**三、資料集來源**

資料集蒐集目的與預期結果在於透過多維度的公開資料微調及訓練模型，提升模型在跨資料集和其他情感分析的泛化能力。此外，透過二分類的情感分析資料集提升模型對於強烈情緒語句的文本分類能力。

* SST-5 Fine-grained classification

本資料集為 SST-5 Fine-grained classification with Rotten Tomatoes Movie Reviews 電影評論情感分類集，情感標籤包含5種情感等級:

|  |  |
| --- | --- |
| Label | Comments |
| 0 | Negative |
| 1 | Somewhat negative |
| 2 | Neutral |
| 3 | Somewhat positive |
| 4 | Positive |

使用該資料集訓練不僅提供多維度情感分析，細粒度的情感標籤亦在訓練過中降低模型可能衍伸之情感等級模糊問題。

* SST2: Stanford - IMDb Dataset of 50K Movie Reviews

本資料集包含五萬餘則 IMDb 電影評論情感分類集，情感標籤有正

面及負面2極情感維度。訓練集及測試集的文本序列數量分別各為

25,000餘則。