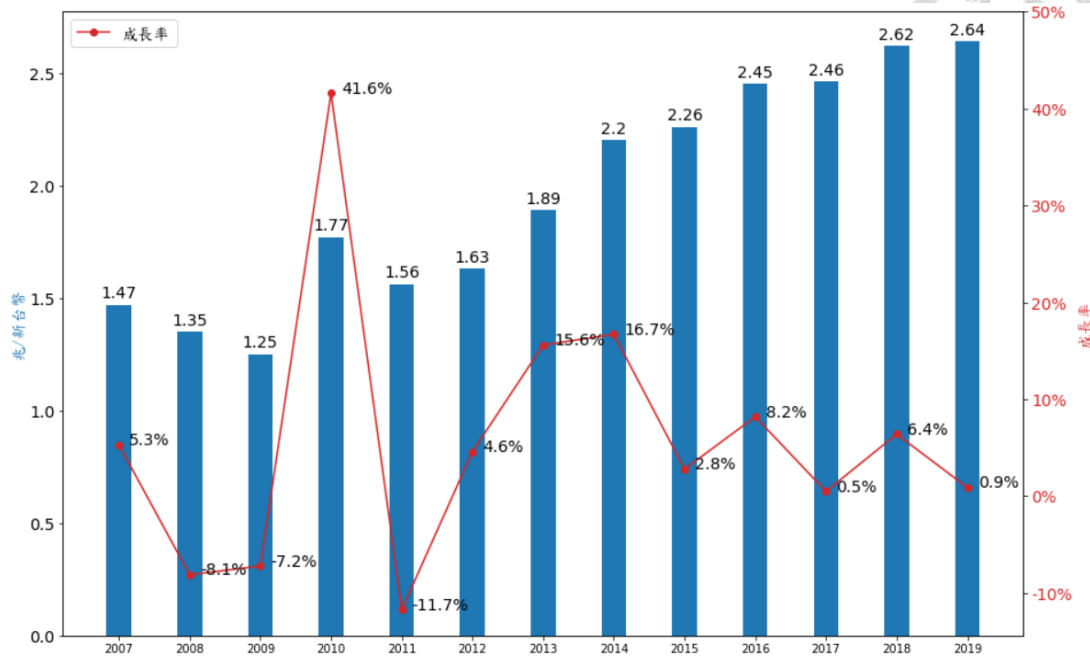
半導體晶片市場的需求，同時 DRAM 產業因為 PC 的快速成長而需求大幅增加，故 2010 年半導體產業蓬勃發展。但 2011 年上半年歷經了日本地震、美國舉債危機導致全球半導體設備與上游原材料供應不穩定，而下半年又面臨希臘債務違約風險、泰國水災等事件造成經濟震盪，以及取代桌機的趨勢導致 PC 成長力道逐漸疲軟，需求多轉往智慧型手機及平板電腦，導致許多公司處於一個青黃不接的時期。而 2012 年至 2014 年智慧型手機數量大幅增長、雲端商機崛起及物聯網應用使經濟回溫，IC 製造產業因記憶體價格復甦而在 2013 年產值大幅成長；晶圓代工的部分，則因台積電掌握 28 奈米先進製程領先優勢，對台灣 IC 製造產業有所助益。近幾年，記憶體市場力道枯竭，智慧型手機需求放緩，以及中美貿易戰加劇，導致全球半導體市場出現負成長，而台灣半導體市場同樣成長趨緩。

綜上所述，本研究將樣本期間鎖定於 2011 年至 2019 年，因為該期間內有一波景氣循環，能夠讓文本內所傳達情緒較為豐富，且本研究旨在分析 2011 至 2017 年的致股東報告書之文字資訊，並預測 2011 年至 2018 年的致股東報告書情緒，同時結合各年的財務資料，希望能幫助投資人成功地捕捉隔年(t+1 年)公司營運情形，本文未包含 2020 年的資料，以避免 COVID-19 所造成的不確定性，導致致股東報告書與實際情形相差太多。同時避免 COVID-19 為 2020 年資料帶來的不確定性。

表三、半導體產業家數、產值及成長率

年份	2011	2012	2013	2014	2015	2016	2017	2018	2019
家數	503	488	478	476	473	448	439	427	422
IC 產業產值 (兆/新台幣)	1.56	1.63	1.89	2.20	2.26	2.45	2.46	2.62	2.64
成長率(%)	-11.7	4.6	15.6	16.7	2.8	8.2	0.5	6.4	0.9

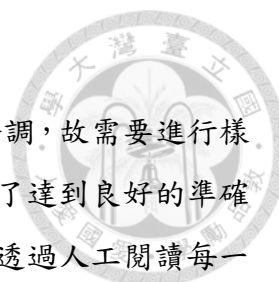
資料來源：財政部-財政統計月報、工業技術研究院(ISTI)、經濟部工業局智慧墊子產業計畫推動辦公室(SIPO)



圖四、台灣 IC 產業產值及成長率

## (二) 樣本篩選

本文樣本篩選步驟如下，首先，下載上市櫃半導體公司之年報，2011 年至 2018 年總共有 1,203 筆樣本，排除公開資訊觀測站無年報資料和致股東報告書係圖片類型的年報 218 筆，最後樣本剩 985 筆。其次，選取致股東報告書之前十行，且不包含標題及表格，一行是以一個句點或者是一個分號作為切割。選取前十行的原因如下：(1) 本文透過觀察致股東報告書，初步發現管理階層對產業概況的瞭解、公司前一年度之表現總結，以及公司未來發展之預期，主要集中於前段。因此透過選取致股東報告書的前十行句子，便能夠獲得豐富且有價值之資訊。(2) 避免主觀地篩選句子以及蒐集較有效率。本文排除標題之原因，是由於標題文字簡短，且通常會伴隨表格出現，若將其視為一句話，容易使前十行句子資訊含量降低；排除表格之原因，是由於各家公司之致股東報告書表格所揭露之事項不同，有些表格所包含的資訊對於投資人並不是那麼有價值，例如：發展新產品的編號，且表格的文字資訊含量少或是已經揭露於前後文中。但本研究並未排除句子中之數字，主要原因係 Siano and Wysocki (2018) 發現在自然語言分析中，夾雜在句子裡的數字是具有資訊價值的。最後，一行句子可能會有太長的問題，如某些公司較少使用句點或分號斷句，這會導致模型訓練時間拉長或是超過模型句子長度的限制，故該部分是透過人工的方式根據連接詞而斷句，將一行句子限縮於 150 字以內。



### (三) 樣本標記

本研究需要分析致股東報告書中文字之正向、中立及負向語調，故需要進行樣本標記的動作。樣本標記本身耗時耗工，因為通常監督式學習為了達到良好的準確率，需要大量資料，這同時意味著需要大量的人工標記。本研究透過人工閱讀每一篇致股東報告書的前十行句子，以判別語調係屬於正向、負向或中立。本小節依下列五項，分別說明本研究觀察到之現象跟判斷標準，(1)當年度營業狀況 (2)連接詞 (3)研究發展 (4)未來展望 (5)無關。

首先關於當年度營業狀況，表達大致上分成兩種，一種係『本集團民國 XXX 年度全年營業收入淨額為新台幣 X,XXX,XXX 元，較前一年的 X,XXX,XXX 仟元增加 XX%；』，屬於有前後年比較的情況，這種句子會給予正向或負向之語調；而另一種為『本公司 XXX 年之營業收入為 X,XXX,XXX 元，稅後淨利為 X,XXX,XXX 元，稅後每股盈餘為 X.XX 元。』，沒有前後年之比較，只有單純敘述者，會給予中立之語調。主要原因係不知道該公司是進步、退步還是持平，故給予其中立之標籤。

有關連接詞，其代表轉折關係，前面負向語句連接後面正向語句，反之亦然，可以發現在某些情形下，後半段比前半段更具有強調意味，因此在這種情形下，情緒標籤將以後半段語句為主。倘若後半段語句較空泛，則會給予中立之語調，舉例如下：『雖然美中貿易與關稅的紛爭，造成政治及經濟上諸多不確定因素的影響，使得營運倍受挑戰，然而 XXX 自下半年庫存開始去化的驅動下，營收與毛利率皆較 XXX 年明顯成長與提升。』—正向

『展望 XXX 年的半導體銷售的成長預估 XX%達到 X,XXX 億美元的規模，可望重新回到成長的周期，惟 XXXX 年全球景氣依舊壟罩在中美貿易的不確定性，以及新冠肺炎疫情擴大，對電子業和半導體業供應鏈帶來挑戰，供貨和需求皆同步受到影響，也為 XXXX 年的復甦埋下不確定的因素。』—負向

『XXXX 年度是競爭激烈的一年，近年來國內封裝大廠持續擴廠，進而影響中小型封裝業之業績成長但經營團隊仍努力維持既有產品及彈性的製造排程，以市場區隔的方式，不斷創新研發新產品。』—中立

關於研究發展，公司時常揭露『持續研發產品先進製程技術，增加產品競爭力。』或是『持續降低生產成本。』類型的空泛句子，目前給予中立之語調，原因係觀察



這幾年的致股東報告書，發現有許多公司充斥著這樣類似句子，這樣的語句對於投資者沒有任何實質的參考價值。目前排除『開發 XXX 產品。』等類似語句，排除原因一是某些公司產品名稱太過專業，投資者難以瞭解該產品是否有市場或效益；原因二是通常這類句子常常列點出現，導致這家公司可能有前十段文字就有四、五行是這種句子，而這類句子同樣對未來公司營運預測沒有太大的實質幫助，因此盡量將這類型的句子排除在外。

有關未來展望，有些公司會於未來營業計劃或未來展望述說公司的規劃，然而如研究發展狀況一樣，不免會出現『拓展營業接單策略，積極開發新客戶。』、『持續創新產品開發，提供多元式產品封裝製程，滿足客戶需求。』等空洞無物的精神喊話，因此這類型的句子處理方式與研究發展一樣，皆給予中立之語調。

關於無關，致股東報告書有時會出現『成功者找方法』、『只為成功找方法，不為失敗找藉口』等董事長信心喊話，抑或是『由衷感謝股東的長期愛護肯定、客戶的認同信任及供應商的全力支持。』等感謝的話，將這類型的句子標記為中立。

經過前述之資料蒐集、處理和標記後，再將這些人工標記句子拆分為訓練、驗證和測試資料集，結果如表四。表四顯示，訓練資料集共有 5,239 行句子，其中 2,436 行 (47%) 為正向；843 行 (16%) 為負向；1,933 行 (37%) 為中立。驗證資料集有 1,638 行句子，其中 748 行 (46%) 為正向；261 行 (16%) 為負向；629 行 (38%) 為中立。測試資料集有 1,310 行句子，其中 599 行 (46%) 為正向；199 行 (15%) 為負向；512 行 (39%) 為中立。

表四、致股東報告書句子拆分為訓練、驗證和測試資料集

	句子數
半導體上市櫃公司 (2011-2018) (公司年)	9,514
減：2018 年句子數	1,327
半導體上市櫃公司 (2011-2017)	8,187
減：驗證資料集 (validation) (20%)	1,638
訓練 (train)與測試資料集 (test) (80%)	6,549
訓練資料集 (佔訓練與測試資料集 80%)	5,239
-正向	2,463
-負向	843
-中立	1,933
測試資料集 (佔訓練與測試資料集 20%)	1,310
-正向	599
-負向	199
-中立	512

## 第四章 實證結果

### 第一節 模型分類結果



本節先介紹本研究評估模型之衡量指標，接著顯示放入不同輸入及參數其產生之模型結果，及透過驗證資料集篩選出最佳的超參數，放入 5-Fold CV，以確認模型沒有因為拆分樣本之偏誤，產生不穩定和不可靠的結果。

首先，本研究使用混淆矩陣(confusion matrix)評估模型的預測表現，運用指標包括：Accuracy、Precision、Recall 和 F1-Score，如表五所示：

表五、混淆矩陣

	實際為正	實際為負
預測為正	True Positive (TP)	False Positive (FP)
預測為負	False Negative (FN)	True Negative (TN)

True Positive (TP)代表模型預測語調與實際語調一致為正向語調；False Positive (FP)代表實際語調係負向但被模型判斷為正向；False Negative (FN)代表實際語調係正向但被模型判斷為負向；True Negative (TN)代表模型預測語調與實際語調一致為負向語調。Accuracy 為衡量被正確地分類樣本(TP+TN)佔所有樣本比例，但該指標在不平衡資料情況下，容易造成準確率高估，如：模型全部猜標籤較多的一方，會比隨機猜準確率高，但是模型實際上並未良好地學習；Precision 代表預測正向語調，實際上確實是正向語調( $TP / (TP + FP)$ )。Recall 代表實際為正向語調，模型成功預測為正向語調之比例( $TP / (TP + FN)$ )；F1-Score 則為綜合考量 Precision 及 Recall 之衡量指標( $2 * Precision * Recall / (Precision + Recall)$ )。由於本研究係多類別分類問題，因此將 Precision 表示為多類別形式。給定一個 k 類別的語調，分母為模型根據文字  $x_i$  預測語調為 k 的數量，分子為實際語調與預測語調同樣為 k 的數量，相除後可得模型預測 k 類別正確的比例，如下等式(2)：

$$\sum_{i=1}^N \frac{y_{i,k} \cap h(x_i)}{[h(x_i) = k]}, h(x_i) \text{ 是預測值}, y_{i,k} \text{ 是實際為第 } k \text{ 類別} \quad (2)$$

Recall 的多類別形式是給定一個 k 類別的語調，分母為實際語調為 k 的數量，分子為實際語調與預測語調同樣為 k 的數量，相除後可得實際為 k 類別數量，模型

能正確「召回」多少實際為 k 類別的數量，如下等式(3)：

$$\sum_{i=1}^N \frac{y_{i,k} \cap h(x_i)}{y_{i,k}}, h(x_i) \text{ 是預測值}, y_{i,k} \text{ 是實際為第 } k \text{ 類別} \quad (3)$$



### (一) 模型選擇

將致股東報告書之原始文字資料放入 BERT 模型中進行訓練，根據樣本標記所述之情形，可以知道有某些句子係因上下句子剛好為轉折關係，且其後半段句子可能較為空泛，抑或是句子本身正負向語調不強烈，因此被歸類為中立。這些涉及的人為主觀判斷，皆可能導致模型判斷不準確，故預測中立語調的衡量指標會略低於預測正向或負向的衡量指標。實證結果如表六。該表顯示，BERT 模型在驗證資料集的評估結果，可以發現經過 3 個 epochs 及 2e-5 的學習率訓練，Evaluation loss<sup>13</sup> 為 0.33，而 Evaluation Accuracy 為 0.87，中立的 Precision 為 0.85，中立的 Recall 為 0.85，中立的 F1-Score 係 0.85；而正向的 Precision 為 0.88，Recall 為 0.88，F1-Score 為 0.88；負向的 Precision 為 0.91，Recall 為 0.90，F1-Score 為 0.91。從結果可以發現，中立的 Precision 指標與 Recall 指標皆略輸於正向及負向的 Precision 指標與 Recall 指標，與預期的結果一致，因此抽取幾個實際為中立，模型卻將其預測為正向或負向的句子，觀察這些句子是怎麼讓模型判斷錯誤的。舉例而言：「某公司 104 年度營收淨額為新台幣 12 億 7 千 7 佰萬元，稅後淨利約為 8 千 8 佰萬元，相較於 103 年度，營收增加近 0.58%，稅後淨利則減少 8 佰 8 拾萬元，稅後 EPS 為 2.99 元。」該句話標記實際為中立，原因是因為稅後淨利減少與營收增加相抵，但模型卻將其判斷為正向語調。「2016 年通訊市場的成長與 PC 市場的衰退仍舊形成一種互相抵消的狀態，無線通訊營收在智慧型手機與記憶體的帶動下，成長 9.6%，可是 PC 與平板電腦的半導體市場卻衰退 8.3%。」該句話標記實際為中立，因為通訊市場的成長與 PC 與平板電腦衰退相抵，但模型卻將此句判斷為負向語調。透過觀察這些模型判斷錯誤的例子，可以發現，模型容易在轉折語氣不強烈的語句下判斷錯誤。

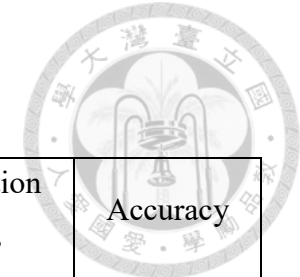
再來，表七是將致股東報告書之原始資料作英文字轉換後餵入模型，模型於驗

<sup>13</sup> Evaluation loss 係指真實值與模型之預測值不一致，透過函數之轉換產生之損失。該差額可以透過各式各樣的函數進行轉換。本研究使用之損失函數係 cross-entropy。由於類神經網路目的係透過最小化損失，以達到良好的訓練，因此較高損失代表其模型分類準確率較低。

證資料集的表現。在經過 2 個 epochs 及  $2e-5$  的學習率訓練，Evaluation loss 下降至 0.31，而 Evaluation Accuracy 上升至 0.89，相比於致股東報告書原始文字，表現些微地提升。中立的 Precision 為 0.88，中立的 Recall 為 0.83，中立的 F1-Score 係 0.86；而正向的 Precision 為 0.88，Recall 為 0.91，F1-Score 為 0.89；負向的 Precision 為 0.90，Recall 為 0.91，F1-Score 為 0.91，相比於原始資料，各個衡量指標有些微地提升，因此，最後選擇輸入為前處理後的致股東報告書，超參數的選擇為 2 個 epochs 及  $2e-5$  的學習率，接著進行交叉驗證。

## (二) 交叉驗證

經過 5 折交叉驗證後，其 5 次平均結果如表八、經前處理的致股東報告書於 5-Fold CV 平均結果表八所示，Evaluation loss 為 0.38，而 Evaluation Accuracy 為 0.86，而中立的 Precision 為 0.85，中立的 Recall 為 0.83，中立的 F1-Score 係 0.84；而正向的 Precision 為 0.85，Recall 為 0.89，F1-Score 為 0.87；負向的 Precision 為 0.89，Recall 為 0.90，F1-Score 為 0.90，結果正如預期一樣，預測中立語句涉及較多人為主觀判斷，因此容易導致模型預測不準確，中立 F1-Score 指標低於正向與負向的 F1-Score 指標。整體來說，模型準確率高達 86%，是一個不錯的結果。同時可以發現，這裡模型預測正向語句與預測負向語句的表現相差不多，與 Loughran and McDonald (2016) 指出管理階層較常使用多個正向辭彙包裝一句負向語句，反而較少使用雙重否定表示肯定，因此通常預測負向語句較預測正向語句更為準確的結果有所出入，因此將在第二節透過視覺化分析 BERT 是否能成功捕捉到否定詞去形容正向辭彙使模型提升預測這類反轉語句的情緒。

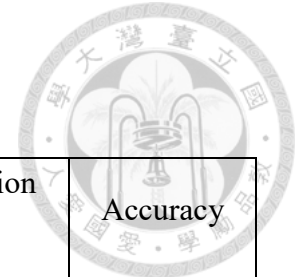


表六、致股東報告書原始文字於驗證資料集的結果

驗證資料集結果	訓練資料集 (n=5,239)			驗證資料集 (n=1,638)			測試資料集 (n=1,310)			Evaluation Loss	Accuracy
	正向	中立	負向	正向	中立	負向	正向	中立	負向		
輸入： - 致股東報告書原始文字	2,463 (0.47)	1,933 (0.37)	843 (0.16)	748 (0.46)	629 (0.38)	261 (0.16)				0.33	0.87
BERT 參數 - BERT_base_chinese 2e-5 學習率 2 個 epochs	Precision			Recall			F1-Score				
	正向	中立	負向	正向	中立	負向	正向	中立	負向		
	0.88	0.85	0.91	0.88	0.85	0.90	0.88	0.85	0.91		

表七、經前處理的致股東報告書於驗證資料集的結果

驗證資料集結果	訓練資料集 (n=5,239)			驗證資料集 (n=1,638)			測試資料集 (n=1,310)			Evaluation Loss	Accuracy
	正向	中立	負向	正向	中立	負向	正向	中立	負向		
輸入： -致股東報告書前處理	2,463 (0.47)	1,933 (0.37)	843 (0.16)	748 (0.46)	629 (0.38)	261 (0.16)				0.31	0.89
BERT 參數	Precision			Recall			F1-Score				
-	正向	中立	負向	正向	中立	負向	正向	中立	負向		
BERT_base_chinese 2e-5 學習率 2 個 epochs	0.88	0.88	0.90	0.91	0.83	0.91	0.89	0.86	0.91		



表八、經前處理的致股東報告書於 5-Fold CV 平均結果

5-Fold CV	訓練資料集 (n=5,239)			驗證資料集 (n=1,638)			測試資料集 (n=1,310)			Evaluation Loss	Accuracy
	正向	中立	負向	正向	中立	負向	正向	中立	負向		
輸入： -致股東報告書前處理	2,463 (0.47)	1,933 (0.37)	843 (0.16)				599 (0.46)	512 (0.39)	199 (0.15)	0.38	0.86
BERT 參數 -基於 Validation 所 選擇之最佳參數	Precision			Recall			F1-Score				
	正向	中立	負向	正向	中立	負向	正向	中立	負向		
	0.85	0.85	0.89	0.89	0.83	0.90	0.87	0.84	0.90		

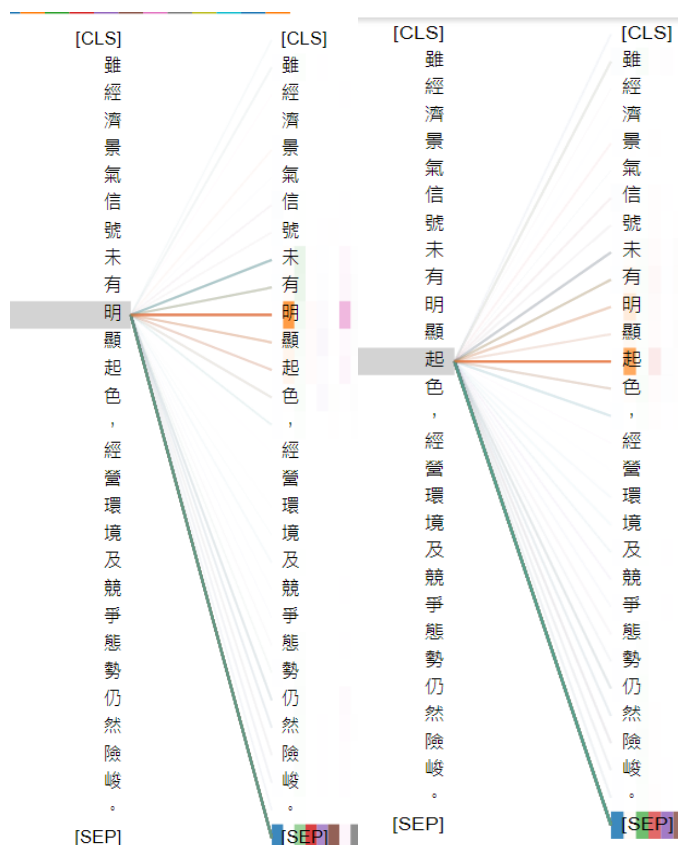




## 第二節 視覺化與語境測試

### (一) BERT 視覺化

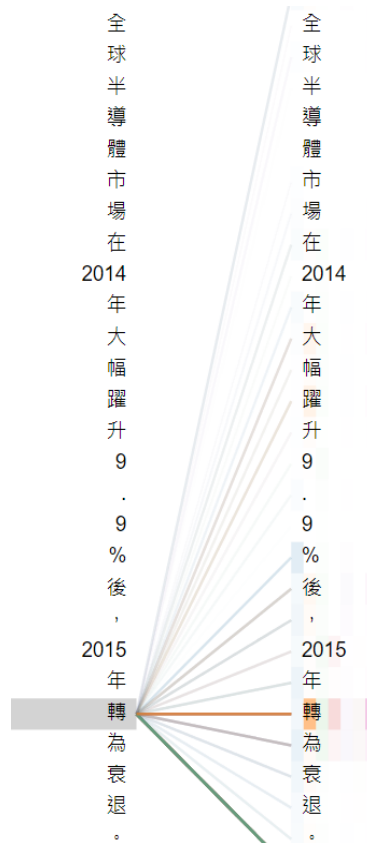
如研究設計所述，本節將模型預測正確的語句取出，透過 BertViz 的套件進行視覺化分析，下圖中的線條代表 head<sup>14</sup>在更新其(左側)詞向量時，關注其他詞彙(右側)的注意力程度。



圖五、BERT 視覺化(1)

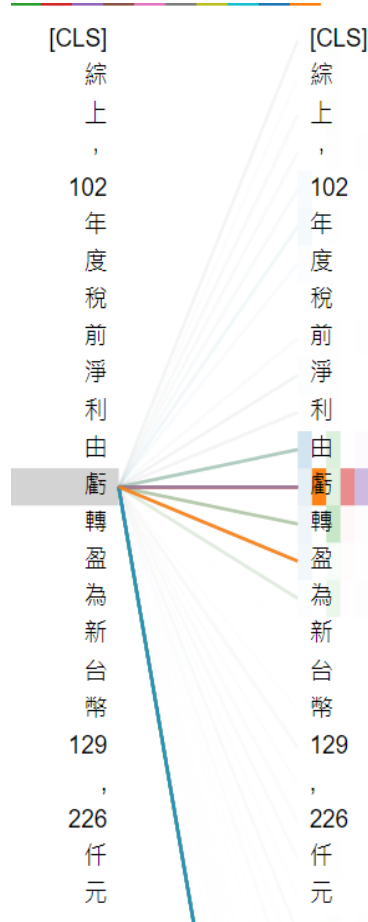
根據圖五，兩張張圖顯示『雖經濟景氣信號未有明顯起色，經營環境及競爭態勢仍然險峻。』，可以發現 BERT 能辨認『未有明顯起色』，知道該否定詞『未』與『明顯起色』相關聯，能夠將否定詞與其形容的正向形容詞組合一起，理解該句是負向語句，解決字典法中無法辨識否定詞在形容哪個詞彙的難題；

<sup>14</sup> 參考附錄一



圖六、BERT 視覺化(2)

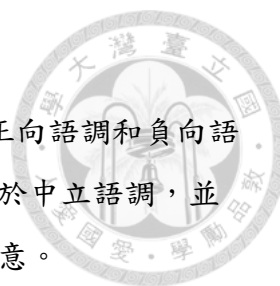
根據圖六，該句是說『全球半導體市場在 2014 年大幅躍升 9.9%後，2015 年轉為衰退』，這句話裏頭的『轉』是扮演轉折詞的效果，是形容『大幅躍升』轉為『衰退』，而該句牽涉上下文的結構，如『大幅躍升』及『衰退』對調位置的話，語句情緒便完全相反，而字典法及樸素貝氏演算法是無法理解這樣的上下文轉折的情境，但 BERT 卻能理解這樣複雜的語言結構，實在是一件令人雀躍的事情。



圖七、BERT 視覺化(3)

根據圖七，『綜上，102 年度稅前淨利由虧轉盈為新台幣 129,226 仟元，基本每股稅後盈餘為新台幣 1 元。』，BERT 將『由虧轉盈』看在一起，成功理解『轉』扮演在該句扮演轉折的角色，並不會因為一個『虧』和一個『盈』而將此句預測為中立。

根據上面幾個例子，可以發現 BERT 透過複雜的運算邏輯和龐大的參數，能夠將否定詞與其形容的形容詞綁在一起，或是將轉折詞與上下文連結在一起，這些步驟是以往字典法或樸素貝氏演算法做不到的事，在本研究的交叉驗證結果同樣地證明在將語調區分為正向、中立和負向的中文文本中，預測正向語調的 Precision 與負向語調的 Precision 結果是差不多的，與 Loughran and McDonald (2016)及 Siano and Wysocki (2020)的實證結果顯示正向衡量指標明顯低於負向衡量指標的結果有所差異。推測與 Loughran and McDonald 差異源於 BERT 本身能夠處理語言結構的問題，並不像字典法只能依照出現的詞彙作為語調的判斷，容易導致負向語調的句子因為否定詞判斷不精確而被模型判斷為正向。與 Siano and Wysocki (2020) 的實



證結果差異可能來自於下列三個原因：

- (1) 語調類別數目：Siano and Wysocki (2020)將句子語調區分為正向語調和負向語調，但在閱讀致股東報告書中，可以發現許多句子其實係屬於中立語調，並不適用於二分法，因此可能導致預測正向語調的表現差強人意。
- (2) 標記類別的方法：在研究設計章節裡，有特別敘述本研究與 Siano and Wysocki (2020)標記樣本方法的不同，其中提到他們做法可能導致標記與句子本身的意義脫離，進而導致預測不準確的結果。
- (3) 致股東報告書的本質：本研究瀏覽近千份的致股東報告書，觀察到特別的現象，即大部分公司前後幾年的致股東報告書，表達方式非常相近，甚至有些是一模一樣的句子，若這類型樣本同時出現於訓練資料集和測試資料集，會導致模型訓練完後，能夠更準確地預測該句的語調。

## (二) 語境測試

表九顯示語境測試結果。在將句子隨機打亂後，進行交叉驗證，得到平均的 Evaluation Loss 為 1.11，Evaluation Accuracy 為 0.47，中立的 Precision 為 0.43，中立的 Recall 為 0.97，中立的 F1-Score 係 0.60；而正向的 Precision 為 0.87，Recall 為 0.15，F1-Score 為 0.26；負向的 Precision 為 0.83，Recall 為 0.18，F1-Score 為 0.30，顯示將語句隨機打亂後，BERT 的預測衡量指標下降許多，Loss 從平均 0.38 上升至 1.11，而 Accuracy 從 0.86 下降至 0.47，其他中立、正向及負向的 F1-Score 原本皆為 0.80 以上，都分別下降至 0.60、0.26 及 0.30，從上述的指標來看，若 BERT 模型如樸素貝氏演算法未考慮文字順序，則其各項指標不應隨著文字順序打亂而大幅下降，因此可得知模型的確有考慮語言結構及語境問題。

表九、隨機打亂的致股東報告書於 5-Fold CV 平均結果

5-Fold CV	訓練資料集 (n=5,239)			驗證資料集 (n=1,638)			測試資料集 (n=1,310)			Evaluation Loss	Accuracy
	正向	中立	負向	正向	中立	負向	正向	中立	負向		
輸入： -致股東報告書前 處理隨機打亂	2,463 (0.47)	1,933 (0.37)	843 (0.16)				599 (0.46)	512 (0.39)	199 (0.15)	1.11	0.47
BERT 參數 -基於 Validation 所 選擇之最佳參數	Precision			Recall			F1-Score				
	正向	中立	負向	正向	中立	負向	正向	中立	負向		
	0.87	0.43	0.83	0.15	0.97	0.18	0.26	0.60	0.30		



### 第三節 預測盈餘能力

#### (一) 敘述性統計

本研究將利用 Validation 所選擇的超參數結合 2011 年至 2017 年的句子，重新訓練一個模型，進一步去預測 2011 年至 2018 年的致股東報告書。若模型判斷第 k 個句子語調為中立，則標記為 0，若為正向則標記為+1；若為負向則標記為-1。最後將第 i 個公司致股東報告書情緒定義為該公司的所有句子語調加總平均。

$$\text{語調}_i = \frac{1}{K} \sum_{k=1}^K \text{語調}_{i,k}, K = \text{該公司總句子數}$$

公司語調將介於+1 於-1 間，若公司語調為+1，則代表其致股東報告書所有句子皆為正向，反之亦然。因此可知，當語調愈高，其致股東報告書愈正向。表十顯示 2011 年至 2018 年底各變數的平均數和分位數。平均來說，這 8 年的致股東報告書普遍較為樂觀，平均數為 0.36，而中位數為 0.40，且語調平均數與 0 顯著不同，在 t-test 的 p 值為 0.000。此 8 年的年報酬率平均為 0.12 且中位數為 0.02，由此可知這 8 年的半導體的股價普遍是往上漲的。

表十、敘述性統計—盈餘預測之相關變數

變數	平均數	pr(=0)	Q1	中位數	Q3	標準差
語調	0.36	0.000	0.10	0.40	0.67	0.38
本期淨利	0.03		-0.01	0.05	0.10	0.15
年報酬率	0.12		-0.22	0.02	0.30	0.57
市值	15.27		14.09	15.21	16.19	1.65
股價波動	0.12		0.08	0.10	0.14	0.07
淨利波動	0.06		0.02	0.04	0.07	0.07
應計項目	-0.05		-0.11	-0.05	-0.01	0.12
股價淨值比	1.72		1.01	1.33	1.85	1.49
營運項目	1.06		0.69	1.10	1.39	0.39
成立年數	20.23		15.00	19.00	25.00	8.41

註：該表格顯示 2011 至 2018 年共有 957 個觀察值。欄位 pr (=0) 是 t-test 的 p 值，是檢驗該變數是否等於 0。接下來將解釋變數的定義：語調是 2011 年至 2018 年的致股東報告書語調平均。本期淨利是選自 TEJ 合併總損益欄位並除以當年底資產總額。年報酬率是選自 TEJ 年報酬率 (%) 欄位並除以 100。市值是選自 TEJ 季底普通股市值欄位並做自然對數的轉換。股價波動是根據 TEJ 2011 年至 2018 年，計算各年 12 個月的股價月報酬率標準差而獲得。淨利波動是以五年為一窗期，將五年的合併總損益除以其各自年份的資產總額，並取標準差。應計項目是透過 TEJ 合併總損益欄位減來自營運活動的現金流量欄位，並除以資產總額。股價淨值比是每年季底普通股市值加上負債總額並除以資產總額計算而得。營運項目是選自 TEJ 的營運項目(財報)並計算其營運項目的數量。成立年數是公司的成立年數，選自 TEJ 的成立年數欄位。

## (二) 回歸結果

首先，將 BERT 模型預測的語調與下一年淨利跑 Pearson correlation 發現，語調的確和下一年的淨利呈現顯著正相關，相關係數為 0.2992，P 值為 0.000。接著將語調結合其他控制變數對下一年淨利作線性回歸。根據表十一結果顯示，本研究所關注的語調與下一年淨利關係並不如預期，雖然係數為正向的 0.0104，但 P 值為 0.188，並沒有呈現顯著的關係。而 Li (2010a) 的研究結果顯示美國 MD&A 語調與隔年淨利呈顯著正相關，其他控制變數係數方向大致與本研究的實證結果相同，例如：本期淨利和年報酬率與下一年的淨利呈現顯著正相關，或是應計項目與未來績效呈現顯著負相關。因此，本研究將嘗試說明語調與下一年淨利不顯著的原因。

表十一、語調預測隔年淨利之線性回歸結果—BERT

依變數：下一年淨利				
變數	係數	標準誤差	t 值	p 值
語調	0.0104	0.008	1.317	0.188
本期淨利	0.4714***	0.029	16.414	0.000
年報酬率	0.0231***	0.006	3.638	0.000
市值	0.0095***	0.002	4.875	0.000
股價波動	-0.0005	0.046	-0.011	0.991
淨利波動	-0.0088	0.049	-0.182	0.856
應計項目	-0.1587***	0.027	-5.838	0.000
股價淨值比	0.0144***	0.002	6.808	0.000
營運項目	-0.0026	0.008	-0.338	0.736
成立年數	-0.0012***	0.000	-3.332	0.001
2012	0.0188*	0.011	1.649	0.099
2013	0.0205*	0.012	1.720	0.086
2014	-0.0101	0.012	-0.864	0.388
2015	0.0294***	0.011	2.688	0.007
2016	0.0190	0.012	1.638	0.102
2017	0.0039	0.012	0.319	0.750
2018	0.0164	0.011	1.461	0.144
樣本數	957	R-squared	0.55	

註：該表顯示致股東報告書語調和其他控制變數與下期淨利的線性回歸結果。依變數為下一年(t+1)淨利除以當年年底(t)資產總額。  
 \*\*\* $p < 0.01$ ; \*\* $p < 0.05$ ; \* $p < 0.1$

首先，檢查語調與下一年淨利不顯著的原因是否來自於 BERT 預測致股東報告書的語調不準確而導致，故將檢驗 BERT 模型預測 2011 年至 2018 年的致股東報告書準確率，以確認不顯著原因是否來自於 BERT 模型的準確度太低。但經過 BERT 模型與人工判讀這 8 年的致股東報告書的結果比較後，可以發現 BERT 模型準確率高達 90%；本研究決定進一步使用人工判讀的語調，再跑一次線性回歸，結



果如表十二，結果語調與下一年淨利同樣沒有呈顯著相關。因此可以排除是因為BERT模型的準確率低才導致語調與下一年淨利不顯著。再來，推測可能造成的原因是，從致股東報告書選擇前十行句子的方式過於單薄，理想的作法是將致股東報告書的文字資訊，隨機選取10行或15行，再進行訓練及預測。最後，可能是台灣的致股東報告書及美國MD&A本身資訊含量的差異，導致語調和隔年淨利結果本身並不呈顯著關聯。

表十二、語調預測隔年淨利之線性回歸結果—人工

依變數：下一年淨利				
變數	係數	標準誤差	t 值	p 值
語調	0.0094	0.008	1.198	0.231
本期淨利	0.4708***	0.029	16.379	0.000
年報酬率	0.0232***	0.006	3.650	0.000
市值	0.0096***	0.002	4.933	0.000
股價波動	-0.0012	0.046	-0.027	0.979
淨利波動	-0.0095	0.049	-0.194	0.846
應計項目	-0.1580***	0.027	-5.817	0.000
股價淨值比	0.0145***	0.002	6.825	0.000
營運項目	-0.0023	0.008	-0.306	0.760
成立年數	-0.0012***	0.000	-3.351	0.001
2012	0.0186	0.011	1.627	0.104
2013	0.0204*	0.012	1.714	0.087
2014	-0.0102	0.012	-0.869	0.385
2015	0.0293***	0.011	2.671	0.008
2016	0.0187	0.012	1.612	0.107
2017	0.0032	0.012	0.259	0.796
2018	0.0160	0.011	1.421	0.156
樣本數	957	R-squared	0.55	

註：該表顯示人工判讀的致股東報告書語調和其他控制變數與下期淨利的線性回歸結果。依變數為下一年(t+1)淨利除以當年年底(t)資產總額。

\*\*\* $p < 0.01$ ; \*\* $p < 0.05$ ; \* $p < 0.1$

## 第四節 額外測試

本研究主要是探討 BERT 在中文致股東報告書的表現，然而前文提及不同的文字探勘方法，因此將在此節利用中研院中文斷詞系統<sup>15</sup>配合字典法衡量句子的情緒，作為 BERT 模型的比較基準。首先，樣本與 BERT 樣本同樣是來自於 2011 年至 2017 年的句子，接著透過中文斷詞系統斷詞，並以人工方式標記各個詞彙的情緒，正向為+1、中立為 0 及負向為-1，作成一情緒辭典。最後，分別使用字典法以及 TF-IDF (Term Frequency-Inverse Document Frequency)和 BERT 比較情緒分析的準確率。

### (一) 字典法

本研究參考 Henry (2008) 的作法，該篇論文計算情緒分數的方式為假設每個詞彙權重皆為 1，並把句子正向詞彙數量減除負向詞彙數量，並除以正向及負向詞彙的總數，即是該句子的情緒分數。由於本研究包含中立的情緒，因此分母部分是除以正向、中立和負向的總數，且本研究將正向情緒標記為+1，中立情緒標記為 0 以及負向情緒標記為-1，因此將所有辭彙情緒相加即可。計算每行句子的情緒分數後，本研究將句子情緒分數取四分位數，如表十三所示，並依照驗證資料集的表現，選取適合的分位數作為區分正向、中立及負向的門檻。本研究位最後選擇中位數和第一四分位數作為拆分門檻。

表十三、情緒分數敘述性統計—字典法

句子數	平均數	標準差	最小值	Q <sub>1</sub>	Q <sub>2</sub>	Q <sub>3</sub>	最大值
8,177	0.068	0.148	-1.000	0.000	0.076	0.143	1.000

將測試資料集的句子情緒分數低於Q<sub>1</sub>標記為負向，而介於Q<sub>1</sub>與Q<sub>2</sub>的句子標記為中立，剩下則為正向，得到的結果如表十四所示，發現字典法準確率僅達到 0.60，尤其在中立的句子上表現得差強人意。但值得注意的是，根據前述 BERT 的情緒分析結果，正向和負向語句的 Precision、Recall 皆高達 0.85 以上，而在字典法的結果中，可以發現正向與負向語句的衡量指標皆下降至 0.60 ~ 0.70 的範圍內。因此本研究將從測試資料集挑選出字典法預測負向語句，但 BERT 與實際情緒為正向或

<sup>15</sup> <http://ckipsvr.iis.sinica.edu.tw/>

中立的情形，作為比較這兩種文字探勘方法差異的實際範例。

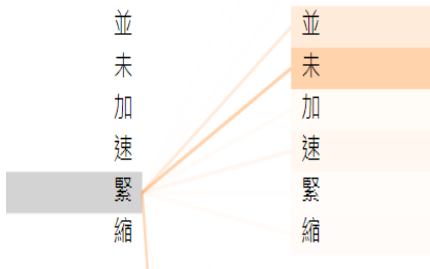
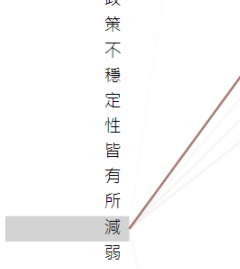


表十四、字典法情緒分析結果

句子數	Accuracy	Precision			Recall			F1-Score		
		正向	中立	負向	正向	中立	負向	正向	中立	負向
1,310	0.60	0.63	0.53	0.66	0.70	0.46	0.70	0.66	0.49	0.68

表十五、字典法與 BERT 之比較

範例一	而 XX 也成功的扭轉逆境，邁向新局：完全脫離了 Elpida 破產的陰霾。
字典法	將『成功』、『扭轉』、『邁向』視為正向詞彙，而『逆境』、『脫離』、『破產』、『陰霾』視為負向詞彙，剩下詞彙皆為中立。由於負向詞彙多於正向詞彙導致情緒分數小於 0.000，因此字典法將其判斷為負向語句。
BERT	<p>圖中可以發現，詞彙之間的修飾能夠成功被 BERT 捕捉，然而字典法卻只依照詞彙意義分辨情緒，容易曲解本身句子傳達的內容。</p>
範例二	主要國家貨幣政策並未加速緊縮，以及已開發中國家及新興市場國家的政策不穩定性皆有所減弱，是支撐 2017 年全球經濟得以穩健復甦的重要關鍵。

字典法	將『支撐』、『復甦』視為正向詞彙，而『緊縮』、『不穩定性』、『減弱』視為負向詞彙，剩下詞彙皆為中立。此句同樣是源於負向詞彙多於正向詞彙導致字典法將其判斷為負向語句。			
BERT	<p>[CLS]</p> <p>主 要 國 家 貨 幣 政 策 並 未 加 速 緊 縮 ，</p>	<p>[CLS]</p> <p>主 要 國 家 貨 幣 政 策 並 未 加 速 緊 縮 ，</p> 	<p>已 開 發 中 國 家 及 新 興 市 場 國 家 的 政 策 不 穩 定 性 皆 有 所 減 弱 ，</p> 	<p>已 開 發 中 國 家 及 新 興 市 場 國 家 的 政 策 不 穩 定 性 皆 有 所 減 弱 ，</p>
<p>圖中，可以發現 BERT 有捕捉到『並未』這個否定詞是在形容『緊縮』，而『減弱』這個動詞是在形容『不穩定性』，這兩種情況說明句子中時常存在負負得正的情形，而 BERT 能夠成功將詞彙之間的關係串聯，字典法卻需要一些更深入的研究。</p>				

## (二) TF-IDF

本研究參考 Loughran and McDonald (2011) 的作法，該篇論文指出單純計算某一情緒詞彙出現的數量，容易受到句子長度的影響，導致使用計算詞彙數量的方式衡量文字資訊的內容是有缺陷的。因此，作者說明考慮權重的作法以及权重帶來的優勢，第一是衡量詞彙的重要性，通常以詞彙出現的比率衡量；第二是對句子長度去規模化，擺脫以往計算數量時，容易受到句子的長度而影響；第三是衡量詞彙在整個文本中的獨特性，例如『公司』經常出現於不同家公司的致股東報告書，然而並非每間公司營運狀況皆會『上升』，因此『上升』被提及的次數相比於『公司』被提及的次數來得更少，因此 TF-IDF 將判斷『上升』比『公司』更為重要。本研究使用 TF-IDF 的定義為考量某一給定詞彙 (i) 在某個句子 (j) 出現的頻率 ( $tf_{i,j}$ )，



和包含該詞彙的句子數量 ( $df_i$ )，計算出每個詞彙在各個句子相對應的權重( $w_{i,j}$ )，權重公式如(4)所示：

$$w_{i,j} = tf_{i,j} * idf_i \quad (4)$$

權重要素為：

(1)  $tf_{i,j}$ ：一給定詞彙 ( $i$ ) 在某個句子 ( $j$ ) 出現的頻率，代表該詞彙的重要性。

(2)  $idf_i$ ： $\log\left(\frac{N}{df_i}\right)$ ， $N$  為句子的數量， $df_i$  為包含該詞彙  $i$  的句子數量。

考慮權重後，將每個句子 ( $j$ ) 的情緒分數計算公式如(5)所示：

$$\frac{\left(\sum w_{i,j} * \left[ [i = \text{正向詞彙}] \right] - \sum w_{i,j} * \left[ [i = \text{負向詞彙}] \right] \right)}{\sum w_{i,j}} \quad (5)$$

最後，與字典法相同，取情緒分數的四分位數，如表十六，並同樣依據驗證資料集的表現，選取適合的分位數作為區分正向、中立及負向的門檻。

表十六、情緒分數敘述性統計-TF-IDF

句子數	平均數	標準差	最小值	Q <sub>1</sub>	Q <sub>2</sub>	Q <sub>3</sub>	最大值
8,177	0.058	0.163	-1.000	0.000	0.061	0.138	1.000

TF-IDF 最後同樣是選擇中位數以及第一四分位數作為區分正向、中立和負向情緒的門檻，得到的結果如表十七所示。整體來說，結果並未與字典法差異太大，唯一值得關注的是，負向情緒的 Recall 從 0.70 上升至 0.83，但 Precision 才犧牲下降了 0.04，這說明 TF-IDF 能夠更準確地辨認負向語句。因此，本研究將挑出，實際為負向但被字典法分類成正向或中立，且在 TF-IDF 被分類為負向語句，作為比較字典法、TF-IDF 以及 BERT 的範例。

表十七、TF-IDF 情緒分析結果

句子數	Accuracy	Precision			Recall			F1-Score		
		正向	中立	負向	正向	中立	負向	正向	中立	負向
1,310	0.60	0.63	0.54	0.62	0.66	0.44	0.83	0.64	0.48	0.71



表十八、三個方法之比較

範例一	全年獲利為 0.57 億元，亦較前一年度衰退 58%。
字典法	將『獲利』視為正向詞彙，而『衰退』詞彙視為負向，一正一負，因此字典法判斷為中立。
TF-IDF	由於『衰退』的權重大於『獲利』，因此 TF-IDF 將其判斷為負向。
BERT	<p>圖中可以發現，BERT 在關注『較』此字時，同時注意『全年獲利』和『衰退』，能辨認主詞是誰。</p>

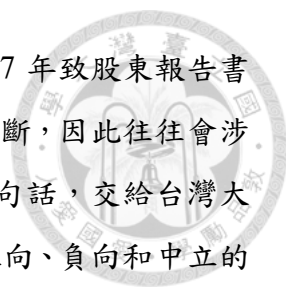
經過方法之間的比較後，得知 TF-IDF 造成負向 Recall 上升的原因可能為致股東報告書負向詞彙數量較少，導致負面詞彙的權重大多數大於正向詞彙的權重。本研究計算正向詞彙權重平均值和中位數分別為 8.30 和 8.62，而負向詞彙的平均值和中位數則分別為 8.52 和 8.91，可以發現普遍負向詞彙的權重大於正向詞彙的權重，因此當一句話正向詞彙與負向詞彙數量相差不多時，TF-IDF 將認為負向詞彙較正向詞彙更為重要，進而將其判斷為負向語句。但相比於 BERT 模型，BERT 能夠理解語意結構，判斷否定詞與詞彙的關係，因此能夠更準確地分辨語句的情緒。

## 第五章 研究結論與限制



由於過往在會計領域的文字探勘方法，常常是依照人工判讀文章辨別文章的情緒，而這樣的蒐集的方式耗時且樣本數量少，因此隨著電腦運算資源的進步，文字探勘方法開始出現演進，像是字典法、樸素貝氏演算法和支撐向量機，但這些文字探勘方法往往都有其致命的限制，例如：字典法只考量相關情緒詞彙出現的頻率，或是難以辨別否定詞是在修飾哪個詞彙；樸素貝氏演算法假設詞彙之間為獨立，並不考慮詞彙的順序性，因此忽略語言的結構性。隨著科技的進步，近年自然語言處理開始嶄露頭角，2019 年 BERT 問世，能夠透過少量會計領域的文本資料結合原本已經訓練好的大數據，訓練出一個專屬會計的大數據模型。

本研究利用 BERT 模型對致股東報告書進行情緒分析，並依照中文致股東報告書的狀況做微調，例如：中文致股東報告書包含有中英夾雜的現象，藉此項更動提高 BERT 模型對致股東報告書情緒分析的準確率。最後，在交叉驗證下，本研究 BERT 模型準確率高達 86%；進一步預測 2011 年至 2018 年的致股東報告書，準確率也同樣高達 90%。在獲得如此良好的表現後，不禁好奇 BERT 模型是否解決以往文字探勘方法的缺點，才能獲得如此表現，因此依 Li (2010b)和 Loughran and McDonald (2016)敘述字典法與樸素貝氏演算法的缺點，本研究透過視覺化和語境測試檢驗 BERT 是否有成功地解決上述所提及的缺點。根據視覺化和語境測試的結果來看，BERT 不但成功捕捉否定詞扮演的角色，且同時能夠捕捉轉折詞的效果；BERT 同樣有考慮語言的結構，當文字順序隨機打亂以後，BERT 模型的預測能力大幅下滑。最後本研究利用 2011 年至 2017 年資料訓練的 BERT 模型，去預測 2011 年至 2018 年的致股東報告書，進而去預測隔年(t+1)的盈餘，作為應用 BERT 模型的範例。雖然最後線性回歸結果顯示，語調與下一年淨利並沒有呈現顯著正相關，但可知其導致的原因並非來自於 BERT 模型預測的不準確，可能是來自於選取前十行過於單薄或是台灣致股東報告書與隔年盈餘並不存在顯著之關聯。未來會計研究能朝著兩個方向更深入鑽研，一是模型訓練的方式，由於財報出現的詞彙較為限縮，若使用詞彙的方式訓練 BERT 是否取得更好的成果，又或是將財務資訊與文字資訊結合，一同訓練 BERT 以觀察這樣的結構是否會取得更優秀的結果。二是將 BERT 應用於不同的會計主題，例如：偵測財報舞弊、查核意見是否受新聞媒體的影響等等，觀察其在不同的主題上表現是否有差異。



本研究在情緒分析上有部分的限制，一是標記 2011 年至 2017 年致股東報告書前十句話的語調情緒，是依賴人工標記，這仰賴會計人的專業判斷，因此往往會涉及主觀的想法，目前是透過隨機抽樣大約 200(佔總句數約 2%)句話，交給台灣大學會計研究所的同學進行複核，且先前並沒有給予同學們判斷正向、負向和中立的標準，避免灌輸本研究主觀的想法，複核結果只有 4 句話(佔抽樣筆數的 2%)有疑慮，例如：「隨著全球產業的快速演進，廠商之競爭已由單純的速度、品質、技術、彈性及製造成本等提升至全球平台的運籌及產業的資源整合，近年來電源管理一直是半導體領域之熱門市場之一，且電源管理 IC 是所有電子產品必備的關鍵元件，隨著環保與節能的議題發酵，節能科技為目前重要發展課題之一，電源管理 IC 扮演不可或缺的關鍵角色。」，本研究將其標記為正向，原因是後半段說明電源管理需求將提升，但其實該句話並沒有說明該公司產品是不是屬於電源管理或是其他有關該公司的資訊，因此若將這句話看成中立，同樣沒有太大的問題。「營收大幅減少的主要原因在於不具競爭力之藍芽晶片停產下市，同時減少該產品之研發人力及費用，營業費用跟隨著減少，致營業損失較前一年度減少 68.1%。」，本研究將其標記為負向，原因是產品的下市導致營收大幅減少，而且由於本研究樣本來自於半導體產業，若營業費用減少來自於研發人力或研究發展費用的削減，將視為中立，因為研究人力及研發費用對該產業格外重要，削減這兩類費用不確定對未來的營運是好或是壞。但由於同學們事先不知道本研究的判斷標準，因此認為此句因為營業損失較前一年度減少，因而標記為正向。

限制二是硬體資源不足，BERT 有分為 12 層及 24 層的不同模型結構，簡單來說，越高的層數代表模型能夠達到越準確特徵表示 (representation)，然而這同時意味著，其所消耗的運算資源越大，由於本研究使用 Google Colab 提供的 GPU 資源，沒有辦法成功嘗試較大的 BERT 結構，因此選擇使用 12 層的結構。

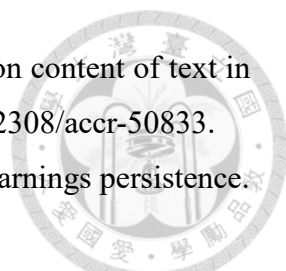
三是礙於時間壓力，本研究僅選擇致股東報告書前十行，該作法可能過於單薄，而導致樣本偏誤的情形。理想的作法應該是隨機選取致股東報告書的其中十行或更多行，使樣本更具有代表性。因此未來研究若能依此方式蒐集樣本，可能得出的結論更具意義。



## 參考文獻



- Antweiler, W., and Murray, F. (2004). Is all that talk just noise? The information content of internet stock message boards. *The Journal of Finance*, 59(3), 1259–1294.
- Biddle, G. C., Hilary, G. & Verdi, R. S. (2009). How does financial reporting quality relate to investment efficiency? *Journal of Accounting and Economics*, 48, 112–131.
- Bochkay, K., & Levine, C. B. (2019). Using MD&A to improve earnings forecasts. *Journal of Accounting, Auditing & Finance*, 34(3), 458-482.  
doi:10.1177/0148558X17722919.
- Bryan, S. H. (1997). Incremental information content of required disclosures contained in management discussion and analysis. *The Accounting Review*, 72(2), 285-301.
- Campbell, J. L., Chen, H., Dhaliwal, D. S., Lu, H., & Steele, L. B. (2014). The information content of mandatory risk factor disclosures in corporate filings. *Review of Accounting Studies*, 19, 396-455. doi:10.1007/s11142-013-9258-3.
- Cole, C. J., and Jones, C. L. (2004). The usefulness of MD&A disclosures in the retail industry. *Journal of Accounting, Auditing & Finance*, 19(4), 361-388.
- Delvin, J., Chang, M. W., Lee, K., and Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformer for language understanding. arXiv:1810.04805.
- Ding, K., Peng, X., and Wang, Y. (2019). A machine learning-based peer selection method with financial ratios. *Accounting Horizons*, 33(3), 75-87.  
doi:10.2308/acch-52454.
- Elwany, E., Moore, D., and Oberoi, G. (2019). BERT goes to law school: Quantifying the competitive advantage of access to large legal corpora in contract understanding. In workshop on document intelligence at NeurIPS 2019.
- Henry, E. (2008). Are investors influenced by how earnings press releases are written? *The Journal of Business Communication*, 45(4), 363-407.  
doi:10.1177/0021943608319388
- Hiew, J. Z. G., Huang, X., Mou, H., Li, D., Wu, Q., and Xu, Y. (2019). Bert-based financial sentiment index and LSTM-based stock return predictability. Submitted to NeurIPS 2019, under review. arXiv:1906.09024.
- Hoberg, G., and Phillips, G. (2015). Text-based network industries and endogenous product differentiation. *Journal of Political Economy* : Forthcoming.

- 
- Huang, A., Zang, A. & Zheng, R. (2014). Evidence on the information content of text in analyst reports. *The Accounting Review*, 89, 2151-2180. doi:10.2308/accr-50833.
- Li, F. (2008). Annual report readability, current earnings, and earnings persistence. *Journal of Accounting and Economics*, 45(2-3), 221-247.
- Li, F. (2010a). The information content of forward-looking statements in corporate filings—A naive Bayesian machine learning approach. *Journal of Accounting Research*, 48, 1049-1102.
- Li, F. (2010b). Textual analysis of corporate disclosures: A survey of the literature. *Journal of Accounting Literature*, 29, 143-165.
- Li, F., Lundholm, R. J., and Minnis, M. (2012). A measure of competition based on 10-K filings. Chicago booth research paper 11-30, *Journal of Accounting Research*.
- Li, M., Li, W., Wang, F., Jia, X., and Rui, G. (2020). Applying BERT to analyze investor sentiment in stock market. *Neural Computing & Applications*. doi:10.1007/s00521-020-05411-7
- Loughran, T., and McDonald, B. (2011). When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks. *Journal of Finance*, 66(1), 35-65.
- Loughran, T., and McDonald, B. (2014). Measuring readability in financial disclosures. *Journal of Finance*, 69(4), 1643-1671.
- Loughran, T., and McDonald, B. (2016). Textual analysis in accounting and finance: A survey. *Journal of Accounting Research*, 54(4), 1187-1230.
- Miller, B. P. (2010). The effects of reporting complexity on small and large investor trading. *The Accounting Review*, 85, 2107- 2143.
- Pan, S. J. and Yang, Q. (2009). A survey on transfer learning. *IEEE Transactions on Knowledge and Data Engineering*, 22(10), 1345-1359. doi: 10.1109/TKDE.2009.191.
- Peters, M. E., Neumann, M., Iyyer, M., Gardner, M., Clark, C., Lee, K., and Zettlemoyer, L. (2018). Deep contextualized word representations. arXiv:1802.05365, 2018.
- Price, S. M., Doran, J. S., Peterson, D. R., and Bliss, B. A. (2012). Earnings conference calls and stock returns: The incremental informativeness of textual tone. *Journal of Banking & Finance*, 36(4), 992-1011.
- Radford, A., Narasimhan, K., Salimans, T., and Sutskever, I. (2018). Improving language understanding by generative pre-training.

Rogers, L. J., Buskirk A. V., and Zechman, S. L. C. (2011).

Disclosure tone and shareholder litigation. *The Accounting Review*, 86(6), 2155-2183.

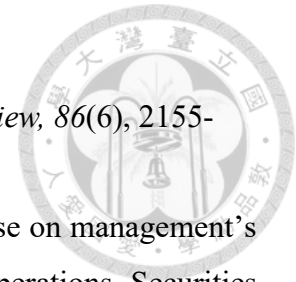
Securities and Exchange Commission (SEC). (1987). Concept release on management's discussion and analysis of financial condition and results of operations. Securities Act Release No. 6711. Washington, D.C.: SEC.

Securities and Exchange Commission (SEC). (2003). Interpretation: commission guidance regarding management's discussion and analysis of financial condition and results of operations. Securities Act Release No. 8350. Washington, D.C.: SEC.

Siano, F., and Wysocki, P. (2018). The primacy of numbers in financial and accounting disclosures: Implications for textual analysis research.

Siano, F., and Wysocki, P. (2020). Transfer learning and textual analysis of accounting disclosures: Applying big data methods to small(er) data sets. doi: 10.2139/ssrn.3560355

Sun, Y. (2010). Do MD&A disclosures help users interpret disproportionate inventory increase? *The Accounting Review* 85(4), 1411-1440.

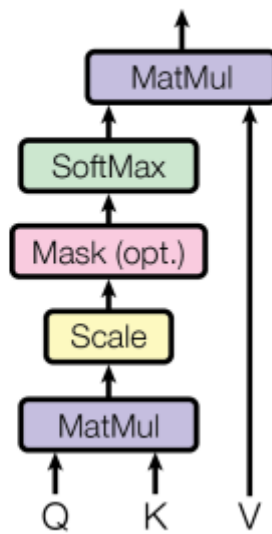


## 附錄

### 附錄一 BERTViz

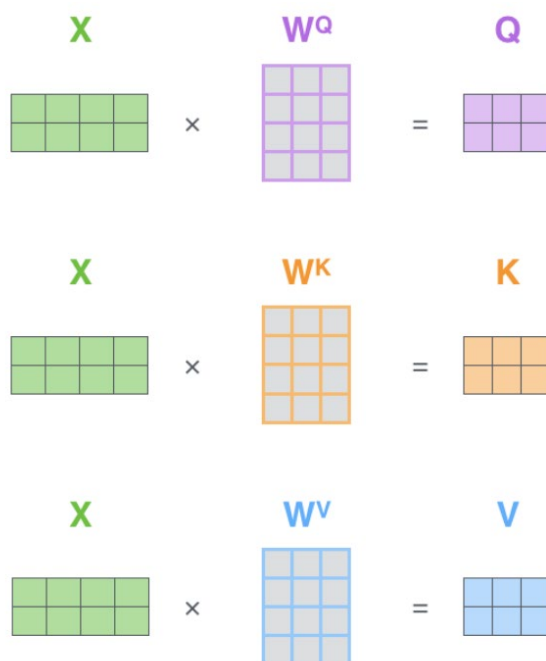
附錄一說明 BERTViz 背後運作的機制，首先要先介紹的是 BERT 模型使用的 Self-Attention，從 Vaswani, Shazeer, Parmar, Uszkoreit, Jones, Gomez, Kaiser, and Polosukhin (2017) 可以看到其結構如下圖。

#### Scaled Dot-Product Attention



其中 Q 代表 Query；K 代表 Key；V 代表 Value，這三個矩陣皆與輸入相對應，而 MatMul 代表矩陣相乘，Mask 代表隨機遮蔽，SoftMax 為歸一化的指數函式。下圖<sup>16</sup>將說明 BERT 是如何計算其 Attention。

<sup>16</sup> 資料來源：<https://jalammar.github.io/illustrated-transformer/>



首先，BERT 會將文字 X 透過不同的權重 W 對應到矩陣 Q、K 和 V，接著計算矩陣 Q 及矩陣 K 的內積，並除以  $\sqrt{d_k}$ ， $d_k$  為 K 矩陣的維度，目的是讓 gradient 更加穩定。接著再利用 SoftMax 將計算結果歸一化，最後乘以各自的矩陣 V 得到加權向量並加總起來。其數學式表示如下：

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{Q \cdot K^T}{\sqrt{d_k}}\right) V$$

下表<sup>17</sup>將上述的步驟彙整

### Attention 計算步驟

輸入	Thinking	Machine
詞向量	$x_1$	$x_2$
Q	$q_1$	$q_2$
K	$k_1$	$k_2$
V	$v_1$	$v_2$
分數	$q_1 \cdot k_1 = 112$	$q_1 \cdot k_2 = 96$
除以 $8(\sqrt{d_k})$	14	12
SoftMax	0.88	0.12
SoftMax*V	$v_1^*$	$v_2^*$

<sup>17</sup> 參考 <https://jalamar.github.io/illustrated-transformer/> 的表示

加總

$$z_1 = v_1^* + v_2^*$$



上述都是一個 Attention head 的情形，而 Vaswani et al.(2017) 論文中有使用多個 Self-Attention 的表示，稱為 Multi-Head Attention，其方式為連接多個 Attention，其數學表示如下：

$$\begin{aligned} \text{MultiHead}(Q, K, V) &= \text{Concat}(\text{head}_1, \text{head}_2, \dots, \text{head}_h)W^O, \text{ where } \text{head}_i \\ &= \text{Attention}(QW_i^Q, KW_i^K, VW_i^V) \end{aligned}$$

Multi-Head Attention 意思為有多個 Self-Attention，因此對應到不同的  $Q_i$ 、 $K_i$  及  $V_i$ , for  $i = 1, 2, \dots, 12$ ，最後產出不同的  $Z_i$ , for  $i = 1, 2, \dots, 12$ ，並進一步將其連接起來乘以權重，得到最後的  $Z^*$  矩陣。而 BertViz 便是將 Multi-head Attention 視覺化出來，線條代表 head 在更新詞向量 (左側) 時，關注其他詞彙的注意力程度 (右側)，因此連接的線條越粗代表在更新該詞向量時，對應的詞彙對該詞向量影響非常大。12 個 Multi-Head 分別對應到不同顏色的線條。