第十組

M11115020 吳宇祥 M11015084 魏向晨

M11115079 張家維 M11115099 郭建鴻

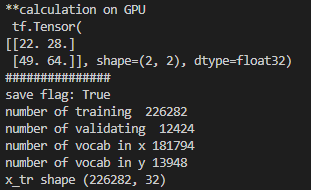
M11115073 張晉嘉 M11115103 施昱民

實作心得:

先安裝tensorflow==2.5.1

1. Training 訓練資料集 -d後面可以填wiki\_tfidf或wiki\_sent 此步驟為用維基的資料集去訓練標籤

python train\_tf.py -m 'bigru\_bahdanau\_attention'  -d 'wiki\_tfidf'

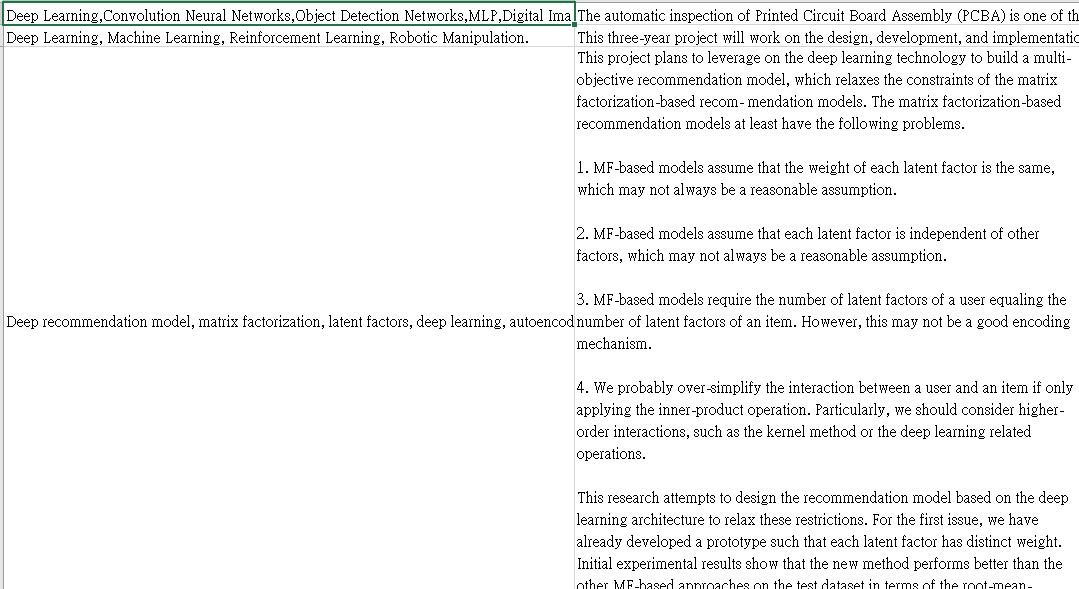


訓練資訊會印在terminal

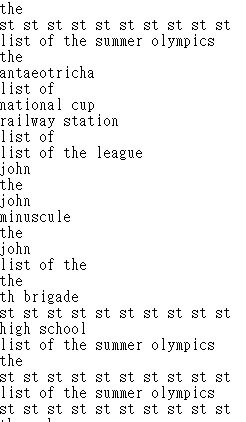
我們除了論文資料及外還有嘗試自己找的資料集

從政府的計畫內容系統GRB 下載資料 並將資料改成論文的input

指標分數是還蠻高的但是實際看預測標籤好像都沒有預測正確的，因此我們推斷這個model可能還是有domain knowledge的限制，我們跑了另外一個從維基抓下來的資料集，指標分數仍然不錯，在觀察預測的label就有比較明顯的抓出一些關鍵字了。因此我們推測是因為訓練資料是用wiki來作訓練，導致一些專門領域的資訊還是沒有學習到，但是對於大部分常見的議題這個模型的預測標籤就有呈現出來~

**政府計畫資料**  


實際預測結果:

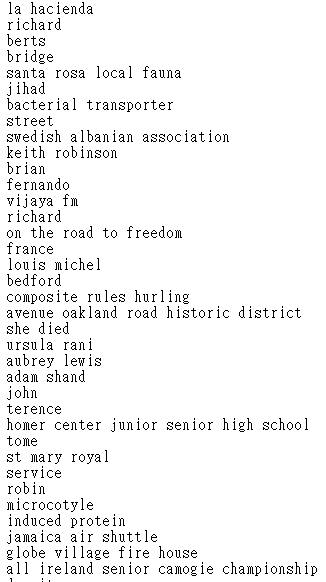


可以發現其實都猜得不太好都是用沒有意義的字來當答案，我們推測可能跟訓練資料集沒有考慮到domain knowledge的關係。

**維基新資料**:這是額外從維基百科的官方載下來的資料



模型猜測結果:



可以觀察到出現的字比較有意義了，這樣好像驗證了我們的推測，但是作者沒有多加說明，之後可以再深入探討。