

ntust-nlp-2 at ROCLING-2021 Shared Task: BERT-based semantic analyzer with word-level information

盧克函

Ke-Han Lu

國立臺灣科技大學

National Taiwan University
of Science and Technology

khlunlp@csie.ntust.edu.tw

陳冠宇

Kuan-Yu Chen

國立臺灣科技大學

National Taiwan University
of Science and Technology

kychen@ntust.edu.tw

摘要

本論文提出基於 BERT 架構之維度式情感辨識器，透過增加基於詞層級的資訊，我們的模型在「ROCLING 2021 共享任務：教學文本之維度式情感辨識」四項指標中拿到了三項最佳成績。通過一連串的實驗，我們比較了不同預訓練方法對於結果的影響，也證明了我們基於預訓練模型所提出的方法，顯著的改進了模型的表現，最後我們針對實驗結果，提出關於模型和資料集的深入分析與討論。

Abstract

In this paper, we proposed a BERT-based dimensional semantic analyzer, which is designed by incorporating with word-level information. Our model achieved three of the best results in four metrics on "ROCLING 2021 Shared Task: Dimensional Sentiment Analysis for Educational Texts". We conducted a series of experiments to compare the effectiveness of different pre-trained methods. Besides, the results also proved that our method can significantly improve the performances than classic methods. Based on the experiments, we also discussed the impact of model architectures and datasets.

關鍵字：情感辨識、預訓練語言模型、BERT

Keywords: Semantic analysis, Pre-trained Language Model, BERT

1 緒論

情感辨識一直是自然語言處理領域中相當熱門的研究方向，可分為分類式及維度式的情感辨識任務 (Calvo and Kim, 2013)。分類式的情感辨識為將情緒分為不同類別，如：快樂、悲傷、憤怒、厭惡等；維度式的情感辨識則是將情緒表示為不同維度中的不同尺度，如：可以將一個詞 (Wei et al., 2011; Malandrakis

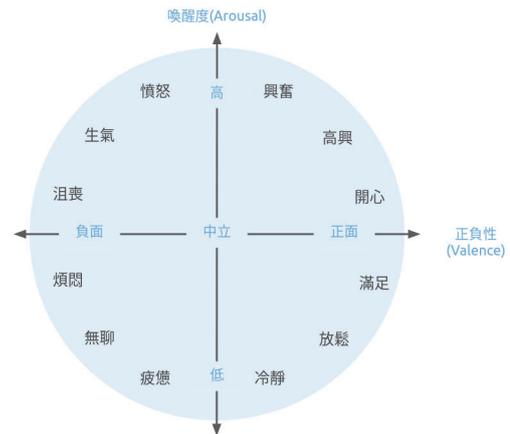


圖 1: 正負性－喚醒度空間

et al., 2013; Wang et al., 2016; Du and Zhang, 2016; Wu et al., 2017; Yu et al., 2020) 或一串句子 (Kim et al., 2010; Paltoglou et al., 2013; Goel et al., 2017; Zhu et al., 2019; Wang et al., 2020) 的情緒表示於以正負性 (Valence) 與喚醒度 (Arousal) 組成的二維空間中 (Russell, 1980)，如圖1所示，其中正負性代表情緒的正面、負面程度，而喚醒度是指情緒的激動程度。

在教育現場，老師們常需要評估學生的學習狀況，常見的方法有出席率、考試成績、課堂發言次數等結構性的資料，而非結構性的資料，像是學生的自我評述，不僅可能紀錄著上課與課後的點點滴滴，也可能含有學生對這門課與老師的評價與回饋，豐富的情感資訊通常蘊含其中。為了探究機器是否可以準確地自動分析這類文本，ROCLING 2021 提出了一個共享任務：教學文本之維度式情感辨識 (ROCLING 2021 Shared Task: Dimensional Sentiment Analysis for Educational Texts)，目標為對學生的評語文本進行維度式的情感評分，也就是需將一段句子映射到正負性－喚醒度的空間中。

在這個共享任務裡，主辦方所提供的資料集

為 CVAW 4.0(Yu et al., 2016)、CVAP 2.0(Yu et al., 2017) 及 CVAT 2.0(Yu et al., 2016)，分別為單詞、片語、句子所對應的正負性與喚醒度，我們將在第2章介紹資料集的蒐集方式與特性。評估模型的方式為預測分數與標記分數的平均絕對誤差 (Mean Absolute Error) 與皮爾森相關係數 (Pearson Correlation Coefficient)。第3章介紹我們所提出的基於預訓練 BERT(Devlin et al., 2019) 架構之維度式情感辨識器，除了句子以外，我們將詞層級的正負性與喚醒度作為額外的訓練目標去微調最終的模型。我們使用了不同預訓練方法，包含 BERT(Devlin et al., 2019)、RoBERTa(Liu et al., 2019) 與 MacBERT(Cui et al., 2020) 以驗證我們方法的有效性。第4章展示了我們模型參數設定及實驗細節。最後，在第5章我們分析了不同預訓練方法對於此任務的影響，也證明了我們所提出的方法有效的增強了模型的表現。在測試集的表現上我們的方法可以在四項指標中獲得其中三項指標最好的成績。此外，我們更進一步地對模型及資料集做了深入的討論。

2 資料集

本研究僅使用由主辦方提供的 CVAW 4.0、CVAP 2.0 及 CVAT 2.0，上述資料集為搜集文本後由數位標記員分別根據相同標準標記，在剔除標記差距過大的資料後整理而成，表1為從資料中選取的範例。

CVAW 4.0: 包含 5,512 個常見與情緒有關的單詞。

CVAP 2.0: 共 2,998 個以副詞、動詞、形容詞所組成之片語，包含不同程度 (如：極度、非常、略微等)、帶有懷疑 (如：也許、可能、應該等)、反向 (如：不、沒有、不能等) 等不同排列組合的修飾。

CVAT 2.0: 共 2,969 則對於書本、旅館評論、汽車討論、電腦討論、新聞、政治評論等六類的評價內容，大部份是網路上蒐集的句子，平均長度為 57.515，99.8% 句子中包含 CVAW 4.0 或 CVAP 2.0 所出現的單詞或片語。

3 方法

3.1 Baseline 模型

我們將維度式情感辨識任務視為一種迴歸任務，採用基於 BERT 架構的模型，以 CLS Token 的輸出，輸入迴歸器得到預測的結果。

更明確地，我們首先將一個輸入的中文句子表示為一連串的 Token 序列 $W = \{w_{cls}, w_1, \dots, w_{|T|}, w_{sep}\}$ ，其中 w_{cls} 與 w_{sep} 為兩個特殊的 Token 分別插入在句首與句尾。接著，經過嵌入層 (Embedding Layer)， W 可以轉換為一連串的向量表示法 (Embeddings) $E = \{e_{cls}, e_1, \dots, e_{|T|}, e_{sep}\}$ 。然後，我們將 E 輸入數層的 Transformer 後，以對應於 w_{cls} 之輸出 h_{cls} 作為此一輸入文本的向量表示法，藉由經過一層前饋神經網路 (Feed-forward Neural Network)，得到迴歸預測的數值：

$$H = \text{BERT}(E) \\ [y^v, y^a] = \text{FFN}(h_{cls})$$

$[y^v, y^a]$ 為模型所預測之正負性及喚醒度。為了達成此一目的，在模型訓練方面，我們使用均方根誤差 (Mean Square Error, MSE) 計算預測與答案之間的誤差，並以此作為模型訓練的目標函數：

$$\mathcal{L}_t = \text{MSE}([y^v, y^a], [\hat{y}^v, \hat{y}^a])$$

其中， \hat{y}^v 和 \hat{y}^a 表示正負性及喚醒度的正確答案。

3.2 Enhanced 模型

我們注意到在訓練資料集 CVAT 2.0 中，有 99.8% 的句子含有 CVAW 4.0 及 CVAP 2.0 所出現的詞，且目標文本 (測試集) 可能存在與訓練資料分佈不同的情況，因此若將情感詞作為額外的資訊，應可以提升預測結果的準確性。

為了讓模型不只是學到如何為句子評分，也同時能學到句子中出現的詞的情感分數，我們基於 Baseline 系統，額外地增加了以詞的正負性、喚醒度為迴歸目標的增強式模型，模型架構如圖2所示。更明確地，我們首先抽取句子中最多 k 個長度小於 m 的詞或片語 $\{w_1, \dots, w_k\}$ ，其方法如演算法1所示，這些詞或片語會以 SEP token 隔開，串接在原始的句子詞嵌入序列之後 $\hat{E} = \{e_{cls}, e_1, \dots, e_{|T|}, e_{sep}, e_{w_{11}}, e_{w_{12}}, \dots, e_{w_{km}}, e_{sep}\}$ ，其中 $e_{w_{ij}}$ 表示為第 i 個詞的第 j 個 Token，由於每個詞經過分詞器可能表示為多個 Token，我們將每個詞所對應到的 Transformer 輸出取平均，再經過共享參數的前饋神經網路，得到

資料集	文本	正負性	喚醒度
CVAW 4.0	開心	7.2	6.6
	愉快	7	4.8
	溫柔	6.6	4.2
	廢物	2.9	6.0
	於事無補	1.7	3.4
CVAP 2.0	極度失望	1.630	7.244
	非常失望	2.000	7.00
	略微失望	2.313	5.870
CVAT 2.0	(Book) 故事情節的發展，我心中甜蜜，酸楚，失落，幸福	5.125	4.571
	(Hotel) 服務好，前臺接待員挺熱情	6.667	4.111
	(News) 新竹縣一名 55 歲男子，曾持刀刺傷父親也曾放火燒屋，他這個月 6 日不滿父親拒喝湯，竟然把熱湯潑在父親身上，再動手痛毆父親，造成老父 2 根肋骨斷裂住院。	1.875	6.875
測試集	不知道為什麼要趕課，一直跳投影片，都不知道重點在哪	1.75	7.08
	今天教了許多以前沒有學過的東西，所以上起課來很新鮮	6.8	5.2
	覺得課程進度有點快，內容難以消化	3.0	4.0

表 1: 各個資料集的範例資料。註：在比賽結束後才得到測試集的標記。

演算法 1

```

procedure EXTRACTEMOTION( $T$ )
     $S \leftarrow \{\}$ 
     $P \leftarrow \{\dots\}$   $\triangleright$  依長度排列之片語和詞
     $i \leftarrow 0$ 
    while  $i < \text{len}(T)$  and  $S.\text{size}() < k$  do
         $j \leftarrow \min(m, \text{len}(T) - i)$ 
        while  $j > 0$  do
             $s \leftarrow T[i : i + j]$ 
            if  $s \in P$  then
                 $S \leftarrow S \cup \{s\}$ 
                 $i \leftarrow i + j - 1$ 
                break
            end if
             $j \leftarrow j - 1$ 
        end while
         $i \leftarrow i + 1$ 
    end while
    return  $S$ 
end procedure
    
```

該詞的預測結果：

$$\begin{aligned}
 H &= \text{BERT}(\hat{E}) \\
 h_{w_i} &= \text{Mean}([h_{w_{i1}}, \dots, h_{w_{im}}]) \\
 [y_{w_i}^v, y_{w_i}^a] &= \text{FFN}(h_{w_i}), i \in \{1, \dots, k\} \\
 \mathcal{L} &= (1 - \alpha)\mathcal{L}_t + \alpha \sum_{i=1}^k \text{MSE}([y_{w_i}^v, y_{w_i}^a], [\hat{y}_{w_i}^v, \hat{y}_{w_i}^a])
 \end{aligned}$$

最終，我們以超參數 α 調節 \mathcal{L}_t 與我們新加入

的目標函數之訓練比重。

4 實驗設置

4.1 資料前處理

將所有資料 (CVAT 2.0、CVAP 2.0 與 CVAW 4.0) 使用 Microsoft Office Word 從繁體中文轉為簡體中文，並將 CVAT 2.0 中 20% 的資料隨機切割作為發展集，剩餘 80% 資料作為訓練集，訓練集共有 2,375 句，發展集則有 594 句，Enhanced 模型使用全部的 CVAP 2.0 和 CVAW 4.0 做為額外的訓練資料，後續實驗中我們將固定此發展集，以評估不同模型之間的表現。我們也將正負性和喚醒度的標記答案從 1 至 9 分正規化為 -4 至 4 分，以利模型訓練。

4.2 模型設定

本論文中使用 Huggingface (Wolf et al., 2020) 所提供的 Transformer 架構，以哈工大訊飛聯合實驗室所釋出的預訓練 BERT (chinese-bert-wwm-ext)、RoBERTa (chinese-RoBERTa-wwm-ext) 及 MacBERT (chinese-MacBERT-base) 初始化模型參數。

Baseline 及 Enhanced 模型同樣以學習率 (Learning Rate) 10^{-4} 與批次大小 (Batch Size) 80 訓練 7 個 Epoch，以在發展集上均方根誤差最低的模型作為最優模型，我們注意到不同的隨機種子對訓練結果的影響很大，我們最終展示以三個不同的隨機種子訓練的模型成績之平均。

Enhanced 模型抽取句子中最多 6 個長度小於 6 的詞或片語 (即 $k = 6$ 且 $m = 6$)，輸入時

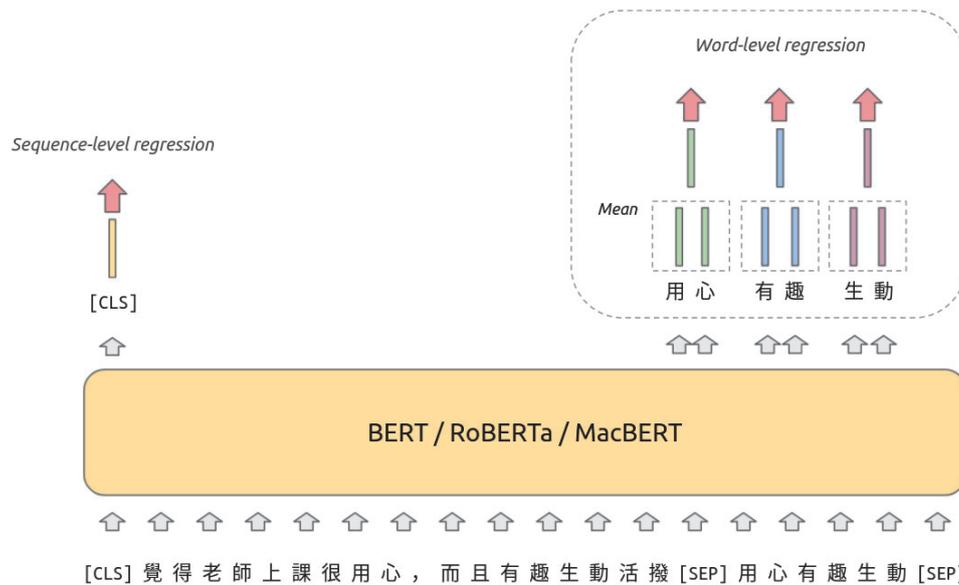


圖 2: Caption

模型	Valence _{MAE}	Valence _r	Arousal _{MAE}	Arousal _r
Baseline-BERT	0.518(0.019)	0.871(0.011)	0.694(0.010)	0.585(0.007)
Baseline-RoBERTa	0.497(0.009)	0.886(0.004)	0.691(0.001)	0.590(0.003)
Baseline-MacBERT	0.497(0.015)	0.886(0.007)	0.704(0.007)	0.577(0.009)
Enhanced-BERT	0.489(0.007)	0.890(0.003)	0.676(0.001)	0.610(0.004)
Enhanced-RoBERTa	0.485(0.003)	0.895(0.004)	0.675(0.000)	0.614(0.001)
Enhanced-MacBERT	0.480(0.003)	0.898(0.001)	0.679(0.009)	0.613(0.015)

表 2: 以三個不同的隨機種子訓練的模型在發展集上的成績平均 (標準差)。

長度不足的部份以 PAD token 補足，輸出時僅計算非 PAD token 之向量平均與損失。超參數 α 設定為 0.5，調整與傳統目標函數的比例。

4.3 集成學習

基於對實驗的觀察，我們使用兩組集成模型作為最後測試集的預測。首先，我們用不同的隨機種子切割訓練集與發展集，並訓練了各 5 個不同的 BERT、RoBERTa、MacBERT 共 15 個模型。我們將 5 個 MacBERT 的預測結果平均作為第一個測試結果；另外，在所訓練的 15 個模型中，剔除每筆資料最大與最小的極值後取平均，作為第二組測試結果。

5 實驗結果與討論

表 2 展示 Baseline 模型及 Enhanced 模型在使用不同的預訓練參數所得到的成績，我們以平均絕對誤差 (越低越好) 與皮爾森相關係數 (越高越好) 兩個指標來評估模型的表現。

首先，在所有的結果中，Enhanced 模型皆優於 Baseline 模型，代表我們所新增的以詞

輔助的迴歸目標，使用了額外的資訊，可以預期無論是正負性或喚醒度都可以帶來正向的影響。

接著我們討論不同維度之間的差異，在正負性的兩個指標中，BERT 模型在 Baseline 及 Enhanced 模型中皆略遜於另兩個模型，RoBERTa 則相同或略遜於 MacBERT。在喚醒度的部份，MacBERT 在三者之中表現最不理想，RoBERTa 有最好的成績。BERT 因為是最早期的模型，所以我們期待基於 BERT 改進的 RoBERTa 與 MacBERT 表現應該會較優，但從實驗顯示 MacBERT 沒辦法在喚醒度和正負性有相當的表現。

從結果中可以看到正負性的成績皆優於喚醒度的成績，代表喚醒度對我們所設計的模型來說是比較困難的任務。經過人工對資料集的分析，我們認為喚醒度在標記答案時較為主觀，相較於正負性來說，較難客觀評估情緒的激動程度，相似的句子可能會有相當不同的標記，即使 MacBERT 可以在 Baseline 及 Enhanced 模型的正負性中取得最好的成績，在喚醒度部份則表現較差，我們猜測原因可能與預訓練

Model	Valence _{MAE}	Valence _r	Arousal _{MAE}	Arousal _r
CYUT-run1	1.695	-0.017	1.177	0.040
CYUT-run2	1.685	0.007	1.252	-0.021
NCU-NLP-run1	0.625	0.900	0.938	0.549
NCU-NLP-run2	0.611	0.904	0.989	0.582
ntust-nlp-1-run1	0.684	0.912	0.906	0.607
ntust-nlp-1-run2	0.586	0.901	0.885	0.585
SCUDS-run1	0.953	0.694	1.054	0.375
SCUDS-run2	0.975	0.667	1.039	0.354
SoochowDS-run1	2.421	0.073	1.327	0.051
SoochowDS-run2	1.073	0.584	1.125	0.228
Enhanced-MacBERT(X5)	0.654	0.880	0.905	0.581
Enhanced(X15)	0.667	0.913	0.866	0.616

表 3: 提交的測試集結果。註: Enhanced-MacBERT(X5) 與 Enhanced(X15) 分別為 ntust-nlp-2-run1 與 ntust-nlp-2-run2

資料	正負性	喚醒度
CVAT 2.0	4.80(1.34)	4.84(1.04)
測試集	5.32(1.69)	4.51(1.32)
Ensemble-MacBERT(X5)	5.39(1.10)	4.60(0.54)
Ensemble(X15)	5.41(1.05)	4.65(0.55)

表 4: 資料集與預測結果中的所有資料的平均 (標準差)。

的方式有關, MacBERT 在預訓練時將詞彙替換為相似的詞彙以達到更好的語言模型效果, 但在情感辨識任務中, 由於標記的主觀性, 相似的詞, 例如表1中的「開心」與「愉快」, 在語意上相近, 但在喚醒度卻相差很大, 這可能使 MacBERT 無法發揮原本的優勢, 而我們所設計的 Enhanced 模型, 以詞層級的資訊讓模型學會分辨不同相似詞的情緒關係, 修正了 MacBERT 本身的這項缺點。最終提交結果時, 我們使用集成學習的方式 (請參考4.3節) 產生了兩組結果, 表3可以看出使用多種模型的效果最佳, 但在正負性的平均絕對誤差卻相對表現較差。

經過對測試資料與其標記答案的分析, 我們注意到訓練資料和測試資料之間存在明顯的差異。在 CVAT 2.0 中蒐集的內容為書本、旅館評論、汽車討論、電腦討論、新聞、政治評論等六類內容, 其中不乏較激進、狂喜、憤怒的言論; 相對地, 在測試資料中普遍為學生的課堂自評, 情感上相對訓練資料是比較溫和、集中的。表4顯示了以正確答案計算所得的平均值與標準差, 我們將 Ensemble-MacBERT(X5) 與 Ensemble(X15) 兩個模型對於測試集的預測結果, 同樣計算平均值與標準差, 並列於表4中。透過上述統計, 我們發現兩個模型所預測的結果中, 正負性平均較 CVAT 2.0 大, 反應了學生自評中較多正向

句的觀察。而喚醒度標準差明顯較 CVAT 2.0 小很多, 代表模型傾向給予測試資料更集中的評分, 符合我們前述學生自評情緒較溫和的預期。

當我們進一步地以人工方式檢視資料, 發現最終釋出的測試集標記答案與訓練資料有很大的不同, 兩個資料集分開來看可以看出各個資料集的分佈類似, 但兩者標記的尺度存在不小的差異。舉例來說, 如表1所示, 訓練資料集中「新竹縣一名 55 歲男子, 曾持刀刺傷父親也曾放火燒屋, 他這個月 6 日不滿父親拒喝湯, 竟然把熱湯潑在父親身上, 再動手痛毆父親, 造成老父 2 根肋骨斷裂住院。」與測試資料集中「不知道為什麼要趕課, 一直跳投影片, 都不知道重點在哪」兩者正負性和喚醒度接近, 但以訓練資料的角度來看, 測試資料的標記明顯較為誇大了正負性與喚醒度。

最後, 我們比較表2與表3, 可以發現發展集及測試集平均絕對誤差有明顯的退步, 但皮爾森相關係數卻能保持接近的成績, 我們認為原因就是因為標記尺度不同, 所以模型雖然能依訓練資料的尺度評分, 但平均絕對誤差則會大幅退步, 不過相關性仍能保有一定的預測水準。

6 結論

我們針對維度式的情感辨識任務「ROCLING 2021 共享任務：教學文本之維度式情感辨識」提出了基於預訓練語言模型 BERT 架構的迴歸模型，測試結果顯示我們的模型可以達到有競爭力的表現，我們也透過一連串的分析，證明我們所提出的模型設計能有比較好的表現，也注意到仍有許多可以改進的問題留待未來繼續研究。

7 Acknowledgment

This work was supported by the Ministry of Science and Technology of Taiwan under Grant MOST 110-2636-E-011-003 (Young Scholar Fellowship Program). We thank to National Center for High-performance Computing (NCHC) for providing computational and storage resources.

References

- Rafael A. Calvo and Sunghwan Mac Kim. 2013. Emotions in text: dimensional and categorical models. *Computational Intelligence*, 29(3):527–543.
- Yiming Cui, Wanxiang Che, Ting Liu, Bing Qin, Shijin Wang, and Guoping Hu. 2020. Revisiting pre-trained models for Chinese natural language processing. In *Findings of the Association for Computational Linguistics: EMNLP 2020*, pages 657–668, Online. Association for Computational Linguistics.
- Jacob Devlin, 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, Volume 1 (Long and Short Papers)*, pages 4171–4186, Minneapolis, Minnesota. Association for Computational Linguistics.
- Steven Du and Xi Zhang. 2016. Aicyber’s system for ialp 2016 shared task: Character-enhanced word vectors and boosted neural networks. In *2016 International Conference on Asian Language Processing (IALP)*, pages 161–163.
- Pranav Goel, Devang Kulshreshtha, Prayas Jain, and Kaushal Kumar Shukla. 2017. Prayas at EmoInt 2017: An ensemble of deep neural architectures for emotion intensity prediction in tweets. In *Proceedings of the 8th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*, pages 58–65, Copenhagen, Denmark. Association for Computational Linguistics.
- Sunghwan Mac Kim, Alessandro Valitutti, and Rafael A. Calvo. 2010. Evaluation of unsupervised emotion models to textual affect recognition. In *Proceedings of the NAACL HLT 2010 Workshop on Computational Approaches to Analysis and Generation of Emotion in Text*, pages 62–70, Los Angeles, CA. Association for Computational Linguistics.
- Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. 2019. Roberta: A robustly optimized bert pre-training approach. Cite arxiv:1907.11692.
- Nikolaos Malandrakis, Alexandros Potamianos, Elias Iosif, and Shrikanth Narayanan. 2013. Distributional semantic models for affective text analysis. *IEEE Transactions on Audio, Speech, and Language Processing*, 21(11):2379–2392.
- Georgios Paltoglou, Mathias Theunis, Arvid Kappas, and Mike Thelwall. 2013. Predicting emotional responses to long informal text. *IEEE Transactions on Affective Computing*, 4(1):106–115.
- James A. Russell. 1980. *A circumplex model of affect*. 1161, Journal of Personality and Social Psychology, 39(6).
- Jin Wang, Liang-Chih Yu, K. Robert Lai, and Xuejie Zhang. 2016. Community-based weighted graph model for valence-arousal prediction of affective words. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 24(11):1957–1968.
- Jin Wang, Liang-Chih Yu, K. Robert Lai, and Xuejie Zhang. 2020. Tree-structured regional cnn-lstm model for dimensional sentiment analysis. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 28:581–591.
- Wen-Li Wei, Chung-Hsien Wu, and Jen-Chun Lin. 2011. A regression approach to affective rating of chinese words from anew. In *Affective Computing and Intelligent Interaction*, pages 121–131, Berlin, Heidelberg. Springer Berlin Heidelberg.
- Thomas Wolf, Lysandre Debut, Victor Sanh, Julien Chaumond, Clement Delangue, Anthony Moi, Pierric Cistac, Tim Rault, Rémi Louf, Morgan Funtowicz, Joe Davison, Sam Shleifer, Patrick von Platen, Clara Ma, Yacine Jernite, Julien Plu, Canwen Xu, Teven Le Scao, Sylvain Gugger, Mariama Drame, Quentin Lhoest, and Alexander M. Rush. 2020. Transformers: State-of-the-art natural language processing. In *Proceedings of the 2020 Conference on Empirical*

Methods in Natural Language Processing: System Demonstrations, pages 38–45, Online. Association for Computational Linguistics.

Chuhan Wu, Fangzhao Wu, Yongfeng Huang, Sixing Wu, and Zhigang Yuan. 2017. THU_NGN at IJCNLP-2017 task 2: Dimensional sentiment analysis for Chinese phrases with deep LSTM. In *Proceedings of the IJCNLP 2017, Shared Tasks*, pages 47–52, Taipei, Taiwan. Asian Federation of Natural Language Processing.

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 the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 540–545, San Diego, California. Association for Computational Linguistics.

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 the IJCNLP 2017, Shared Tasks*, pages 9–16, Taipei, Taiwan. Asian Federation of Natural Language Processing.

Liang-Chih Yu, Jin Wang, K. Robert Lai, and Xuejie Zhang. 2020. Pipelined neural networks for phrase-level sentiment intensity prediction. *IEEE Transactions on Affective Computing*, 11(3):447–458.

Suyang Zhu, Shoushan Li, and Guodong Zhou. 2019. Adversarial attention modeling for multi-dimensional emotion regression. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 471–480, Florence, Italy. Association for Computational Linguistics.