

# SCUDS at ROCLING-2021 Shared Task: Using Pretrained Model for Dimensional Sentiment Analysis Based on Sample Expansion Method

**Hsiao-Shih Chen**  
School of Big Data  
Management, Soochow  
University, Taiwan  
07170235@gm.scu.edu.tw

**Pin-Chiung Chen**  
School of Big Data  
Management, Soochow  
University, Taiwan  
07170140@gm.scu.edu.tw

**Shao-Cheng Huang**  
School of Big Data  
Management, Soochow  
University, Taiwan  
07170123@gm.scu.edu.tw

**Yu-Cheng Chiu**  
School of Big Data  
Management, Soochow  
University, Taiwan  
jean199925@gmail.com

**Jheng-Long Wu**  
Department of Data Science,  
Soochow University, Taiwan  
jluwu@gm.scu.edu.tw

## 摘要

情感分析已成為現今非常熱門的研究議題，而在教育文本的情感分析仍是一大重點。根據文獻指出，產生極為相似的語句，將有效幫助機器學習的學習成效，因此可被控制的訓練樣本擴充法將成為預測模型提升效果的主要元素。本研究將提出採用詞性篩選結合 Word2Vec 模型計算相似字詞來擴充訓練樣本，以達到相似字詞替換後，仍保有較高品質的情感表達。而在預測模型方面，本研究採用 DistilBERT 預訓練模型作為基礎，再進行句子的 Valence-Arousal 情感分數學習與預測。根據實驗結果顯示，採用本研究所提出的擴充訓練樣本時，可以獲得降低 80% MAE 和提升 28% PCC。

## Abstract

Sentiment analysis has become a popular research issue in recent years, especially on educational texts which is an important problem. According to literature, the similar sentence generation can help the prediction performance of machine learning. Therefore, the process of controlled expansional samples is a key component to prediction models. The paper proposed a sample expansion method which combined part-of-speech filter and similar word finder of Word2Vec. The generate samples have high quality with similar sentiment representation. The DistilBERT pretrained model is used to learn and predict Valence-Arousal scores

from the expansion samples. Experimental result displays that the using the expansion samples as training data into prediction model has outperforms original training data without expansion, and obtains 80% mean square error reducing and 28% pearson correlation coefficient increasing.

關鍵字：情感分析、預訓練模型、樣本擴充法  
Keywords: Sentiment analysis, Pretrained model, Sample expansion method

## 1 緒論

### 1.1 背景與動機

隨著社群媒體的快速發展，以及網路資源、行動裝置的普及，造就現今社會時時刻刻都能透過各大社群網站獲取大量資訊，包括新聞網站、影音平台與直播平台等都有提供留言或即時聊天的功能，讓世界各地的人能在偌大的網路虛擬世界進行交流。在各大論壇上，曾經發表過的言論，都會被記錄下來，這些大量的文字資訊，經常在自然語言領域，當作是模型訓練的語料庫。

情感分析是一種用來識別文本中情感資訊的重要技術。通常會針對一個文本作兩極性的分類。分類的目的在於判斷文字中表達的情緒是正向 (positive)、負向 (negative) 或中性 (neutral)。在早期情感分析領域提出研究主要是在探測商品評論與電影影評的兩極觀點 (Turney, 2002)。而另一種方法是以維度的方式將情感狀態以連續數值表示在多個維度上，像是 Valence-Arousal (VA) (Russell,

1980)。Valence 代表環境帶給人類的情緒影響，當值為正向時，代表歡樂、和平與喜悅等正面情緒；值為負時，代表生氣、無趣與悲傷等負面情緒。Arousal 代表情緒的強烈、壓力程度，當值為高時，代表快樂、興奮；值為低時，代表憂鬱、無聊。本研究將採用 ROCLING 2021 Shared Task 所提供的語料庫，其中包含以句子為單位的 CVAT 2.0 (Yu et al., 2016)、以單詞為單位的 CVAW 4.0 (Yu et al., 2016) 以及以片語為單位的 CVAP 2.0 (Yu et al., 2017)。由於本研究要解決句子型態的情感狀態，即預測句子的 VA 值，則使用 CVAT 2.0 語料庫作為本研究的主要訓練資料，包含來自六個不同類別的網路評論，每個句子手動標記成二維的向量，分別代表情緒的正負面及語氣的興奮程度，共計 2,969 筆句子。而 CVAW 4.0 語料庫有 5,512 個單詞，以及 CVAP 2.0 語料庫則有 2,998 個片語。

由於以句子為單位的語料庫僅有 2,969 筆，其樣本數略顯不足。根據過去的相關訓練樣本數不足的研究，已經證實擴充訓練樣本後，能夠提供更多樣性的訓練樣本，使得機器學習模型可以從中學學習更多的樣本，避免機器學習模型只能參考少部分樣本分布，使得機器學習模型在預測的精確度有所提升 (Shorten and Khoshgoftaar, 2019)。而在自然語言處理相關研究中，深度學習 (Deep Learning) 的 BERT 模型 (Devlin et al., 2018) 已經取得相當大幅度的改善，透過預訓練的方式，從超大量語料庫中學習後，再僅透過一個額外的輸出層進行後續任務微調或特定架構修改等就能夠獲得預測效果的改善。而另一個傳統的機器學習預測模型，則是使用 TF-IDF (Term Frequency-Inverse document Frequency) (Ramos, 2003) 將文字轉換成特徵向量後，再透過全域求解的 SVM (Support Vector Machine) (Cortes and Vapnik, 1995) 演算法來建立模型。為了有效預測以句子為單位的教育文本之情感分數，本研究將採取多種訓練樣本擴充法搭配機器學習及深度學習的模型進行訓練與預測。所以自動化預測句子的情感分數，相關單位可以透過分數的高低來判斷該句子可能想要表達的情緒，甚至能分析學生的學習狀況。

## 1.2 目的

圖 1 為本研究的實驗流程圖，為了要提高模型預測的準確度，本次任務將 CVAT 2.0 語料庫透過斷詞及詞性標注後，再找出情感單詞的近義詞和計算情感單詞與近義詞的相似度分數，並重新計算替換過後的句子 VA 值，因此透過替換方式所擴充出相似句子，可用於訓練樣本的資料量。再者，透過 DistilBERT 預訓練模型及 SVM 模型進行情感分數的訓練與預測。本研究藉由上述所提出的流程來訓練與預測教育文本 (學生自我評價) 的句子在維度情感分析任務中應有的分數。本研究的目的是為建構一套樣本擴充法與採用強大的預測模型，可以從網路評論的訓練樣本學習到規則，進而用來預測教育文本的資料集，即學生自我評價的句子所表達的 Valence 和 Arousal 值。

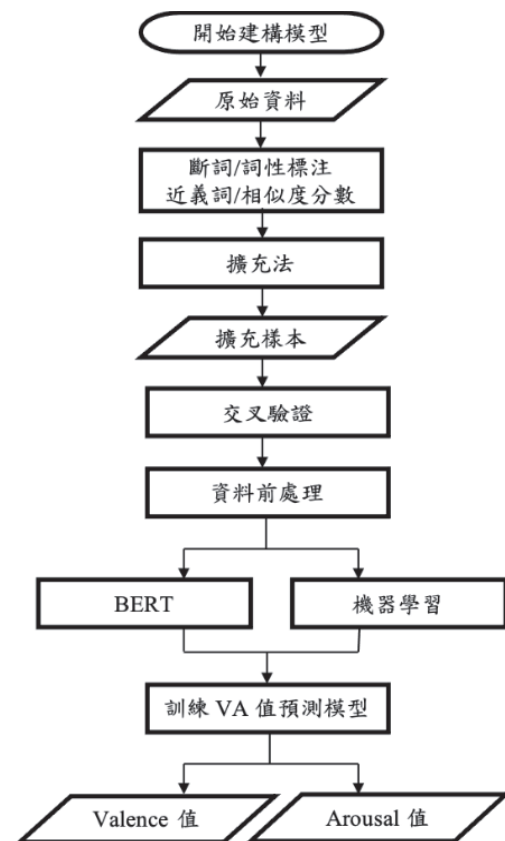


圖 1：研究實驗流程圖

## 2 系統方法

本研究的系統方法包括資料擴充法以及預測模型建構。資料擴充法使得模型有更多的訓練資料量可以用來學習，而再訓練機器學習

與深度學習的模型獲得預測模型，並透過交叉驗證法得出表現最好的模型來做為最終預測使用。

## 2.1 資料擴充法

首先參考 EDA (Easy Data Augmentation techniques) 模型 (Wei and Zou, 2019)，此為一種傳統的文本擴充方法，可以以四種方式來進行資料擴充，包含：同義詞替換、隨機插入、隨機交換與隨機刪除。接著參考 Word2Vec 模型 (Mikolov et al., 2013) 的字詞相似度概念，使用一個向量來表示一個詞，將文本空間中的某個 word，透過語言模型的訓練後，將該詞相關的資訊嵌入 (embedding) 於所屬向量空間，因此可以透過餘弦相似度 (cosine-similarity) 方法來計算詞與詞之間的相似程度，進而做到找尋相近詞的功用。

### 2.1.1 斷詞與詞性標注

多數與英文相關的語言，字詞與字詞之間都有明顯的空白做為分隔，所以很容易就能判斷出單詞。然而，中文則是一長串文字的組合，字詞與字詞間沒有明顯的符號來做區隔，因此斷詞為文字預處理不可或缺的過程。本研究使用 Python Synonyms 套件<sup>1</sup>，來進行斷詞與詞性標注，這樣可以針對特定的詞性作為替換的規則，斷詞與詞性標注範例如圖 2 所示。

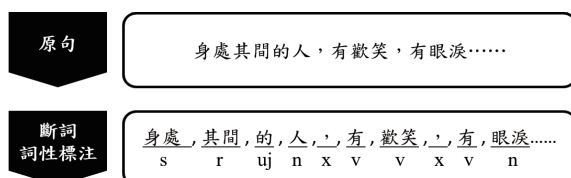


圖 2：斷詞及詞性標注

### 2.1.2 近義詞與餘弦相似度分數

為了達成產生相似句子的目的，本研究將採取找尋合適的近義詞，以替換句子斷詞後的單詞，此可產生一句新的擴充句子。例如單詞「心碎」可與其近義單詞計算餘弦相似度，分數愈接近 1 則代表該詞與「心碎」的語意愈

相近，如「心痛」的餘弦相似度值為 0.76，其餘相似單詞的相似度如圖 3 所示。

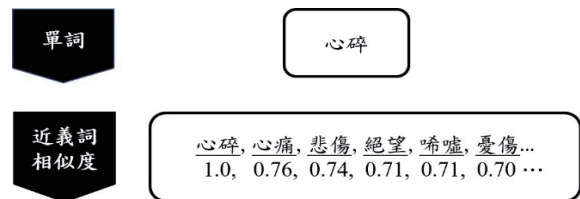


圖 3：近義詞及相似度分數

本研究預計透過四種方法來擴充原始資料，並會使用同一原始句子來做說明，此範例的原始句子如：

- 身處其間的人，有歡笑，有眼淚，有快樂，有悲傷，有甜蜜，有心碎，可是應該是不會後悔的吧。
- Valence：5.75、Arousal：5。

以下小節是擴充法說明：

### 2.1.3 擴充法 1

當句子斷詞後的單詞出現在 CVAW 4.0 單詞語料庫時，找尋其要被替換的單詞之 VA 值，相差正負 0.1 內，便可替換該詞，以產生新的擴充句子，其擴充句子的 VA 值將與原始句子相同。以下為擴充法 1 所產生的擴充句子（粗體表示替換過的單詞）與其對應的 VA 值：

- 身處其間的人，有**歡喜**、有眼淚，有快樂，有**傲慢**，有**強壯**，有心碎，可是應該**是犯錯乞討**的吧。
- Valence：5.75、Arousal：5。

### 2.1.4 擴充法 2

基本操作同擴充法 1 相似，差別在於指定**動詞**為範圍找出可被替之同義詞，且符合餘弦相似度大於 0.7，並採取隨機替換。而情感分數方面，以原始詞與被替換詞的各別 VA 值之差乘以 5% 作為新 VA 值調整。以下為擴充法 2 所產生的擴充句子與其對應的 VA 值：

- 身處其間的人，有**歡樂**，有眼淚，有快樂，有悲傷，有甜蜜，有**憂傷**，可是應該**是無法懊悔**的吧。
- Valence：5.79、Arousal：4.99

<sup>1</sup> 中文近義詞工具包 Synonyms. 網址：  
<https://github.com/chatopera/Synonyms>



### 2.1.5 擴充法 3

基本操作與擴充法 2 相似，差別在於指定**名詞與動詞**為範圍找出可被替之同義詞。以下為擴充法 3 所產生的擴充句子與其對應的 VA 值：

- 身處其間的人，有**歡樂**，有眼淚，有快樂，有悲傷，有甜蜜，有**心痛**，可是應該是**無法生氣**的吧。
- Valence：5.73、Arousal：5.1

### 2.1.6 擴充法 4

基本操作與擴充法 2 相似，差別在於指定**名詞、動詞、名詞+動詞**為範圍，找出被替換詞之同義詞，一次替換一個詞至一次替換五個詞，排列所有可能。以下為擴充法 4 所產生的擴充句子與其對應的 VA 值：

- 身處其間的人，有**歡樂**，有**流淚**，有快樂，有悲傷，有甜蜜，有**憂傷**，可是應該是**不能慚愧**的吧。
- Valence：5.81、Arousal：4.86

## 2.2 交叉驗證

為了得到可靠且穩定的模型，本研究使用交叉驗證法來評估模型的學習效果，避免資料的分布不均所造成的學習效果不佳。本研究以 5-fold 交叉驗證將資料分成訓練集與驗證集，就是把資料切成五個子集合，其中一個子集合需被保留作為驗證，其餘四個子樣本用來訓練，重複五次，每個子集合都必須輪流作為驗證集一次。然而，先前提到的資料擴充法所產生的擴充句子，將作為訓練樣本來加強模型訓練，所以在每個 fold 中，只會在訓練樣本額外加上擴充句子進行訓練，並不會將擴充資料版本納入驗證子樣本進行驗證，以避免擴充資料影響整體模型驗證效果。

## 2.3 預測模型訓練

### 2.3.1 DistilBERT 模型與超參數設定

本研究使用 CVAT 2.0 句子語料庫作為訓練集，將輸入句字進行斷詞 (Tokenizer)，也就是轉換成 BERT 模型當中單個 Token 的編號，並且統一每個輸入的長度，未滿長度者以 0 補齊，最多為 512 個單詞。再者透過 Hugging Face 的套件 (Wolf et al., 2020) 所提供的預訓練模型

DistilBERT (distilbert-base-multilingual-cased) 作為預測模型基礎。

並透過調整模型超參數以取得最佳模型，本研究分別定義 Batch Size, Learning Rate 和 Epochs 等三個超參數，總共訂定六組超參數，表 1 為 DistilBERT 所使用的超參數設定。

Experiment	Batch size	Learning rate	Epochs
EXP1	16	5e-2	50
EXP2	16	5e-3	50
EXP3	16	5e-5	50
EXP4	8	5e-2	50
EXP5	8	5e-3	50
EXP6	8	5e-5	50

表 1：DistilBERT 模型之超參數設定

### 2.3.2 SVM 機器學習模型與超參數設定

SVM 模型訓練則將所有輸入句子採用 CKIP-Transformers (Li et al., 2020) 工具進行斷詞，接著訓練資料建立詞彙表，再計算各樣本的 TF-IDF 值。而 SVM 透過調整不同超參數以取得最佳模型，SVM 的超參數設定如表 2 所示。

Kernel	Cost	Degree	Gamma
Linear	1	-	auto
RBF	1	-	auto
	1	-	scale
	20	-	0.5
Sigmoid	1	-	auto
	1	-	scale
	10	-	0.1
Poly	1	3	auto
	1	3	scale
	10	3	0.5
	10	3	1
	10	3	2
	10	3	0.1

表 2：SVM 模型之超參數設定

## 2.4 評估指標介紹

根據 ROCLING 2021 Shared Task 所示，為了評估模型的表現，使用平均絕對誤差 (Mean Square Error, MAE) 與皮爾森相關係數

(Pearson Correlation Coefficient, PCC)，如公式(1)與公式(2)所示。

- MAE：計算各次真實值與預測值誤差取絕對值後再求平均值。

$$MAE = \frac{1}{n} \sum_{i=1}^n |A_i - P_i| \quad (1)$$

- PCC：表示真實值與預測值是否共同產生變化的關聯程度。

$$PCC = \frac{1}{n} \sum_{i=1}^n \left( \frac{A_i - \bar{A}}{\sigma_A} \right) \left( \frac{P_i - \bar{P}}{\sigma_P} \right) \quad (2)$$

$A_i$  為手動標記評分， $P_i$  為模型預測評分， $n$  為測試樣本數量， $\bar{A}$  和  $\bar{P}$  分別為手動標記評分與模型預測評分的總平均， $\sigma$  為標準差。

### 2.4.1 最佳模型挑選準則

先前提到為了避免模型訓練偏差，使用 5-fold 交叉驗證來穩定模型，而評估指標計算方法則是將得到的五次 MAE 與 PCC 之結果各自取平均值作為此超參數最終的評估指標，以挑選出最佳模型。

## 3 實驗結果與比較

### 3.1 擴充句子統計量

將 CVAT 2.0 句子語料庫 (2,969 筆) 做為版本一的訓練資料版本，版本二至版本五分別使用擴充法 1 至 4 來生成的訓練擴充樣本分別加上版本一，最後版本六為合併版本一至版本五的所有資料，並刪除重複生成句子 (共計移除 9,977 筆)。表 3 為各版本之擴充量統計數據。

資料版本	規則	總數量 (+擴充量)
版本一	原始擴充資料集	29,69 (+0)
版本二	使用擴充法 1	5,444 (+2,475)
版本三	使用擴充法 2	7,008 (+4,039)
版本四	使用擴充法 3	6,921 (+3,952)
版本五	使用擴充法 4	31,433 (+28,464)
版本六	合併第一至第五版， 並刪除重複句子	34,892 (+31,923)

表 3：訓練資料版本規則與統計量

### 3.2 DistilBERT 模型之預測效果

本研究使用 DistilBERT 預訓練模型作為此 VA 預測模型的起點，表 4 與表 5 為統整每個資料版本中表現最佳超參數組合，以及對應的 MAE 與 PCC，其中最佳超參數是以最低的 MAE 或最高的 PCC 篩選而得。然而進一步觀察得知無論是 Valence 還是 Arousal 表現最好的皆為資料版本六，和版本一 (無擴充資料集) 相比，大幅降低了 MAE 評估指標，Valence 從 0.6723 降至 0.1378，Arousal 則從 0.7265 降至 0.1436，兩個預測目標幾乎都降了 80%。

資料版本	Valence		
	Experiment	MAE	PCC
版本一	EXP3	0.672	0.757
版本二	EXP3	0.224	0.957
版本三	EXP3	0.310	0.915
版本四	EXP3	0.264	0.931
版本五	EXP3	0.296	0.917
版本六	EXP3	<b>0.138</b>	<b>0.971</b>

表 4：DistilBERT 之 Valence 預測結果

資料版本	Arousal		
	Experiment	MAE	PCC
版本一	EXP3	0.727	0.469
版本二	EXP6	0.247	0.897
版本三	EXP3	0.341	0.829
版本四	EXP3	0.292	0.867
版本五	EXP3	0.306	0.850
版本六	EXP3	<b>0.144</b>	<b>0.939</b>

表 5：DistilBERT 之 Arousal 預測結果

### 3.3 SVM 機器學習之預測效果

由 DistilBERT 預訓練模型可得知資料版本六的預測效果大幅領先，因此在 SVM 模型的訓練將直接採用資料版本六進行訓練，根據不同的超參數組合，最佳模型的超參數設置為：Kernel 採用 RBF，Gamma 設定 0.5、Cost 設定 20，其設定在預測 VA 值皆得到最好的效果，如表 6 所示。與最差的 SVM 模型比較下，最佳的 SVM 模型在 Valence 改善約 0.10 的 MAE，

Arousal 則是改善約 0.05 的 *MAE*。有此可知，再不同的超參數組合下，對 SVM 模型沒有產生極大差異，顯示 SVM 模型相當穩定。

Target	<i>MAE</i>	<i>PCC</i>
Valence	0.235	0.952
Arousal	0.218	0.919

表 6：SVM 之預測結果

### 3.4 SVM 與 DistilBERT 模型之預測結果比較

表 7 為將 SVM 和 DistilBERT 模型所訓練的最佳成果比較，可以看到 DistilBERT 模型的 *MAE* 都比 SVM 要來得低，在 Valence 的部分，兩模型的 *MAE* 相差約 9.7%；在 Arousal 的部分則相差約 7.4%，因此可以得知 DistilBERT 的表現相較於 SVM 傳統機器學習模型來的佳，且大幅改善。

Model	Valence		Arousal	
	<i>MAE</i>	<i>PCC</i>	<i>MAE</i>	<i>PCC</i>
DistilBERT	0.138	0.971	0.144	0.939
SVM	0.235	0.952	0.218	0.919

表 7：DistilBERT 與 SVM 模型預測結果之比較

### 3.5 不同 BERT 的預訓練模型之比較

由於官方提供的資料集與本研究的擴充資料版本皆為繁體中文，DistilBERT 是一種多語言的預訓練模型，為了探究預訓練模型的文本語言是否會影響 VA 的預測效果，所以本研究利用中央研究院開發繁體中文的預訓練模型 CKIPlabBERT<sup>2</sup> (ckiplab/bert-base-chinese) 再次作為預訓練模型，並訓練本研究的擴充資料。首先採用資料版一作為基礎，比較無擴充訓練樣本的預測效果。圖 4 表示在無擴充資料版本一的比較結果，CKIPlabBERT 都比 DistilBERT 模型來得優秀，在 VA 的 *MAE* 分別

差 0.152 和 0.027，而 *PCC* 則是分別差 0.1 和 0.051。

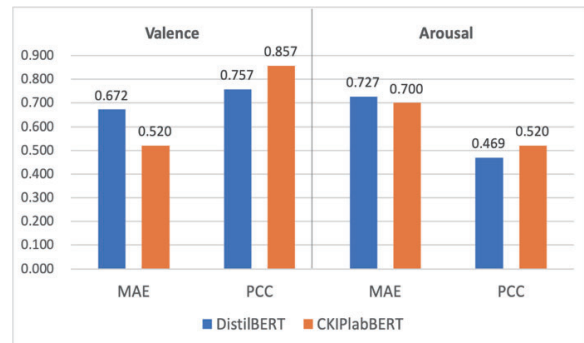


圖 4：版本一預訓練模型比較

然而，圖 5 表示有擴充資料的版本六的比較結果，反而是 DistilBERT 比 CKIPlabBERT 來的優秀，在 VA 的 *MAE* 分別差 0.016 和 0.035，而 *PCC* 則是分別差 0.004 和 0.003。但相較於無擴充資料版本一時，其差距變得非常小。

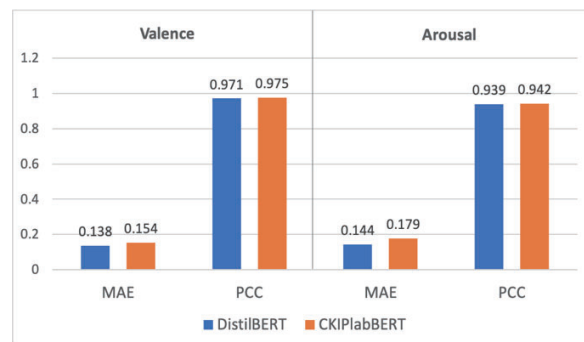


圖 5：版本六預訓練模型比較

### 3.6 最佳模型選擇

模型訓練完成後，以驗證集作為基準，從 SVM、DistilBERT 和 CKIPlabBERT 中挑選最佳模型作為最終繳交使用之預測模型。本研究選擇 DistilBERT 和 CKIPlabBERT 兩模型搭配先前表現最好的超參數組合 EXP3 和 EXP6，作為預測模型進行比較，並使用無擴充資料的版本一作為預測集。由表 8 與表 9 可見，在 DistilBERT 模型中，VA 值的驗證效果皆為超參數組合 EXP3 較好，而在預測集的結果，其 *MAE* 又更低，兩者相差約 9%。而在 CKIPlabBERT 模型中，Valence 的驗證集效果

<sup>2</sup> CKIP LAB 所開發之 CKIP Transformers 為繁體中文的 transformers 模型，網址：  
[https://huggingface.co/CKIP\\_lab](https://huggingface.co/CKIP_lab)

為超參數組合 EXP3 較好，其在預測集的結果亦為 *MAE* 較低，兩者相差了約 6.87%；Arousal 的驗證集則是超參數組合 EXP6 較好，其在預測集的結果亦為 *MAE* 較低，兩者相差了約 8.72%。然而兩個模型在預測 VA 值的效果上，超參數組合不論是 EXP3 或 EXP6 皆是 DistilBERT 模型表現得較出色，其整體在預測集的 *MAE* 都是 CKIplabBERT 模型的約 0.5 倍，所以根據前述的最佳模型挑選準則，最後選擇了 DistilBERT 模型搭配超參數組合 EXP3 與 EXP6 作為兩個最終模型。

Model / Experiment	驗證集		預測集	
	<i>MAE</i>	<i>PCC</i>	<i>MAE</i>	<i>PCC</i>
DistilBERT / EXP3	0.138	0.971	0.045	0.998
DistilBERT / EXP6	0.142	0.967	0.045	0.998
CKIplabBERT / EXP3	0.154	0.975	0.085	0.999
CKIplabBERT / EXP6	0.174	0.970	0.084	0.998

表 8：DistilBERT 與 CKIplabBERT 模型在驗證集與預測集的 Valence 預測效果之比較

Model / Experiment	驗證集		預測集	
	<i>MAE</i>	<i>PCC</i>	<i>MAE</i>	<i>PCC</i>
DistilBERT / EXP3	0.144	0.939	0.044	0.997
DistilBERT / EXP6	0.159	0.935	0.052	0.996
CKIplabBERT / EXP3	0.179	0.944	0.094	0.997
CKIplabBERT / EXP6	0.171	0.938	0.084	0.997

表 9：DistilBERT 與 CKIplabBERT 模型在驗證集與預測集的 Arousal 預測效果之比較

### 3.7 測試集評估結果

根據 ROCLING 2021 Shared Task 提供的測試集，所提交的兩個預測模型為 DistilBERT 搭配超參數組合 EXP3 與 EXP6，其預測結果如

表 10 所示，在 Valence 的 *MAE* 最低為 0.953 和 *PCC* 最高為 0.694；在 Arousal 的 *MAE* 最低為 1.039 和 *PCC* 最高為 0.375。

Model / Experiment	Valence		Arousal	
	<i>MAE</i>	<i>PCC</i>	<i>MAE</i>	<i>PCC</i>
DistilBERT / EXP3	0.953	0.694	1.054	0.375
DistilBERT / EXP6	0.975	0.667	1.039	0.354

表 10：DistilBERT 模型與 EXP3 和 EXP6 超參數組合在測試集的評估結果

然而，ROCLING 2021 Shared Task 釋出測試集答案後，本研究也將 CKIplabBERT 模型進行預測，並進行評估，其評估結果如表 11 所示，在 Valence 很明顯可以看出比 CKIP\_LAB\_BERT 比 DistilBERT 模型表現來得好，*MAE* 降低約 20%，*PCC* 增加約 17%，而 Arousal 的 *MAE*，CKIplabBERT 降低約 5%，*PCC* 也增加約 17%。

Model / Experiment	Valence		Arousal	
	<i>MAE</i>	<i>PCC</i>	<i>MAE</i>	<i>PCC</i>
CKIplab_BERT / EXP3	0.767	0.814	0.983	0.480
CKIplabBERT / EXP6	0.857	0.766	0.982	0.443

表 11：CKIplabBERT 模型與 EXP3 和 EXP6 超參數組合在測試集的評估結果

由於在提交測試集的預測結果時，無從得知測試集答案，就標準實驗流程而言，根據本研究中預測集所表現最好的 DistilBERT 模型作為當前的最佳模型，並提交該最佳模型在測試集的預測結果，固然在測試集的效果是 CKIplabBERT 模型較佳，但也無法作為最終被挑選為最終預測模型。

## 4 結論

本研究已經提出了一種有效的訓練樣本擴充方法。根據情感字詞的特性，透過預訓練的 Word2Vec 模型找尋相似字詞，以及與情感較



相關的詞性進行篩選，限縮替換範圍，將可以使替換之字詞聚焦於情感方面，避免過度發散，意味著有效的控制替換字詞的品質，也就可以產生高品質的擴充句子。除了訓練樣本擴充可帶來效果外，本研究也再次證實，相較於機器學習模型，較為新進的 DistilBERT 模型可以有效的提升 VA 的預測效果，甚至根據測試集的評估可以得知，專門為繁體中文所預訓練的 CKIPLabBERT 模型可以更進一步獲得改善。總之，未來研究仍可以朝向資料擴充法的探索，像是採用 Generative Adversarial Network 作為仿生句子的訓練，達到更具多樣性的擴充句子，並保有合適的 VA 情感分數。

### 參考文獻

- Connor Shorten and Taghi M. Khoshgoftaar. 2019. A survey on Image Data Augmentation for Deep Learning. *Journal of Big Data*, 6, Article number: 60. <https://doi.org/10.1186/s40537-019-0197-0>.
- Corinna Cortes and Vladimir Vapnik. 1995. Support-Vector Networks. *Machine Learning volume 20*, pages 273–297.
- James A. Russell. 1980. A circumplex model of affect. *Journal of Personality and Social Psychology*, 39(6):1161. <https://doi.org/10.1037/h0077714>.
- Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. *BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding*. *Computation and Language*. arXiv:1810.04805. Version 2.
- Juan Ramos. 2003. Using TF-IDF to Determine Word Relevance in Document Queries.
- Jason Wei and Kai Zou. 2019. *EDA: Easy data augmentation techniques for boosting performance on text classification tasks*. *Computation and Language*, arXiv:1901.11196. Version 2.
- Liang-Chih Yu, 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. In *Proceedings of NAACL/HLT-16*, pages 540-545.
- Liang-Chih Yu, Lung-Hao Lee, Jin Wang, and Kam-Fai Wong. 2017. IJCNLP-2017 Task 2: Dimensional Sentiment Analysis for Chinese Phrases. In *Proceedings of IJCNLP-17, Shared Tasks*, pages 9-16.
- Peter Turney. 2002. Thumbs Up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Reviews. In *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics, Philadelphia, Pennsylvania*, pages 417-424. <https://doi.org/10.3115/1073083.1073153>.
- Peng-Hsuan Li, Tsu-Jui Fu, and Wei-Yun Ma. 2020. Why Attention? Analyze BiLSTM Deficiency and Its Remedies in the Case of NER. In *Proceedings of AAAI 2020. Computation and Language*, arXiv:1908.11046. Version 3.
- Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. *Efficient estimation of word representations in vector space*. *Computation and Language*, arXiv:1301.3781. Version 3.
- Thomas Wolf, Lysandre Debut, Victor Sanh, Julien Chaumond, Clement Delangue, Anthony Moi, Pierric Cistac, et al. 2020. Transformers: State-of-the-Art Natural Language Processing. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*. Association for Computational Linguistics, pages 38-45. <https://aclanthology.org/2020.emnlp-demos.6>.