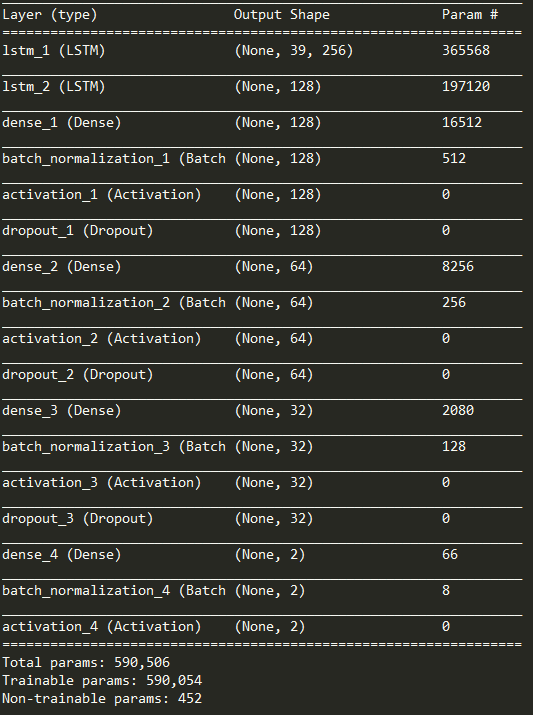
學號：R06942082 系級： 電信碩一 姓名：黃釋平

1. (1%) 請說明你實作的 RNN model，其模型架構、訓練過程和準確率為何？  
   (Collaborators: 鄭立晟\_R05942148、姚嘉昇\_R06922002)

答：

經過多次的嘗試，依然達不到標準，所以參考同學的model，不同的是我對資料做了兩次normalize，x=(x-mean)/var(x)，一次是word自己的vector (dim=100)做，第二次是進model之前對所有vector做。

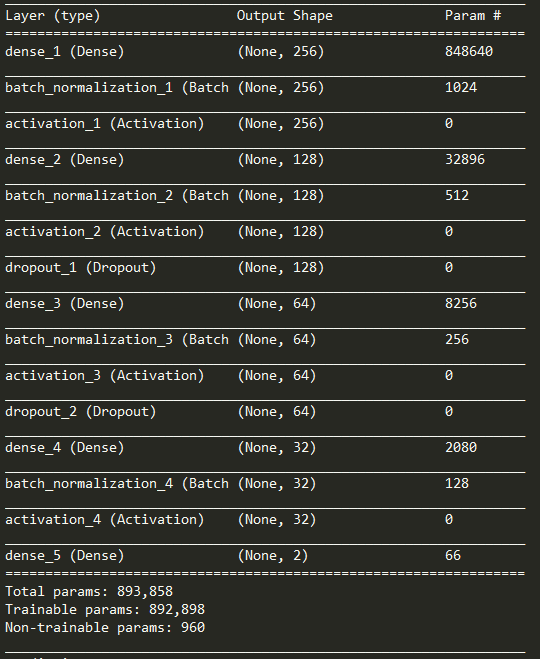
神奇的是，我在這樣的架構下試著多加Dense(16)、Dense(8)、Dense(4)三層，使其遞減到2，結果上傳卻降低，且應該沒有overfitting。

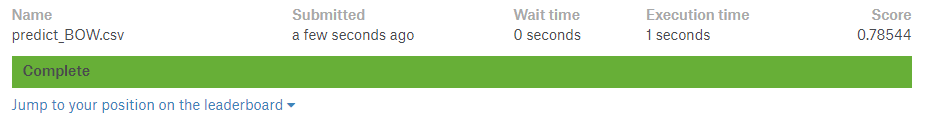


Best result: epoch=4, validation:0.820, training:0.834, kaggle:0.822

1. (1%) 請說明你實作的 BOW model，其模型架構、訓練過程和準確率為何？  
    (Collaborators: )

答：





使用corpora來做BOW，這是目前調到最好的參數，**validation=0.786**，**上傳結果0.785**，差異不大，但略比LSTM差，可能是**沒考慮到字的順序**。

1. (1%) 請比較bag of word與RNN兩種不同model對於"today is a good day, but it is hot"與"today is hot, but it is a good day"這兩句的情緒分數，並討論造成差異的原因。  
    (Collaborators: )

答：

|  |  |  |
| --- | --- | --- |
| Kaggle Test | LSTM | bag of word |
| "today is a good day, but it is hot" | **[ 0.267 0.732]** | **[ 0.435 0.533]** |
| "today is hot, but it is a good day" | **[0.0003 0.9997]** | **[ 0.465 0.513]** |

RNN很確定判斷為好，兩句確定性差異較大;  
BOW雖然也都判斷為好，但只接近五成，且兩句結果差不多  
可見**字的順序確實影響到判斷結果**。

1. (1%) 請比較"有無"包含標點符號兩種不同tokenize的方式，並討論兩者對準確率的影響。  
    (Collaborators: )

答：

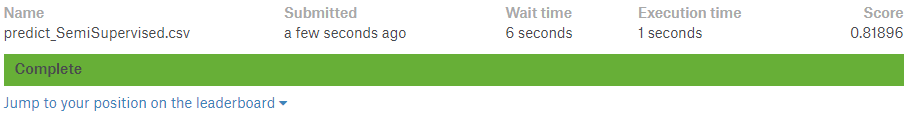
|  |  |  |
| --- | --- | --- |
| Kaggle Test | LSTM | bag of word |
| 保留標點符號 | 0.82211 | 0.78544 |
| 去標點符號 | 0.81038 | 0.78158 |

去除標點符號後平均表現皆變差，也許**符號對情緒的判斷是有參考價值的**。

1. (1%) 請描述在你的semi-supervised方法是如何標記label，並比較有無semi-surpervised training對準確率的影響。  
    (Collaborators: )

答：

|  |  |  |
| --- | --- | --- |
| Kaggle Test | Non | Semi-Supervised |
| LSTM | 0.82211 | 0.81896 |



做了Semi-Supervised後，validation accuracy飆高到0.94，training accuracy則是0.86，都有提高，但上傳結果卻稍微降低一點。