NTHU\_NLP Assignment 3 Report

Name: 張芷晴 School: 成功大學 School\_id: NM6131051

Running Environment: Ubuntu 13.2.0

CPU: 2\* Intel Xeon Gold 6138 20C 2.0GHz

GPU: Tesla V100S-PCIE-32GB

Python Version: 3.11.5

* Which (pre-trained) model do you use? Why to choose the model? (5%)

一開始先用助教給的code中預設的bert-base-uncased來測試。

後來在hugging face找一些model來試試看，效果都沒有一開始的好。

從testResult(各model詳細的SpearmanCorrCoef和Accuracy表格在最末頁)來看，表現最好的是amberoad/bert-multilingual-passage-reranking-msmarco和google-bert/bert-base-uncased。

推測amberoad/bert-multilingual-passage-reranking-msmarco效果好的原因是該模型可能具有更好的語境建模能力，因為它的預訓練目的是段落重排序，這需要更精確的句子間關聯建模能力。

而表現較不好的是Davlan/bert-base-multilingual-cased-ner-hrl，推測是因為他們是專門針對 NER 的模型，所以不適合此次的任務。

* Compared with models trained separately on each of the sub-task, does multi-output learning improve the performance? (8%)

請chatgpt和copilot幫我生成單一子任務訓練模型，結果顯示multi-output learning的準確率略高於單一子任務訓練1%。

* Why does your model fail to correctly predict some data points? Please provide an error analysis. (8%)

一張含有 文字, 螢幕擷取畫面, 圖表, 字型 的圖片

自動產生的描述

「'0': NEUTRAL, '1': ENTAILMENT ,'2': CONTRADICTION」

模型在 NEUTRAL 和 ENTAILMENT 兩類上的分類表現比較好,正確率較高。

但在 CONTRADICTION 類別上,模型有較多的錯誤,主要是將 CONTRADICTION 錯誤分類為 NEUTRAL。

ENTAILMENT 類別也有一定程度的錯誤,主要是被錯誤分類為 NEUTRAL。

推測可能原因：

CONTRADICTION 和 NEUTRAL 之間的特徵區分可能不够清晰,模型難以準確區分。ENTAILMENT 和 NEUTRAL 在某些樣本上可能存在一定的重疊,使得模型產生錯誤。

* How do you improve your model performance? (9%)

把資料前處理做好、更換不同的model、增加epoch次數、增加model的layer。

* ... Anything that can strengthen your report. (10%)

以下是我嘗試不同的模型的testset result:

|  |  |  |
| --- | --- | --- |
| Model | SpearmanCorrCoef | Accuracy |
| google-bert/bert-base-uncased | 0.8217 | 87.80% |
| google-bert/bert-base-multilingual-cased | 0.8009 | 85.77% |
| nlptown/bert-base-multilingual-uncased-sentiment | 0.8148 | 85.69% |
| dslim/bert-base-NER | 0.8221 | 86.60% |
| amberoad/bert-multilingual-passage-reranking-msmarco | 0.8266 | 87.56% |
| Davlan/bert-base-multilingual-cased-ner-hrl | 0.7852 | 85.29% |

* Reference:

<https://transformers.run/c2/2021-12-08-transformers-note-1/>

<https://blog.csdn.net/qq_36182852/article/details/121479452>