

智慧化企業整合_Project3
SR-NET 與 ESRGAN 生成圖像之比較

指導教授:邱銘傳

110034561 洪聖博

2022/1/7

一、背景介紹

(一)、背景說明

近幾年來網際網路迅速發展，除了能夠為商業服務增值，亦能提供景點的評論回饋，以供消費者參考與選擇。而消費者除了能透過如 Google Maps 等線上地圖服務提供者了解景點的基本資訊，諸如景點名稱、營業時間等，也能經由其獲得廣大使用者的評論。然而評論內容數大且繁雜，使用者難以一一檢閱以確保餐廳是否符合期望。因此本研究目的即在為廣大使用者提供景點評論的深度見解，以期能加速使用者尋找景點的過程。現今網路蓬勃發展，許多不具備資訊背景的使用者也能夠輕鬆且快速的上網。據此，許多面相資訊交流與傳遞的平台有如雨後春筍般興起，Google Maps 即為一例。Google 扶龐大的使用者基數，為旗下的網路地圖服務帶來了許多增值服務，如人潮預測與景點評論等，漸漸成為人們尋找景點時的重要依據之一。

(二)、5W1H 分析

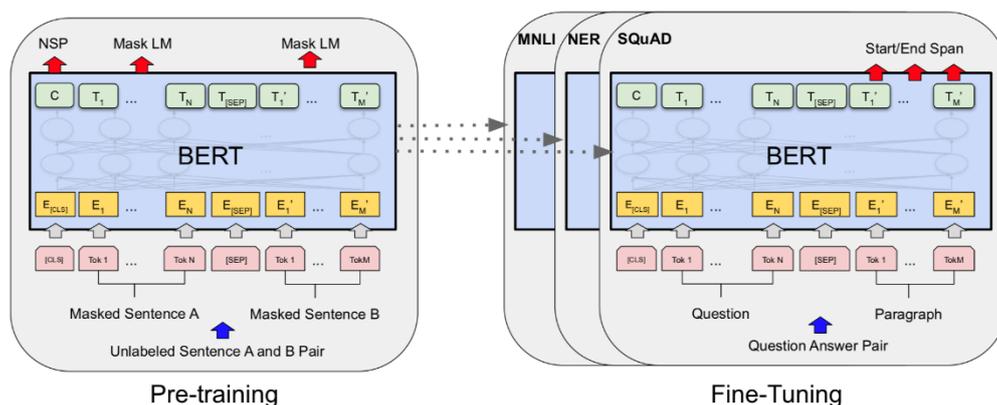
Why	判斷留言為正向或負向。
What	利用 Google 評論尋找景點的觀光客。
Where	提供建使用者想要尋找景點的時候議。
When	能在 Google 景點評論區進行判斷。
Who	避免字數過多的評論或過多的評論。
How	透過文本情感分析與文本分類等技術輔以網站開發，為觀光客提供評價。

二、方法介紹

(一)、BERT

傳統的語言模型 (Language Model) 往往需要大量的標註資料，然而這是不切實際且需要大量經費的。據此，Google 提出 BERT: Bidirectional

Encoder Representations from Transformers 模型，其透過大量無標記的資料自我學習文本 上下關係。不像過往的 word2vec 等模型無法獲取文本上下文的關係，導致字詞 的真正意義難以被萃取出來，BERT 因為採用 Bidirectional Encoder 而得以獲取 單一字詞前後相連的字詞。事實上，BERT 即為 Transformer 中 Encoder 的部分，所以許多研究將 BERT 作為新一代字詞 embedding 的生成器。



(二)、情感分析

本研究之研究架構，可分為(1)資料輸入、(2)資料前處理、(3)模型建構、(4)進行訓練。本研究目標為不透過完整標註的資料集即成功訓練情感分析模型。因此資料選用 Google Maps 上的評論，並將每篇評論的文字斷句成短句後，給予其 weakly label 即為該則評論的評分，接著將訓練資料提供給 BERT 以得到 sentence embedding，最後將 embedding 串接 Fully Connected Neural Network 進行評分預測。下圖為評論轉換為訓練資料的過程。可以發現並非所短句皆具有合理的 rating 標籤，然而可將其視為 noise，因為應該有更多其他「不好玩」的句子獲得低分，而得以正確訓練。

以下圖說明情感分析模型架構，左邊為輸入資料，而右邊為預測值。



三、個案研究

(一)、資料集介紹

本次 Project 所使用的 dataset 是透過 Python 網路爬蟲抓取 Google 景點評論，原始評論有 4500 則，而每一則評論內容都不相同、評分的高低也不相同。

以下為一些評論的範例：

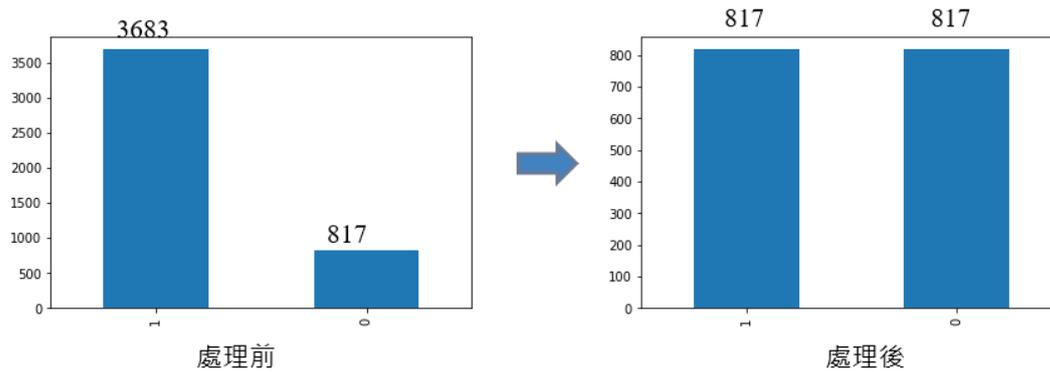
Peter Pan	5	2012/9/8帶著住在德國不萊梅附近的交換學生，遊墾丁：浮潛，香蕉船，水上摩托車，潘揚都是第一次體驗。小子用讚讚
Ying Lin	5	導覽人員的細心解說讓我們學到很多關於墾丁地質的屬性，也才知原來大尖山與帆船石都是外來的而不是當地就存在的。
Hsin Yang	5	沒人都美口
陳世杰	4	在墾丁炎炎夏日，國家森林是避暑的好地方，有機會可以看見獼猴，梅花鹿
詹貝貝	5	巴士海峽 晨起落山風 關山觀日落 後壁湖紅樓茅 窗綠佳落水 滴水南仁湖 圓綠一線天 牛郎會織女 綠灣鴛鴦水 幽會墾丁公
Yimin Che	5	去了很多地方 墾丁就是帶著懶懶慢慢的南洋放鬆感
jimmy	4	騎屏鵝公路 心情舒暢。
阮玫心	4	走走散散心，曬曬太陽，看看海浪，玩玩海水，.....
Rose Hu	5	南岬草原縱走，值得一遊。很多秘境等待探訪。還碰到了復育的梅花鹿群、還有豐富的植物生態！鳥類...值得一遊
謝美蘭	5	臺灣第一個國家公園
邱軒	4	6年多沒來墾丁變滿多的，希望不要過度開發破壞了生態大自然美觀
烏人法爾	5	臺灣成立的第一個國家公園，也是唯一一個裡頭有進行大量商業活動的國家公園。

(二)、資料前處理

本研究之資料前處理包含下列步驟。

1. 忽略所有評論文字為空的評論
2. 忽略所有由 Google 翻譯而來的評論
3. 移除所有表情符號
4. 處理句子前與後的空白與換行字元
5. 透過「。；！？\n」等符號進行斷句
6. 移除過短 (≤ 2 字) 或過長 (≥ 20 字) 的短句
7. 設定短句的情感評分標籤為該則評論的評分
8. 由於各評分的短句分布不均，所以進行資料擴增，將資料較少的評分數 (如 1 星至 3 星) 隨機抽樣複製直至各分數評論數平均

資料總共為 4500 筆，但由於正面評論遠遠高於負面評論，資料量差距過大，可能會導致訓練失敗。所以從原始資料中的 817 則負面評論設為基準，隨機抽取 817 則正面評論，使正負評論資料平衡。



(三)、BERT 模型建立

(一)、透過 BERT 生成 embedding

本研究採用 Pretrained BERT 的 chinese_L-12_H-768_A-12 版本，每個 embedding 寬度為 768，而 word embedding 個數和句子長度相同。

(二)、Pooling

因為 BERT 生成出來的 word embedding 個數和短句長度相同，為了將此資料輸入進接下來的 NN，因此本研究將這些 embedding 進行 pooling，並使用官方建議的 REDUCE_MEAN Pooling 策略。

(三)、Fully Connected Neural Network

本研究將短句的 embedding 輸入一個 Fully Connected Neural Network 以期能夠預測短句的評分，我們先以下列模型為 baseline，接下來幾節會按此為基礎進行超參數優化。此 model 的 dropout rate 為 0.3，每層的 activation function 採用 relu，而最後一層的單結點當作預測值。訓練時以 mean squared error 作為 loss function，並且以 Adam 作為優化器。

四、參數優化

	Level 1	Level 2
Activation Function	<u>Relu</u>	Sigmoid
Dropout	0.2	0.4
Pooling	O	X

	Activation Function	Dropout	Pooling	Accuracy
1	<u>Relu</u>	0.2	O	0.9572
2	<u>Relu</u>	0.2	X	0.9174
3	<u>Relu</u>	0.4	O	0.9543
4	<u>Relu</u>	0.4	X	0.9633
5	Sigmoid	0.2	O	0.9572
6	Sigmoid	0.2	X	0.9572
7	Sigmoid	0.4	O	0.9205
8	Sigmoid	0.4	X	0.9694

五、研究結果

本研究總共建構了三大部分，情感分析模型、關鍵指標分數計算模型與平台網站，將從 Google Maps 上獲取的大筆評論數據提供予情感分析模型進行訓練與推練，並藉由上述模型計算各店家的關鍵指標分數，最後將關鍵指標分數呈現於平台網站中，供所有想要快速找尋心儀餐廳的使用者操作。其中情感分析模型以 NTUSD 情緒字詞辭典進行驗證，並達到可實用的程度，並透過下游應用發揮其效用。

六、結論

(一)、成果貢獻

本研究嘗試以未精準標記過的繁體中文文本情緒資料，即成功訓練文本情緒分析模型，輔以 Pretrained BERT 生成 embedding，並試圖驗證其準確率，最終取得尚可的成績。另外也提出餐廳關鍵指標分數的計算方式，並以實際應用場域驗證之。

(二)、研究限制

本研究最大的限制在於目前繁體中文界尚無完整的文本情感分析 Benchmark，雖有英文或簡體中文的基準可供使用，然而因為已經定義了本研究主題的領域為繁體中文的景點評論，使用諸如 ChnSentiCorp 等簡體中文 benchmark 可能無法正確評價此模型的表現。

(三)、未來方向

可以透過 Named Entity Recognition 等方式達成更好的文本分類模型，並以比 BERT 更好的 Language Model 生成更 contextual informative 的 embedding，又或是可以不要以短句進行情感分析，而是以整句進行分析。

(四)、應用

本研究已透過平台網站驗證其應用能力，另外此情緒分析模型亦可應用於廣大的繁體中文自然語言處理場景，諸如線上即時客戶服務、聊天機器人、電子郵件自動回覆機器人等。

七、參考文獻

- <https://medium.com/@ethan.chen927/python%E6%A9%9F%E5%99%A8%E5%AD%B8%E7%BF%92-google%E6%88%91%E7%9A%84%E5%95%86%E5%AE%B6%E6%AD%A3%E8%B2%A0%E8%A9%95%E7%8C%9C%E6%B8%AC-%E9%81%8B%E7%94%A8bert-model-%E7%90%86%E8%A7%A3%E4%B8%8A%E4%B8%8B%E6%96%87%E7%9A%84%E8%AA%9E%E8%A8%80%E6%A8%A1%E5%9E%8B-with-colab-pro-gpu-ec8ef8cb8a25>
- <https://zh-tw.coderbridge.com/series/2ec9cf0af3f74ed99371952f4849ae33/posts/e2e46480d185435680609a40997644be>
- <https://aistudio.baidu.com/aistudio/projectdetail/2048282>
- <https://zhuanlan.zhihu.com/p/166254682>