

## 基於 BERT-DAOA 的意見目標情感分析

### Aspect-Based Sentiment Analysis Based on BERT-DAOA

陳震瑜 Chen-Yu Chen

國立中央大學資訊工程學系

Department of Computer Science and Information Engineering

National Central University

[107522103@cc.ncu.edu.tw](mailto:107522103@cc.ncu.edu.tw)

張嘉惠 Chia-Hui Chang

國立中央大學資訊工程學系

Department of Computer Science and Information Engineering

National Central University

[chia@csie.ncu.edu.tw](mailto:chia@csie.ncu.edu.tw)

#### 摘要

社群媒體網站包含了豐富且多樣性的資訊，使輿情分析變成市場調查的方法之一。情感分析是輿情分析的重要一環，目的在取得意見目標網路聲量之外更精準的消費者愛好。然而，即使是句子中包含談論目標，句子的情感類別與文中被評論的目標的情感類別也不盡然一致。因此，本研究應用 Aspect-based sentiment analysis (ABSA)，分析社群網站對意見目標(Opinion Target)的情感類別。在模型的設計上，我們採用 Google 的 BERT[1]作為字詞的嵌入層方法，並參考了 Huan 等人[2]的意見目標情感分類方法，以及 Parikh 等人[3]的自然語言推論的方法，結合兩者來分析意見目標情感分類。實驗結果顯示，在 BERT 之上搭配 Attention-over-Attention (AOA)注意力模型，效能優於 BERT CLS 效能。

#### Abstract

Social media networks provide rich and diverse information, making opinion analysis and network volume analysis a new method to investigate and understand the market. Sentiment analysis aims to determine the emotional category in a given text. Since there might be several targets being commented on in the text, Aspect-base sentiment analysis (ABSA) has been proposed to explore the sentiment categories of a target in different aspects. In this paper, we explore the idea of ABSA for sentiment analysis of singers on social networks. We utilize

BERT as the embedding layer method for characters and words to explore the relationship between a given sentence and a mentioned target. We consider two attention mechanisms to enhance the performance. The experimental results show that adding the attention layer on top of BERT outperforms the basic BERT-CLS model.

關鍵詞：深度學習，情感分析，意見目標情感分析

Keywords: Deep Learning, Sentiment Analysis, Aspect-based Sentiment Analysis.

## 一、緒論

現今許多網路論壇、串流影音平台等熱門的網路服務，都設置有留言區或是收藏功能等的反饋機制，讓這些網路平台中累積了多樣且豐富的資料，使得網路輿情分析 (Opinion Analysis) 成為了調查與理解市場的方法之一。在商業市場上，透過輿情分析能夠消費者對於企業的服務、產品等的正負面評價，同時也能發掘消費者所在意的優缺點，讓產業能藉由這些資訊進而調整行銷策略、改進產品生產方向與服務品質。

情感分析是輿情分析的重要一環。傳統的情感分析是針對文章(document-level)或整句話(sentence-level) 進行情感分類，當文本中包含多個討論主題時，如此所得的情感結果不見得是我們所關注角色的情感，而是對其他意見目標的情感結論。為了更精準地評估個別意見目標的情感，近年的情緒分析研究即有意見目標情感分析 (Aspect-level sentiment analysis, ABSA)，對於給定的句子不只要預測句中目標 (Target) 的情感類別，還須預測該句屬於何種面向 (Aspect)。舉例而言，對於同一家旅館或餐廳的評論中，使用者可以針對其設備、服務、清潔、地點、菜色等不同層面提出個人觀點，以及對於不同的面向或目標給予情緒上的評論。

本研究的主要目標是針對社群網路上與歌手相關的評論文章進行情感分析，不論評論內容對於評論者而言情緒為何，評論內容對於主要被評論歌手的情感評價才是我們關注的重點。因此，我們以歌手名作為評論目標 (Opinion Target)，透過 ABSA 中的子任務——意見目標的情緒分類 (Aspect-level Sentiment Classification)，運用此方法來更準確的判斷文本中的評論目標 (Opinion Target) 的情感類別，並可做為歌曲點播率及市場反應度預測，進而做為歌手行銷決策的考量。

本研究在模型的設計上，運用透過 2018 年 Google AI 團隊所提出的預訓練 BERT

中文模型[1]來進行參數微調，在 BERT 的使用上與以往的任務不同，本研究不直接採用 BERT 的分類向量 CLS 直接進行情感分類，而是將 BERT 作為來對輸入的語句進行字詞向量的轉換，並在 BERT 模型之上，運用 Huan 等人[2]的意見目標情感分類方法，以及 Parikh 等人[3]的自然語言推論 (Natural Language Inference) 的方法，進一步探討句子與目標之間的相關性，並對意見目標進行情感分類。

本篇論文貢獻主要包含兩個方向：(1)應用意見目標情感分析於社群網路評論，預測主要評論對象之情感分類結果，提供廠商掌握市場動向。(2)運用 BERT 作為嵌入層，並在 BERT 後運用多種不同的架構模型，來解決意見目標情感分類，提供更多樣化的 BERT 使用方法。

## 二、 相關研究

意見目標情感分析(Asspect Based Sentiment Analysis, ABSA)屬於情感分析中的一項任務，意見目標情感分析可以看成多項子任務問題，意見目標類別分類(Asspect Category Extraction)、意見目標擷取(Asspect Target Extraction)與意見目標情感分類(Asspect-level Sentiment Classification)等，而 Schouten 等人[4]與 Zhou 等人[5]也針對上述的 ABSA 相關知識與任務進行統整與介紹。先前的研究大都是單獨解決部分的子問題，但近年來有不少研究將多項子問題結合起來，即先做目標擷取再做目標的情感分類，為了讓模型可以借鏡個別標記所帶來的資訊，透過 multi-task 架構同時處理目標擷取與情感分類的聯合訓練(Joint Learning)方法變成近年來主要的研究趨勢[6][7]。

本研究專注於意見目標的情感分類(Asspect-level Sentiment Classification, ASC)。以往的情感分類(Sentiment Classification)問題主要是判斷整體文章(Document-level)或整句話(Sentence-level)的情感，而 ASC 的任務中則需要考慮到文本中對於目標(Target)的情感類別。在研究上，目標可以是指一個實體(Entity)或實體的某個層面(Aspect)，為了簡單起見，這裡通稱目標 (Target)；在 ASC 的任務中需要給定一個文本與出現在文本中的一個目標，用來判斷句子對目標的情感類別。

過去在 ASC 任務中，已有不少深度學習方法的應用，如：卷積神經網(CNN)與門控機制(Gating Mechanism)[8]、注意力機制 (Attention mechanism)[2]、Transformation Networks [9]等方法。這些方法都是使用 Word Embedding 的方式來做為輸入句子與目標的詞向量轉換，近年來 BERT 也逐漸被運用於 ABSA 的任務。例如 Sun 等人[10]將句子

與目標以 Sentence Pair 的形式作為 BERT 的輸入，透過製造輔助句子(Auxiliary Sentence)，來重新構建輸入的 Sentence Pair 的內容，並使用 BERT 的用於分類任務的向量 CLS 進行分類，獲得比原本 BERT 更好的效能。Xu 等人[11]將 BERT 應用於機器閱讀理解和 ABSA 任務中，同樣是透過目標與句子所組成的 Sentence pair 做為輸入，並透過作者提出的 Post-training 方法，運用與任務和數據領域相關的大型語料庫，來增強預訓練好的 BERT，最後使用 Post-training 後的 BERT 分類任務向量 CLS 進行目標的情感類別預測，並在實驗結果上獲得比原本 BERT 更加好的效能。

上述使用 BERT 的方法主要是以 Sentence Pair 的形式作為模型的輸入，並用額外數據資料、改善模型的輸入，來改進模型的預測效果，顯示 BERT 分類任務的 CLS 向量難以超越之處。本研究則希望能超越 BERT 分類任務的 CLS 向量的效能，應用 BERT 作為字詞向量的轉換，對於輸入句子與目標所組成的 Sentence pair 的輸出結果，搭配不同的 ASC 相關的模型，進一步探討句子與目標之間的相關性。

### 三、 BERT 模型嵌入層方法

本研究使用過大量中文資料預訓練過的 BERT 模型，參數設定 transformer blocks 層數有 12 層、嵌入大小為 768、transformer blocks 的 Self-Attention Head 數量為 12，以及使用參數微調 (Fine-Tuning) 的方法，來使用預訓練好的 BERT 模型將字詞轉換成嵌入向量，在中文模型的嵌入單位是以「字」為單位。並 BERT 之上運用 Huang 等人[2]所提出的 Attention-over-Attention (AOA)的注意力模型與 Parikh 等人[3]在自然語言推論任務提出的 Decomposable Attention 方法（簡稱 DA）的注意力模型，以及比較 AOA 與 DA 模型架構，所提出結合兩者架構的 Decomposable- Attention-over-Attention (DAOA)模型。如圖 1 所示，在模型的設計上總共分成 4 層：

#### 1. 輸入層 (Input Layer) :

以句子(Sentence)與句子中的意見目標(Opinion Target)做為輸入，並將兩者進行以字元(character-based)進行分割，表示符號為分別為  $S = [s_1, \dots, s_n]$ ,  $T = [t_1, \dots, t_m]$ ，代表句子中有  $n$  個字元，而目標則  $m$  字元，且目標詞是句子中的某

一小段連續字元所組成。並將句子與目標處理成 BERT 的 Sentence Pair 輸入格式，符號表示為  $BERT_{input} = [[CLS], s_1, \dots, s_n, [SEP], t_1, \dots, t_m, [SEP]]$ ，輸出的結果為形成一個長度為  $n+m+3$  的序列。

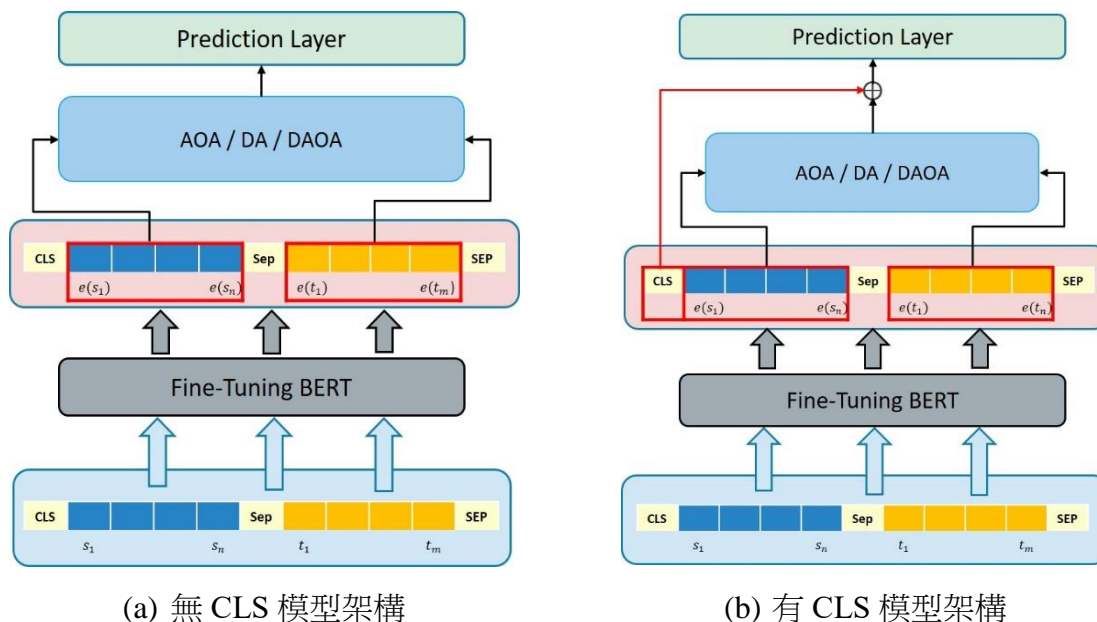


圖 1、BERT 模型架構

## 2. BERT 層(BERT Layer):

本研究使用預訓練好的 BERT 模型將  $BERT_{input}$  轉換成嵌入向量，而嵌入大小為 768。而在 BERT 模型的輸出有分為兩種：pooler output 和 sequence output。前者的輸出為整個 sentence pair 序列的 representation 的向量  $[CLS] \in R^{768}$ ，是專門用於分類任務上的特殊向量；後者則是輸出句中每個字元經過 BERT 計算後的都會獲得一個對應的向量。

本研究使用 sequence output，輸出結果為  $BERT_{output} \in R^{(n+m+3)*768}$ ，並從矩陣中取出句子與目標經由 BERT 轉換後的字向量。取出結果表示  $\bar{S} = [\bar{s}_1, \dots, \bar{s}_n] \in R^{n*768}$ ， $\bar{T} = [\bar{t}_1, \dots, \bar{t}_{i+m-1}] \in R^{m*768}$ 。

## 3. 特徵擷取層:

根據 BERT 層的輸出做為輸入。運用 AOA、DA 以及比較 AOA 與 DA 模型架構，並結合兩者架構提出的 DAOA 注意力模型(詳細模型如 3.1 節所述)，來計算句

子(Sentence)與句子中的意見目標(Opinion Target)的嵌入向量之間的關係。

#### 4. 預測層(Prediction Layer) :

預測層我們設計了兩種方法，第一種是如圖 1 (a) 的架構，運用特徵擷取層的輸出向量  $M_{output} \in R^{d_M}$ ，透過前饋式神經網路分類預測， $d_M$ 表示輸出向量的維度。第二種則是如圖 1 (b)的架構，將 BERT 的 Pooler Output 的分類任務的向量 CLS，符號表示為  $BERT_{CLS} \in R^{768}$ ，與特徵擷取層的輸出結果結合 (公式 1)，再透過前饋式神經網路進行分類預測。

$$C = [BERT_{CLS}, M_{output}] \quad \text{公式 1}$$

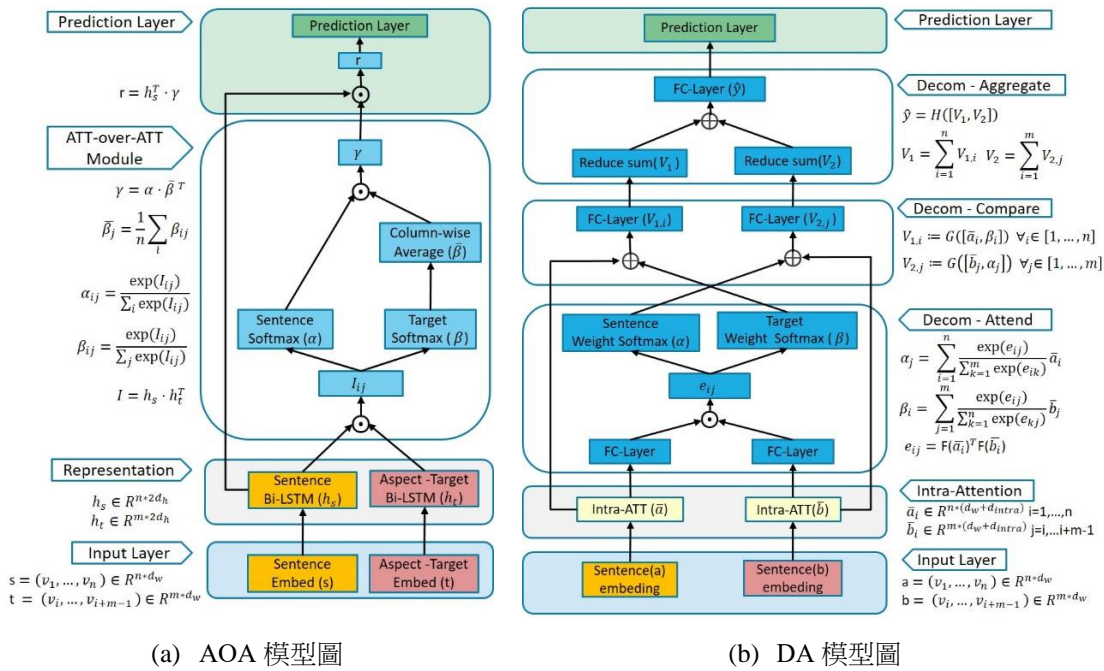


圖 2、AOA 與 DA 模型比較圖

### 3.1 DA 與 AOA 架構比較

再介紹 DAOA 的模型架構前，我們畫出 AOA 與 DA 兩篇論文模型從輸入到輸出的實作圖(圖 2)，並詳細比較兩者的注意力方法。根據圖 2 (a)的 AOA 的注意力計算部分(ATT-over-ATT)是一種非對稱的注意力模型方法，透過計算兩句話之間的注意力權重，並將權重加權於其中一句話。根據圖 2 (b)，DA 的注意力計算分成三個部分

Attend、Compare、Aggregate，且是屬於對稱式的計算，透過前饋神經網路 F 計算兩句話之間的注意力權重，運用 columns-wise 與 row-wise 的 Softmax 對兩句原始句子進行加權與正規化，並在將加權後的句子，與另一方的原始句子進行訊息的交互，運用前饋神經網路 G 進行比較，來探討兩句話詞與詞的相關性，最後在將比較結果透過前饋神經網路 H 進行整合。在注意力計算部分 DA 較 AOA 來的更複雜。

### 3.2 DAOA 架構

本研究的注意力層機制的設計想法上，運用根據圖 2 (b)中 DA 對於句子與意見目標句子之間的注意力權重計算方法(Decom-Attend)與交互比較注意力權重的方法(Decom- Compare)，然而 DA 最後的計算 Decom- Aggregate 只是單純的將比較結果進行總和。因此，我們想透過 AOA 對於資訊彙整的方式來將比較結果進行整合，來加強其中一方的權重，並進行情感預測，其架構如圖 3 所示。

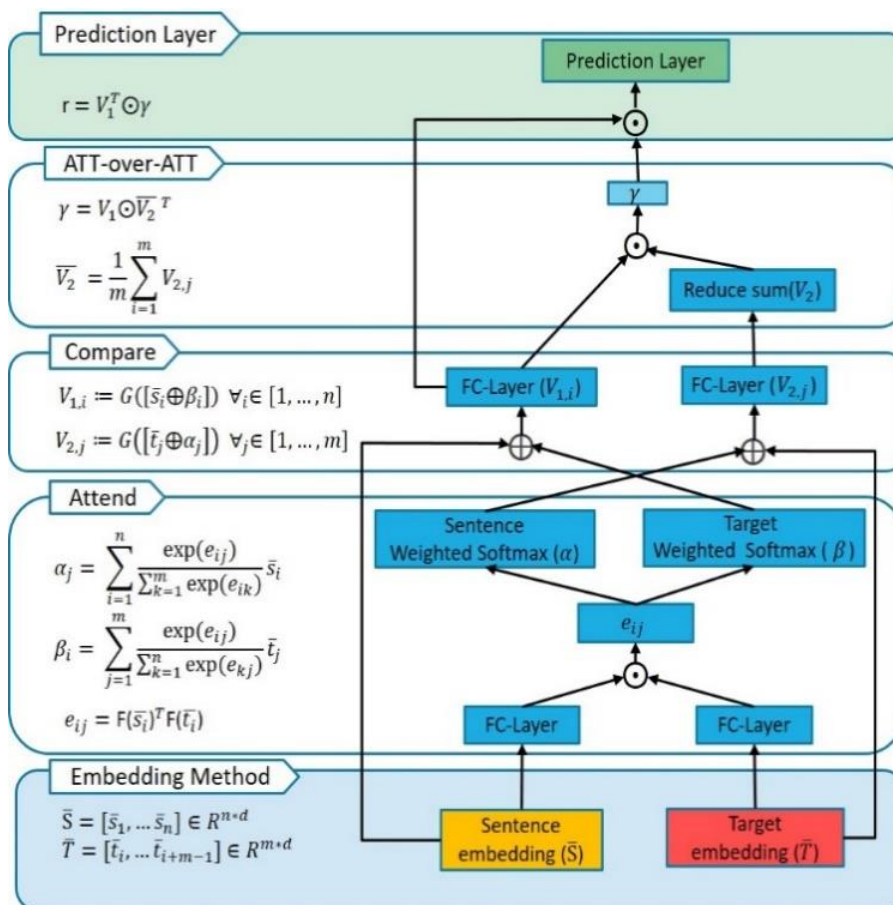


圖 3、Decom-AOA 架構



A. 注意力層(Attend Layer) :

我們首先透過激勵函數為 ReLU 的前饋式神經網路 F 來，來計算兩句話中每個詞的注意力權重  $e_{ij}$ (公式 2)， $\beta_i$  表示 b 句中每個詞對 a 句中第 i 個詞的注意力權重經過加權與正規(normalization) 的總和結果(公式 3)， $\alpha_j$  可用相同方式獲得(公式 4)。

$$e_{ij} = F(\bar{s}_i)^T F(\bar{t}_j) \quad \text{公式 2}$$

$$\beta_i = \sum_{j=1}^m \frac{\exp(e_{ij})}{\sum_{k=1}^n \exp(e_{ik})} \bar{t}_j \quad \text{公式 3}$$

$$\alpha_j = \sum_{i=1}^n \frac{\exp(e_{ij})}{\sum_{k=1}^m \exp(e_{ik})} \bar{s}_i \quad \text{公式 4}$$

B. 比較層 (Compare Layer) :

經過 Attend 加權後的兩個句子則分別其原始句子透過前饋神經網路 G 來進行特徵擷取(公式 5)，這部分的用意是希望能夠透過此種方式交換相互加權後的訊息，進一步的比較兩句中詞與詞的相關性。

$$\begin{aligned} V_{1,i} &:= G([\bar{s}_i \oplus \beta_i]) \quad \forall_i \in [1, \dots, n] \\ V_{2,j} &:= G([\bar{t}_j \oplus \alpha_j]) \quad \forall_j \in [1, \dots, m] \end{aligned} \quad \text{公式 5}$$

C. Att-over-ATT 層 (Att-over-ATT Layer) :

接著透過 Attention-over-Attention (AOA)的方法，將前一層所得的目標權重平均(公式 6)，做為新的權重，並與前一層句子進行比較，透過內積加強句子與目標每個詞之間的相關性(公式 7)。

$$\bar{V}_2 = \frac{1}{m} \sum_{j=1}^m V_{2,j} \quad \text{公式 6}$$

$$\gamma = V_1 \odot \bar{V}_2^T \quad \text{公式 7}$$



#### D. 輸出層(Output Layer) :

最後 DAOA 輸出層的部份，則是經由 Att-over-ATT 所獲得的比較權重，與比較層所獲得的  $V_1$ ，進行矩陣相乘 (公式 8)。其用意是希望，原始的句子已經與目標的注意力權重進行比較，並在運用 Att-over-ATT 所獲得的比較權重，進一步的加強句子與目標每個詞之間的相關性。

$$r = V_1^T \odot \gamma \quad \text{公式 8}$$

#### 四、 資料集準備與分析

本研究所使用之資料集，是來自台灣的「批踢踢 PTT 實業坊」，由於我們關注的主題是歌手的網路聲量，因此我們僅抓取與音樂相關討論版的文章。資料準備方式如下，首先對 PTT 的貼文進行預處理，解決 PTT 貼文特定格式、特殊符號、URL 等，排除會影響文本分析的訊息，並使用歌手字典做為種子，搜尋含有歌手的段落，並透過人工的方式判斷該段落主要被討論的實體，以及該段落所描述的內容對於被標記實體的情感。

為了確保測試資料的標記品質，測試資料經由三組人員以進行標記，最後採用多數決的方式來決定情感類別。如圖 4 範例所示，紅色代表該段落主要被討論的對象，而標記者會根據整段文字的敘述，來判斷該段落對於討論對象情感類別，而情感類別包含「正向、負向、中立」三類的情感。

Target	Target Sentiment	Corpus
張藝興	中立	張藝興回母校設立獎學金未來10年贊助100萬人民幣今日(5月13日)，有網友透過微博稱男團EXO成員張藝興(LAY)向母校湖南師大附中捐款100萬元人民幣和一架鋼琴。據悉，張藝興為湖南師大附中設立藝術獎學金，每年捐款10萬元人民幣幫助有夢想的學弟...
五月天	正評	謝謝你、你、你、你、你。謝謝你們願意成為五月天，謝謝你們願意當五月天。謝謝你們讓我當五月天的家人。想和你們說的感謝不知道還能如何表達。我在五月天音樂中療癒，傷口變成勇氣，把眼淚擦乾，又能勇敢邁進...
徐懷鈺	負評	吳宗憲安慰徐懷鈺訊息曝光！要她問自己錄影時快不快樂2016年04月10日1750記者黃子瑋/台北報導「國民天后」徐懷鈺日前上「綜藝玩很大」被批評為求勝負使出奧步，耍賴硬盧，試圖讓吳宗憲心軟...

圖 4、標記資料範例

表 1、Aspect-Based 資料集統計

Dataset	# of Sent.	Class			# Distinct Targets
		Negative	Positive	Neutral	
Training	1,238	30	426	782	247
Testing	705	17	313	375	189

我們透過 Kappa 值來檢視 3 組人員的情感標記與透過多數決所得情感類別的一致性。各組 Kappa 值分別為 0.66、0.66、0.70，都達到 0.6 以上，根據 Kappa 值的標準，測試資料的標記一致性尚可接受。

## 五、 實驗與模型效能評估

### (一) Bert 模型效能

本階段實驗採用 BERT 的 pooler output 所輸出的分類任務的向量 CLS 之預測結果做為 Baseline，在此簡稱 BERT CLS。首先比較運用 BERT 嵌入層方法，在其架構上搭配不同的注意力模型 AOA、DA 以及 DAOA，來計算轉換後句子與目標之間相關性模型的效能。第二部份，則是在注意力模組輸出後，加入 CLS 向量來探討是否能夠增加效能。各項模型如表 2 所示，表中的每個模型的效能，都是模型執行 10 次後平均的結果。

表 2、BERT 模型之效能

Model	F1-Score				Accuracy (%)
	Negative	Positive	Neutral	W. Avg	
BERT CLS	0.428	0.705	0.758	0.726	72.81
BERT-AOA	0.334	0.715	0.759	0.728	73.16
BERT-DA	0.289	0.689	0.746	0.710	71.49
BERT-DAOA	0.107	0.703	0.758	0.719	72.40
BERT-CLS-AOA	0.321	0.711	0.756	0.726	72.65
BERT-CLS-DA	0.250	0.698	0.753	0.718	72.10
BERT-CLS-DAOA	0.208	0.700	0.760	0.721	72.70

首先，我們先比較表 2，前四項的效能，在模型整體效能的表現上 BERT-AOA 的模型表現最好(Weighted avg. F1 = 72.8)，高於 BaseLine 效能(72.8 v.s. 72.6) 以及其餘的模型效能，且 BERT-AOA 在正向與中立類別的效能分數也是所有模型中表現出色的。不過在負向情感類別的效能分數上 BERT 展現最好的效能 (Negative F1 = 0.428)。

接著，我們加入了 CLS 的向量，來觀察是否能夠改善效能，加入 CLS 向量後對於 BERT-DA (71.0 -> 71.8) 與 DAOA (71.9 -> 72.1) 都可獲得了效能上的改善，但對於的 BERT-AOA 模型在效能的表現上是下降的。

透過實驗，我們發現在 BERT 的 sequence output 上運用太複雜的注意力模型架構可能會導致效能的下降。舉例而言，AOA 的運算比 DA 來的相對簡單許多，而本研究提出的架構 DAOA，雖然在效能上比 DA 來得好(71.9 v.s. 71.0)，但在計算的複雜性上，還是比 AOA 來的更加複雜，因此在 BERT 上應用較為簡單的運算架構，會比複雜的運算架構效能表現上來得更好。而加入 CLS 向量，從實驗結果顯示是有幫助於改善原先效能表現較差的模型。

由於表 2 表現較好的前三個模型在整體效能差異較小，本研究透過盒鬚圖 (box plot) 進一步分析模型執行 10 次的過程中的模型效能差異。圖 5(a)是三個情感類別的 F1-Score 效能，在中立與正向的部分三個模型的最大值以及效能分數的浮動差異不大，但在負向類別，能看出 BERT 的效能明顯高於另外兩者，且效能分數的浮動較其餘模型穩定。圖 5(b)為整體模型的 Weighted avg F1 與 Accuracy，雖然三個模型的 Weighted avg F1 最大值近乎相同，但 BERT-AOA 是三者模型中平均 Weighted avg F1 分數最高且分數浮動最穩定的模型。而在 Accuracy 的部份 BERT-AOA 在模型不管在最大值還是平均分數，都明顯好於其它模型。根據盒鬚圖的分析，得知 BERT-AOA 模型是所有模型中效能最好且最為穩定的模型。

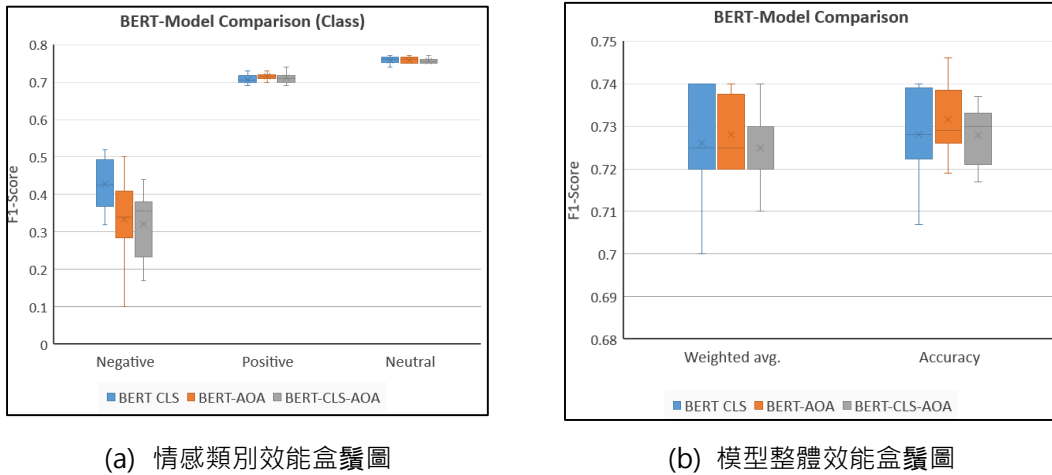


圖 5、BERT 模型盒鬚圖

## (二) 預測錯誤分析

本階段根據 Weighted avg F1 前三名的模型 BERT-AOA, BERT CLS 以及 BERT-CLS-AOA 的共同預測錯誤的結果進行錯誤分析，並列出模型在預測上常見的錯誤範例。在第一、二個範例中，"疑似劈腿"、"孝順"並不是情感詞彙，但根據人的認知判斷，這些詞彙對於目標來說是帶有正負評價的詞彙。第三個範例中，雖然有明顯的情感詞彙，但這些詞彙並不是指向於目標的詞彙，而是描述留言者本人。第四個範例中，出現了強烈的負面詞彙，但是這些詞彙並不指向於目標，而是一種情境上的敘述方式。

Target	Target Sentiment	Corpus	BERT	BERT-AOA	BERT-CLS-AOA
唐從聖	負評	從哥大方認了一認了未婚生女！ <b>唐從聖</b> PO全家福照「今天自由時報即時新聞」藝人 <b>唐從聖</b> 先是被爆未婚生女，又被媒體揭露正與一名護士交往，疑似劈腿，但 <b>唐從聖</b> 昨晚在臉書上貼出一家四口的照片，等於認了未婚生女	中立	中立	中立
羅志祥	正評	<b>羅志祥</b> 講小時候的故事，感覺很辛苦他會這麼 <b>孝順</b> 是知道要回頭！媽媽突然出現，小豬沒猜到很驚訝！...	中立	中立	中立
蘇打綠	中立	在這誠心誠意地徵求 <b>蘇打綠</b> 謝天謝地演唱會門票2張非常喜歡春日光專輯，最喜歡日光，交響夢，早點回家三首歌，如果能夠如願前往聆聽，一定會遵守 <b>蘇打綠</b> 演唱會的所有規定，	正評	正評	正評
吳宗憲	中立	媒體來源蘋果 <b>吳宗憲</b> 為節目哭窮被酸放屁諷馮光遠值8千元再談金鐘50唇槍舌箭「蔡維歡、顏馨宜／台北報導」 <b>吳宗憲</b> （憲哥）在金鐘獎典禮上怒嗆評審，引起網友共鳴，並創下當天典禮最高收視417，吸92萬人收看。但他引言...	負評	負評	負評

圖 6、錯誤分析範例

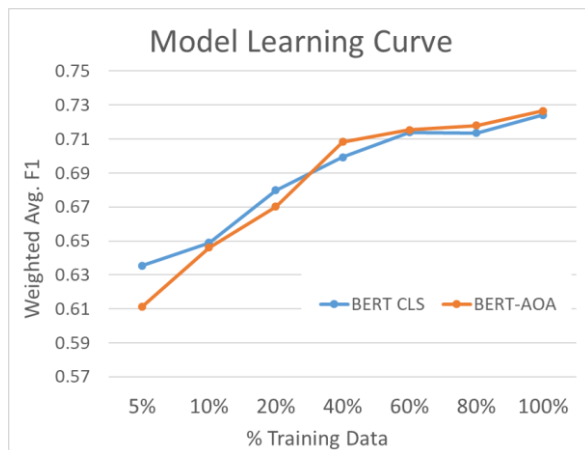


圖 7、模型學習曲線

從上述錯誤分析來看，現有模型對於具有情感涵義的詞彙，以及對於情感詞彙與意見目標的關係，並未能完全理解，難以判斷。

### (三) 模型學習曲線

根據 BERT 模型中 Weighted Avg F1 表現最好的 BERT-AOA 模型與 Baseline 模型 BERT CLS，來畫出模型學習曲線(Learning curve)，並探討模型與資料之間的關係。根據圖 7 中可以觀察到，兩個模型在訓練資料量只有 5%的情況下模型的 F1-score 就能夠到達 60 以上，而 BERT CLS 在數據量少(< 40%)的時候效能高於 BERT-AOA 模型，但在模型隨著資料量的增加 BERT-AOA 效能比 BERT CLS 來的更好，且隨著資料量的增加，BERT-AOA 模型的 F1-score 都可以隨之提升，若是增加更多訓練資料，也許能獲得更好的效能。

### (四) 新增資料效能影響

由於本研究的負向情感訓練資料偏少，使模型在負向類別學習效能表現上不佳，因此我們希望透過增加負向標記訓練資料，是否能否提升模型對於負向類別的效能，以及資料量的增加數是否會再次提升效能。

我們以模型中表現最好的 BERT-AOA 來進行這項實驗。如圖 8 所示，BERT-AOA 隨著負向資料的增加，負向的 F1-score 獲得了提升。不過在負向 F1-score 提

升過程中，並沒有改善其餘類別的 F1-score 與整體模型的 Accuracy。當負向訓練資料增加超過 100 筆時，整體效能略為下滑，顯示負向資料僅有助於改善模型對於負向類別的預測效能。

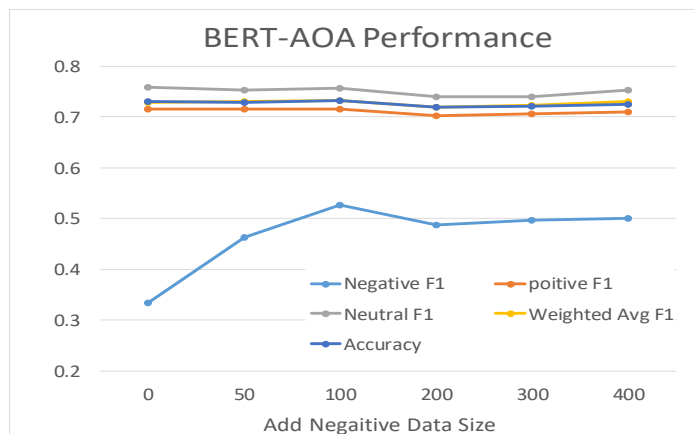


圖 8、BERT-AOA 效能曲線圖

## 六、 結論

本研究中透過 BERT 模型作為嵌入層來對句子與目標的字進行轉換，並搭配不同 NLP 任務的注意力模型來找出句子與目標之間的相關性。實驗結果顯示，BERT 模型之上搭配注意力模型以及加入 BERT 的 CLS 分類向量是有助於 BERT 提升模型的效能，但在 BERT 輸出後搭配過於複雜的注意力模型可能導致模型效能的下降。在本研究的模型中由於訓練資料負向資料過少，本研究嘗增加負向的訓練資料，根據實驗結果，增加負向資料有助於改善模型負向資料的效能，且對於其餘類別以及整體效能的提升有限。

## 參考文獻

- [1] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, “BERT: Pre-training of deep bidirectional transformers for language understanding”, in Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1, Minneapolis, Minnesota: Association for Computational Linguistics, Jun. 2019, pp. 4171–4186.
- [2] B. Huang, Y. Ou, and K. M. Carley, “Aspect level sentiment classification with attention-over-attention neural networks”, vol. abs/1804.06536, 2018.
- [3] A. Parikh, O. Täckström, D. Das, and J. Uszkoreit, “A decomposable attention model

- for natural language inference”, in Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, Nov. 2016, pp. 2249–2255.
- [4] K. Schouten and F. Frasincar, “Survey on aspect-level sentiment analysis”, IEEE Transactions on Knowledge and Data Engineering, vol. 28, no. 3, pp. 813–830, 2016.
- [5] J. Zhou, J. X. Huang, Q. Chen, Q. V. Hu, T. Wang, and L. He, “Deep learning for aspect-level sentiment classification: Survey, vision, and challenges”, IEEE Access, vol. 7, pp. 78 454–78 483, 2019.
- [6] D. Yin, X. Liu, and X. Wan, “Interactive multi-grained joint model for targeted sentiment analysis”, in CIKM ’19, 2019.
- [7] R. He, W. S. Lee, H. T. Ng, and D. Dahlmeier, “An interactive multi-task learning network for end-to-end aspect-based sentiment analysis”, in Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, Florence, Italy: Association for Computational Linguistics, Jul. 2019, pp. 504–515.
- [8] W. Xue and T. Li, “Aspect based sentiment analysis with gated convolutional networks”, in Proceedings of the 56th Annual Meeting of the association for Computational Linguistics (Volume 1: Long Papers), Melbourne, Australia, Jul. 2018, pp. 2514–2523.
- [9] X. Li, L. Bing, W. Lam, and B. Shi, “Transformation networks for target-oriented sentiment classification”, in Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), Melbourne, Australia: Association for Computational Linguistics, Jul. 2018, pp. 946–956.
- [10] C. Sun, L. Huang, and X. Qiu, “Utilizing BERT for aspect-based sentiment analysis via constructing auxiliary sentence”, in Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1, Minneapolis, Minnesota: Association for Computational Linguistics, Jun. 2019, pp. 380–385.
- [11] H. Xu, B. Liu, L. Shu, and P. Yu, “BERT post-training for review reading comprehension and aspect-based sentiment analysis”, in Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1, Minneapolis, Minnesota: Association for Computational Linguistics, Jun. 2019, pp. 2324–2335.