用 BERT 进行情感分析的过程的报告

---- 以携程网的点评信息为例

李明聪

摘要：本文用携程网的酒店点评信息作为数据集，使用 BERT 进行了文本情感倾向的预测。数据集共 7766 个样本，测试集共 5435 个样本，模型在 GPU 上训练耗时 4 分 55 秒，模型预测的 f1 score 为 0.93。

一、数据集概述

1. 数据来源：携程网

2. 分布情况

总评论数	正向	负向
7766	5322	2444

3. 字段说明

字段	说明
label	1 表示正向评论，0 表示负向评论
review	评论内容

4. 示例：

index	label	review
5612	0	房间小得无法想象,建议个子大的不要选择,一般的睡觉脚也伸不直.房间不超过 10 平方,彩电是 14...
7321	0	我们一家人带孩子去过“五.一”，在协程网上挑了半天才选中的酒店，但看来还是错了。1.酒店除了...
3870	1	周六到西山去采橘子,路过这家酒店的时候就觉得应该不错的,采好橘子回来天也晚了,就临时决定住在...

二、工作环境

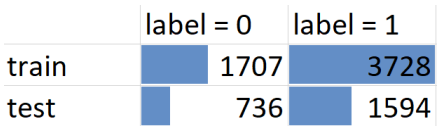
RAM：13G

GPU Memory：16G

三、训练过程

首先划分训练集和测试集。划分后的训练集和测试集的情况如下：

	label = 0	label = 1	sum	%
train	1707	3728	5435	70%
test	736	1594	2330	30%
sum	2443	5322	7765	100%
%	31%	69%	100%	



然后训练模型。训练集上 5435 个样本训练的总耗时为 4 分 55 秒。训练过程中模型返回的信息见下图。

```
input_word_ids (InputLayer)     [(None, 160)]        0
__________________________________________________________________________________________________
input_mask (InputLayer)         [(None, 160)]        0
__________________________________________________________________________________________________
segment_ids (InputLayer)        [(None, 160)]        0
__________________________________________________________________________________________________
keras_layer (KerasLayer)        [(None, 768), (None, 102267649   input_word_ids[0][0]
                                                                 input_mask[0][0]
                                                                 segment_ids[0][0]
__________________________________________________________________________________________________
tf_op_layer_strided_slice (Tens [(None, 768)]        0           keras_layer[0][1]
__________________________________________________________________________________________________
dense (Dense)                   (None, 1)            769         tf_op_layer_strided_slice[0][0]
==================================================================================================
Total params: 102,268,418
Trainable params: 102,268,417
Non-trainable params: 1
__________________________________________________________________________________________________
Epoch 1/3
272/272 [==============================] - 99s 363ms/step - loss: 0.3509 - accuracy: 0.8425 - val_loss:
0.2303 - val_accuracy: 0.9089
Epoch 2/3
272/272 [==============================] - 98s 360ms/step - loss: 0.2349 - accuracy: 0.9002 - val_loss:
0.2120 - val_accuracy: 0.9190
Epoch 3/3
272/272 [==============================] - 98s 360ms/step - loss: 0.1916 - accuracy: 0.9227 - val_loss:
0.2116 - val_accuracy: 0.9144
```

四、预测结果评价

利用测试集的真实 label 和预测的 label 计算出下面 5 个预测结果的评价指标。

Accuracy	recall	precision	ROC	f1 score
0.9030	0.9128	0.9436	0.8973	0.9279

五、未来的工作

1. 本次训练没有调参，所有的超参数都是根据感觉设置的，只 fit 了一次。未来具体使用时，调参过后 F1 score 应该还能上升。

2. 对于数据集的不平衡问题没有进行处理。