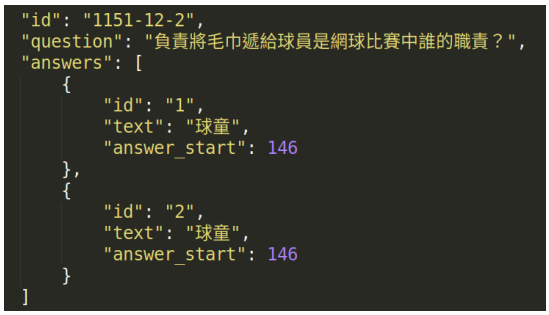
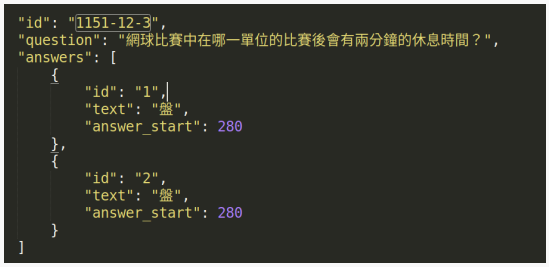
DLHLP HW4-3 Report

組長 Githud ID: lou-tun-chieh

學號：b05902111 系級： 資工四 姓名：婁敦傑

學號：b05902010 系級： 資工四 姓名：張頌平

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

	groundtruth	predict
correct		1151-12-2 球童
incorrect		1151-12-3 觀眾席

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)。

	learning rate = 1e-5	learning rate = 1e-4
kaggle public	0.75257	0.71764
val_f1	0.8099	0.7673

	attention_probs_dropout_prob = 0.1	attention_probs_dropout_prob =0.3
kaggle public	0.75257	0.75807
val_f1	0.8099	0.813

	warmup_steps = 10	warmup_steps = 10000
kaggle public	0.75406	0.75257
val_f1	0.8147	0.8099

上述的實驗結果都是在 finetune bert-base-chinese pretrained model 的結果，我們有調整以上的參數，明顯有差的是 learning rate，在這次 QA 的 task，learning rate 太高會造成結果不好，其餘的參數我們是覺得沒有明顯的上升和降低，我們其實還有調整 max_seq_length 從 384 調高到 512在 Roberta pretrained model ，但結果也是變差，我們最後用 https://github.com/ymcui/Chinese-BERT-wwm 內的 **RoBERTa-wwm-ext-large** pretrained model ，才有明顯的進步，kaggle public 從 0.76 上升到 0.80。