

DLHLP HW 4 - 3

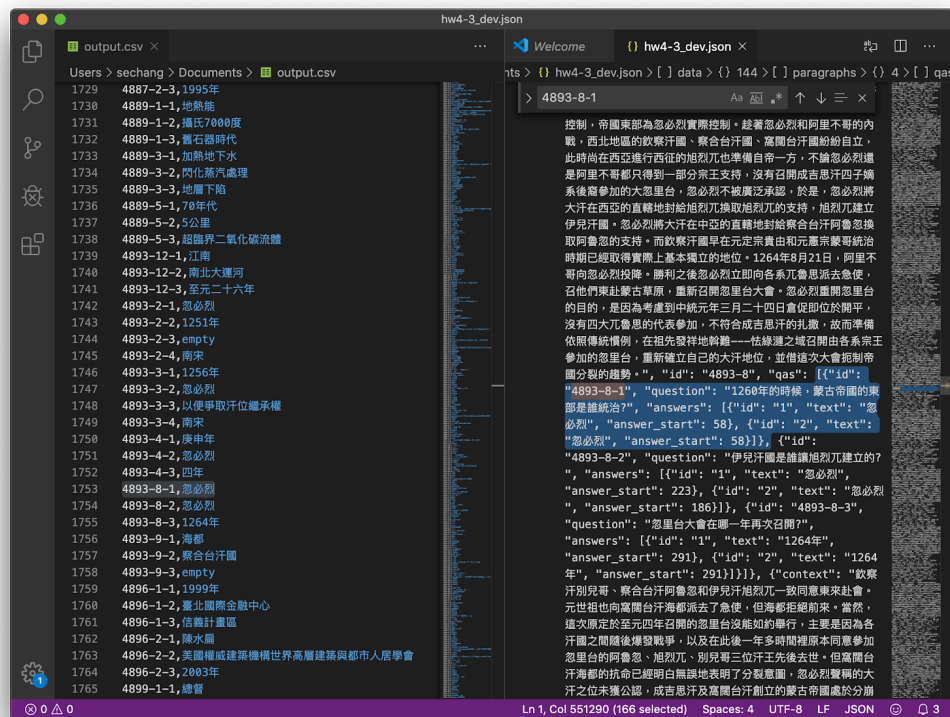
組長 Githud ID: shuoenchang

學號:R08922A02 系級: AI所碩一 姓名:張碩恩

學號:R08922130 系級: 資工所碩一 姓名:丁杰

1. (2%) Please give some examples predicted correctly and incorrectly respectively. At least one for each case is required. Screenshot recommended.

Correctly : 4893-8-1



Incorrectly : 1151-9-2、6075-2-3

```
Users > sechang > Documents > output.csv
47 1151-12-1,主裁判
48 1151-12-2,球童
49 1151-12-3,empty
50 1151-16-1,網球
51 1151-16-2,國家
52 1151-16-3,發球技術
53 1151-19-1,半西方式擺拍
54 1151-19-2,半西方式擺拍
55 1151-19-3,半西方式擺拍
56 1151-2-1,14世紀
57 1151-2-2,網
58 1151-2-3,華爾特·科洛普頓·溫菲爾德
59 1151-24-1,網球
60 1151-24-2,截擊
61 1151-24-3,制勝球
62 1151-4-1,查爾斯·派爾
63 1151-4-2,商業的壓力
64 1151-4-3,非塞利博物館
65 1151-6-1,6.40米
66 1151-6-2,紅土
67 1151-6-3,紅土
68 1151-7-1,一局比賽結束之後
69 1151-7-2,球手做出用球拍擊球的動作但並未擊打到球
70 1151-7-3,為了防止故意讓球的走向
71 1151-9-1,五
72 1151-9-2,零
73 1151-9-3,失去一分
74 1152-12-1,四年
75 1152-12-2,1989年
76 1152-12-3,2014年
77 1152-15-1,8.90%
78 1152-15-2,新北市
79 1152-15-3,汐止區
80 1152-18-1,新北市
81 1152-18-2,新板特區
82 1152-18-3,規劃中
83 1152-20-1,臺北市公車聯營管理中心
```

```
Users > sechang > Documents > output.csv
2641 6048-8-2,普選權的保護者
2642 6048-8-3,為了鞏固自己的地位
2643 6049-1-1,中華人民共和國
2644 6049-1-2,2000年
2645 6049-1-3,戰國時期
2646 6049-2-1,亞熱帶濕潤氣候
2647 6049-2-2,5月
2648 6049-2-3,道數
2649 6055-1-1,加拿大
2650 6055-1-2,曼尼托巴省
2651 6055-1-3,曼尼托巴法案
2652 6055-2-1,溫尼伯
2653 6055-2-2,溫尼伯湖
2654 6055-2-3,邱吉爾港
2655 6055-6-1,冰河時期
2656 6055-6-2,製造箭頭
2657 6055-6-3,1611年
2658 6057-2-1,蘇維埃俄國政府
2659 6057-2-2,1987年
2660 6057-2-3,建造一個人工堤壩將東西兩湖完全分離
2661 6062-1-1,航空運輸
2662 6062-1-2,空軍聯運
2663 6062-1-3,機場聯絡軌道系統
2664 6065-1-1,2395平方公里
2665 6065-1-2,莫倫斯
2666 6065-1-3,南亞熱帶
2667 6075-2-1,二百米級,兩者近距離中都可以
2668 6075-2-2,衝鋒槍
2669 6075-2-3,卡賓槍
2670 6075-3-1,可以
2671 6075-3-2,中國大陸
2672 6075-3-3,56式自動步槍
2673 6075-4-1,衝鋒槍的設計需求最早來自於第一次世界大戰當中的
2674 6075-4-2,德國
2675 6075-4-3,在實行突擊群戰術的步兵上配備了大量衝鋒槍
2676 6075-6-1,德國
2677 6075-6-2,然而,當時衝鋒槍於人們的印象還是一種結構簡單、
```

2. (3%) Which hyperparameter(s) should be modified in order to reach better performance? (e.g. learning_rate, batch_size, warmup_steps, layer_norm_eps, attention_probs_dropout_prob)

根據 bert 的官方 github (<https://github.com/google-research/bert>)，有提出一些在 finetune 時可以參考的 hyperparameter，其中在 issues 裡面看到很多的疑問都是有關 batch size 的，所以我自己的實驗主要是想研究 batch size 的影響。本次 GPU 使用 GTX 1070，memory 只有 8G，所以基本上 batch size 沒辦法放太大，以 max seq length=384 來說，最大只能 7。不過如果在其他參數不變的條件下，把 batch size 往下調，可以很明顯看出來 f1 score 往下掉了不少。我自己的猜測是因為 batch size 越小，每筆資料內含的 noise 無法被消除，在 batch size 大的狀況下，noise 比較容易有機會被去除。

另外一個蠻有趣的實驗結果是，比較 best 和 setting2，兩個模型剛好有一樣的表現，可以知道加大 max seq length 其實對於模型來說會有比較好的成果（其實也很好理解原因，因為被丟棄的句子變少了）。另外在 learning rate，這部分就比較沒有一定的規律了，不同的參數組合上可能會有不同的最佳值，但造成的影響並沒有 batch size 那麼明顯。

最後是 pretrain model 的選擇，使用 wwm 的方式搭配 roberta 所訓練出來的 model，比最原始的 pretrain model 有更好的表現。

	best	setting1	setting2	setting3	setting4	setting5
model name	chinese-roberta-wwm-ext	bert-base-chinese	chinese-roberta-wwm-ext	chinese-roberta-wwm-ext	chinese-roberta-wwm-ext	chinese-roberta-wwm-ext
learning rate	3e-5	3e-5	3e-5	3e-5	5e-5	3e-5
max seq length	384	384	512	384	512	384
batch size	7	7	3	3	3	1
f1 score	0.77605	0.76265	0.77605	0.77090	0.77434	0.76632