

在 Ppt 中的第 10 頁& 11 頁 feature extension 的部分:

使用 bert 模型的理由:

1. 為了取得更加客觀的推薦依據，並省去讀者看評論的時間。因為如果單純以 rating 數值去做推薦，在原數據中只會分成 1~5 的整數指標，如果拿這個 label 去做 top10 高分的推薦，將有很高的機率超過 10 個同分的情況，此時的推薦效果就不夠精準(ex:從一堆結果 random 取樣)，因此若多考慮讀者評論資訊轉換後的情感指數，重新以 $ex : 0.7(\text{情感指數}) + 0.3(\text{原始 rating})$ ，將減少同分狀況，取得更可靠的推薦結果。
2. 而如參考論文(Fault in your stars: An Analysis of Android App Reviews)中所提及，可能有各種原因導致 rating 和評論不一致，用龐大的數據訓練模型，將能盡可能消弭因不一致導致的偏差。

用清理出來的 redundant 資料做 fine tuning 理由:

1. 在第 11 頁的部分增加對 bert 情感分析模型這部分做解釋，我們的想法是，開源的 bert 模型無法針對“書籍”評論來判別情感指數，例如：“這本書內容很生動”、“情節令我沉浸其中”，bert 無法理解該部分的情感指數應該為何，因此使用 redundant 資料的 rating 當作 ground truth(一樣將情感指數劃分成 5 個等級)，去訓練 bert 針對該評論做情感指數預測，因數據量夠大，訓練後的 bert 將能有效預測出該類句子情感指數。