

KOLab at ROCLING-2025 Shared Task: Research on Emotional Dimensions in Chinese Medical Self-Reflection Texts

Chia-Yu Chan, Chia-Wen Wang, Jui-Feng Yeh

Department of Computer Science and Information Engineering, National Chiayi
University

{s1112929,s1112946,ralph}@mail.ncyu.edu.tw

摘要

目前大多數的情感分析技術主要應用於一般性文本，如社交媒體或新聞報導，對於醫學領域中的情感識別仍屬相對空白。自我反思包含個人與內在自我的交流，對於人們未來的生活有正向影響。本篇旨在針對醫學領域相關人員的反思文本，設計出回歸模型，以填補醫學領域在情感分析的空缺。本次任務採用 BERT 模型，配合 Chinese EmoBank 的資料集進行訓練，以 ROCLING 2025 Dimensional Sentiment Analysis - Shared Task 所提供的測試集進行評估，評估結果顯示 Valence 和 Arousal 的 PCC 分別是 0.76、0.58；而 MAE 的分數分別為 0.53、0.82。

Abstract

Currently, most sentiment analysis techniques are primarily applied to general texts such as social media or news reports, and there is still a relative gap in emotion recognition within the medical field. Self-reflection involves communication between individuals and their inner selves, which has a positive impact on people's future lives. This article aims to design a classification model for reflective texts aimed at medical professionals to fill gaps in sentiment analysis within the medical field. This task used a BERT model, trained on a dataset from the Chinese EmoBank, and evaluated using the test set provided by the ROCLING 2025 Dimensional Sentiment Analysis - Shared Task. The assessment results show that Valence and Arousal's PCC scores are 0.76 and 0.58 respectively, while the MAE scores are 0.53 and 0.82, respectively.

關鍵字：BERT、皮爾森相關係數、平均絕對誤差

Keywords : BERT, PCC, MAE

1 Introduction

近年來，隨著深度學習技術日漸發達，本研究使用雙向編碼器 BERT (Bidirectional Encoder Representations from Transformers) 實現一個給醫學領域人員記錄並反思自己的文本。由於 BERT 可以針對許多自然語言處理 NLP (Natural Language Processing) 任務進行微調且應用範圍廣泛，微調 (fine-tuning) 對模型效能有明顯影響，特定領域語料的再訓練 (domain-specific retraining) 也更能提升準確性[10]。本研究結合了自然語言處理及分類式情緒分析以紀錄醫學領域人員的情感分數。使用維度式的情緒分析比傳統情緒分類法 (如正面/負面) 可以提供更細緻的分析[12]。

有別於英文，中文文本在語意上經常面臨許多問題，如分詞、語意等，因此本研究使用 BertTokenizer 進行分詞，採用 WordPiece 方法，適合用於中文 BERT 模型，將文本切分成子詞 (subword)，再透過特徵向量和詞嵌入的方式回歸 Valence 和 Arousal 的分數。

在文本情緒分析的領域中，Calvo 與 Kim (2013) 比較了兩種主要的情緒分析模式，以情緒標籤為基礎的「分類式」和以連續情緒維度 (Valence、Arousal、Dominance) 為基礎的「維度式」。研究顯示，維度模型能在心理學展現優勢。為後續將心理理論整合到自然語言處理的情緒辨識提供了重要基礎[2]。

2 METHODS

2.1 Model Architecture

本研究設計了一個模型來完成本次任務，使用 BERT 中文預訓練模型 (bert-base-chinese) 進

行，並實現分詞。本模型的任務為提取文本中的每個詞的 Valence 和 Arousal 的特徵向量，去進行 Valence 和 Arousal 的分數計算。

2.1.1 BERT

BERT 模型，主要為 NLP 模型中的編碼器。框架包括兩個步驟：預訓練和微調。在預訓練期間，模型在不同的預訓練任務上對未標記的資料進行訓練。而在微調方面，BERT 模型首先用預訓練參數進行初始化，然後使用來自下游任務的標記資料微調所有參數。雖然每個下游任務都有單獨的微調模型，但它們都是用相同的預訓練參數初始化的[4]。

在預訓練中採用 masked language model (MLM) 是 BERT 重要的預訓練任務，用於建立雙向上下文推理能力[3]。其學習目標是預測文本中被隨機遮蔽的詞，利用上下文兩側的信息進行推斷；Next Sentence Prediction (NSP)，為了提升 BERT 模型在捕捉長程依存關係 (long-term dependencies) 方面的能力，訓練過程中引入了下一句預測任務 (Next Sentence Prediction, NSP)，在此任務中，模型需判斷「序列 B 是否為序列 A 的後續內容」。若屬於，序列 A 與序列 B 會從同一文件中依照自然順序抽取；若不屬於，則序列 A 與序列 B 會隨機取樣。透過此設計，模型得以學習文本片段之間的語義連貫性與上下文關聯性，進而加強對篇章結構的理解能力[6]。BERT 的雙向 Transformer 架構能捕捉上下文語義，在 NLP 任務中表現突出[9]。

2.1.2 BertTokenizer

分詞 (Tokenization) 是指將輸入文本劃分為子單位，稱為詞元 (tokens)。這些詞元之後會被用於自然語言處理的後續步驟，例如形態分析 (morphological analysis)、詞性標註 (word-class tagging) 以及句法分析 (parsing) [5]。

Tokenizer 是 NLP 的核心元件之一。由於模型只能處理數字，因此需要標記器 Tokenizer 將輸入的文本轉換為模型可處理的數據。

2.1.3 WordPiece

WordPiece 是 Google 為預訓練 BERT 而開發的標記化算法[11]，其在訓練方面與 BERT 相似，但實際標記化方式不同。WordPiece 是從一個小詞彙表開始，包括模型使用的特殊標記和初始字母表。WordPiece 將文本切分成子詞單位，如「學習中文」可能被分成 ['學', '##習', '中', '##文']，且每個單詞最初是通過將該前綴添加到單詞內的所有字符來拆分的，前綴 ## 表示這個 token 是前一個 token 的延伸。

2.1.4 模型架構

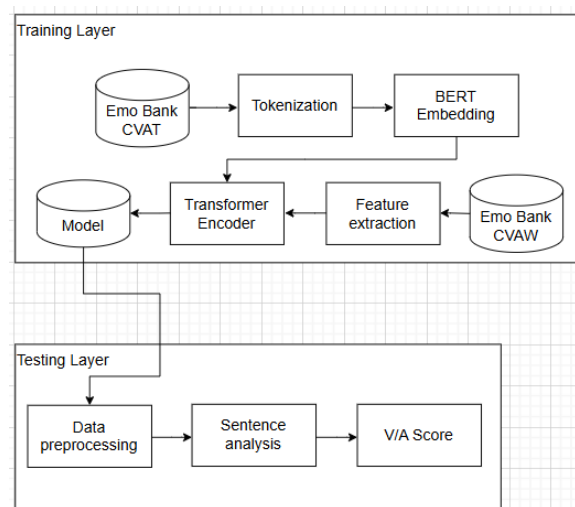


圖 1. 模型架構圖

2.2 Tokenization and Encoding

在本研究中採用 BERT 中文模型 (bert-base-chinese) 內建的 WordPiece 分詞系統對輸入文本進行分詞。模型在輸入文本時首先被轉換為 token 序列，並限制最大長度為 256。並且每個 token 帶有位置的編碼以及注意力遮罩 (attention mask)，以符合 BERT 的輸入格式。

2.3 Feature extraction

在本次任務中使用了 CVAW 作為詞典。當文本輸入並進行分詞後，模型將會檢查文本中的詞彙是否存在詞典中，若存在將提取相對應的 Valence 和 Arousal 的特徵，若不存在將會將數值補充為 [0.0, 0.0]，並生成對應的情緒特徵向量，供模型後續使用。

2.4 Word embedding

詞嵌入 (Word Embedding) 是一種詞彙表示方法，透過將詞語嵌入到實數向量空間中，將離散的文字轉換為連續數值，並使語意相近的詞語在向量空間中具有相近的位置表示[1]。

在本研究中採用的 BERT 模型中，詞嵌入 (Word Embedding) 可以將文字符號轉為模型可處理的向量表示。在 BERT 中包含了三種嵌入，分別是詞嵌入 (Token Embeddings)、分段嵌入 (Segment Embeddings) 和序列位置的嵌入 (Position Embeddings)。詞嵌入負責將詞彙轉換為向量表示，分段嵌入用於區分不同句子的片段，而位置嵌入則為序列中的每個 token 添加位置訊息。三者相加後形成最終的輸入嵌入向量，進一步輸入模型中的編碼器，完成後續的上下文語意建模。

為了更好的得到文本中的情緒特徵，本研究在 BERT 原始的嵌入基礎上額外使用了 CVAW 中文情緒字典的情緒向量 (Valence 與 Arousal 分數)，並將其與 BERT 的嵌入層進行融合。使得模型在獲取上下文語意的同時，輸入來自 CVAW 中文字的情緒特徵，為每個 Token 提供額外的 Valence 和 Arousal 的資訊。兩種向量在嵌入層融合後再經由線性層進行回歸，最終同時輸出 Valence 和 Arousal 的預測值。

2.5 Multi task

本研究將 Valence 與 Arousal 的預測視為一個多任務迴歸問題。對於輸入的文本，模型會同時輸出 Valence 預測值與 Arousal 預測值。故採用了多任務加權去強化模型在學習計算 Valence 跟 Arousal 的權重，並使用均方誤差 (MSE) 作為基礎去計算損失函數。其中均方誤差 MSE 和損失函數 (Loss) 的計算如下列公式所示。

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - y'_i)^2$$

$$Loss = \alpha * MSE(Valence) + \beta * MSE(Arousal)$$

MSE 公式中，其中 n 代表樣本數量， y_i 代表實際分數， y'_i 代表模型預測出的分數。在 Loss 公式中， α 代表 Valence 在模型學習的權重， β 則代表 Arousal 的權重，在本研究中所占權重分別為 0.6 和 0.4，此設計使模型在訓練過程

中能更側重於效價的準確預測，同時兼顧喚醒度的表現，達到兩者間的平衡。

3 EXPERIMENTS

3.1 Datasets

本次任務所使用的資料集為中文情緒資料集 Chinese Emobank [7] 中的中文情緒文本 (CVAT) 及中文情緒字典 (CVAW)。其中的資料包含中文單字 (或文本) 並包含了情緒效價 (Valence) 和喚醒度 (Arousal) 的分數，均以浮點數標註。情緒效價表示情緒的正面和負面情緒的程度，喚醒度表示平靜和興奮的程度。兩個維度的數值範圍均為 1 (非常消極或平靜) 到 9 (非常積極或興奮)。

中文情緒字典 (CVAW) 共包含了 5,512 個單字，中文情緒文本 (CVAT) 共包含了 2,969 句中文文本。

3.2 Authentication and Evaluation

本研究使用了皮爾森相關係數 (PCC) 和平均絕對誤差 (MAE) 作為評估模型的指標。其中 PCC 是一個介於 -1 和 1 之間的值，用於衡量實際值和預測值之間的線性相關性；而 MAE 的理論範圍最低值為 0、最高值為 8。而越低的 MAE 值和越高的 PCC 值，代表了模型預測效能準確越高。兩者指標定義如下列公式所示。

$$MAE = \frac{1}{n} \sum_{i=1}^n |a_i - p_i|$$

$$PCC = \frac{1}{n-1} \sum_{i=1}^n \left(\frac{a_i - \mu_A}{\sigma_A} \right) \left(\frac{p_i - \mu_P}{\sigma_P} \right)$$

在 MAE 的公式中，其中 n 代表樣本數量， a_i 為預測值， p_i 是真實值。在 PCC 的公式中，其中 $\left(\frac{a_i - \mu_A}{\sigma_A} \right)$ 、 μ_A 及 σ_A 分別是 a_i 的樣本的標準分數、樣本平均值和樣本標準差。

4 RESULTS

4.1 Validation dataset

在本次任務中，此模型在驗證集的表現如下表所示。評估結果顯示 Valence 和 Arousal 的

PCC 分別是 0.716、0.508；而 MAE 的分數分別為 0.613、1.079。

Valence PCC	Valence MAE	Arousal PCC	Arousal MAE
0.716	0.613	0.508	1.079

表 1. 驗證集中的表現

4.2 Test dataset

在本次任務中，此模型的在測試集實驗結果如下表所示，評估結果顯示 Valence 和 Arousal 的 PCC 分別是 0.76、0.58；而 MAE 的分數分別為 0.53、0.82。在本次競賽[8]中的最終排名得到了第五名的成果。

Valence PCC	Valence MAE	Arousal PCC	Arousal MAE
0.76	0.53	0.58	0.82

表 2. 實驗結果

5 CONCLUSIONS

在本次的任務中，主要運用了 BERT 中文預訓練模型，其中用了 BertTokenizer 進行分詞，從文本的詞彙中提取特徵向量並融合透過詞嵌入得到的文字向量，從而得到 MAE 及 PCC 的數值。此外，此研究也透過多任務加權的方式來調整模型在 Valence 和 Arousal 分數上學習的權重讓模型可以更好的針對不足的方面加強。在實驗過程中發現模型仍有需加強，使其有更好的結果。未來將運用於醫療領域，可更加清楚的得知醫療相關人員內心真實想法，也可作為醫護人員壓力偵測與心理輔助的基礎模型。

References

Agrawal, Ankur N. 2021. *Introduction to word embeddings*. In *Hands-on Question Answering Systems with BERT*. Berkeley, CA: Apress.

Calvo, R. A., & Kim, S. M. 2013. *Emotions in text: Dimensional and categorical models*. *Computational Intelligence*, 29(3), 527–543.

Cui, Yiming, Wanxiang Che, Ting Liu, Bing Qin, and Ziqing Yang. 2021. *Pre-training with whole word masking for Chinese BERT*. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 29:3504–3514.

Devlin, Jacob, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. *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 (NAACL-HLT 2019)*, pages 4171–4186, Minneapolis, Minnesota. Association for Computational Linguistics.

Grefenstette, Gregory. 1999. *Tokenization*. In *Syntactic Wordclass Tagging*, pages 117–133. Dordrecht: Springer Netherlands.

Koroteev, M. V. 2021. *BERT: A review of applications in natural language processing and understanding*. arXiv preprint arXiv:2103.11943.

Lee, Lung-Hao, Jian-Hong Li, and Liang-Chih Yu. 2022. *Chinese EmoBank: Building valence–arousal resources for dimensional sentiment analysis*. *ACM Transactions on Asian and Low-Resource Language Information Processing*, 21(4):Article 65.

Lee, Lung-Hao, Tzu-Mi Lin, Hsiu-Min Shih, Kuo-Kai Shyu, Anna S. Hsu, and Peih-Ying Lu. 2025. *ROCLING-2025 Shared Task: Chinese dimensional sentiment analysis for medical self-reflection texts*. In *Proceedings of the 37th Conference on Computational Linguistics and Speech Processing (ROCLING 2025)*.

Rietzler, Alexander, Sebastian Stabinger, Paul Opitz, and Stefan Engl. 2019. *Adapt or get left behind: Domain adaptation through BERT language model finetuning for aspect-target sentiment classification*. arXiv preprint arXiv:1908.11860.

Wu, Yichao, Zhengyu Jin, Chenxi Shi, Penghao Liang, and Tong Zhan. 2024. *Research on the application of deep learning-based BERT model in sentiment analysis*. arXiv preprint arXiv:2403.08217.

Wu, Yonghui, Mike Schuster, Zhifeng Chen, Quoc V. Le, Mohammad Norouzi, Wolfgang Macherey, ... and Jeff Dean. 2016. *Google's neural machine translation system: Bridging the gap between human and machine translation*. arXiv preprint arXiv:1609.08144.

Yu, Liang-Chih, Lung-Hao Lee, Shuai Hao, Jin Wang, Yunchao He, Jun Hu, K. Robert Lai, and Xuejie Zhang. 2016. *Building Chinese affective resources in valence–arousal dimensions*.