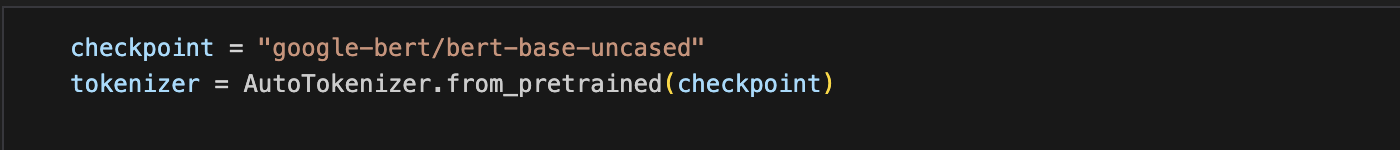
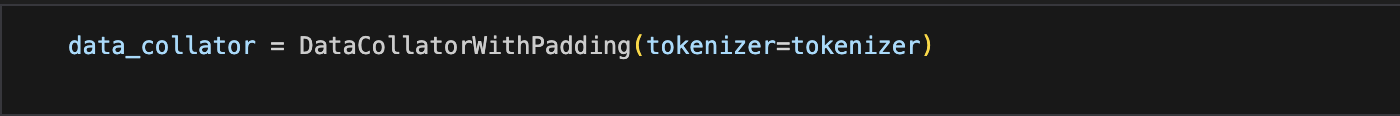
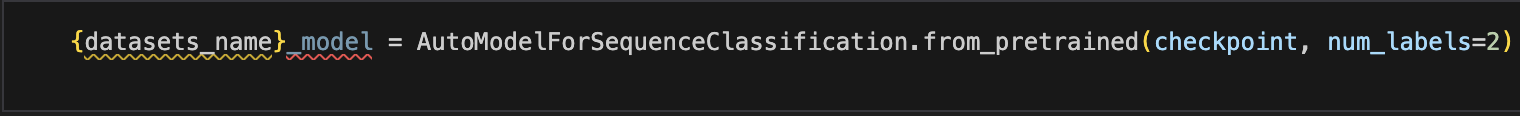
**GAI Project 3 Report**

**PEFT on GLUE benchmarks  
資訊114 F7401254 張暐俊**

1. **Model analysis**

本次使用的模型是google bert底下的bert-base-uncased，以下為導入之模型、tokenizer與data\_collator。  






以下的結果是使用Lora的方式train出來的結果，下表為各種datasets使用之超參數。

|  |  |  |  |
| --- | --- | --- | --- |
| **TrainingArguments** | **SST2** | **CoLA** | **MRPC** |
| **num\_train\_epochs** | 10 | | |
| **learning\_rate** | 5e-4 | | |
| **per\_device\_train\_batch\_size** | 16 | | |
| **per\_device\_eval\_batch\_size** | 16 | | |
| **gradient\_accumulation\_steps** | 1 | | |
| **warmup\_steps** | 500 | | |
| **weight\_decay** | 0.01 | | |
| **evaluation\_strategy** | Epoch | | |
| **save\_strategy** | Epoch | | |
| **save\_total\_limit** | 10 | | |
| **seed** | 42 | | |

|  |  |  |  |
| --- | --- | --- | --- |
| **LoraConfig** | **SST2** | **CoLA** | **MRPC** |
| **r** | 8 | | |
| **lora\_alpha** | 16 | | |
| **lora\_dropout** | 0.01 | | |
| **bias** | none | | |
| **task\_type** | SEQ\_CLS | | |

下圖為三種dataset的train losing rate 跟 validation losing rate結果。

一張含有 行, 文字, 繪圖, 圖表 的圖片

自動產生的描述

1. **PEFT Discussion**
2. Bitfit最主要的重點是他在訓練時只更新bias的參數，因此在這種訓練的方式下，可以大幅度地縮小需要訓練的parameter，同時因為是訓練bias的參數，也能在更新極少量參數的情況達到不錯的效果。

一張含有 文字, 字型, 行, 螢幕擷取畫面 的圖片

自動產生的描述

下表為我能找到對於我訓練使用的三種資料集來說，最好的超參數設定。主要在調整learning rate的部分著手。SST-2及MRPC皆相較於CoLA來說更輕鬆就達到網路上大多資料引用的paper所達到的標準，而CoLA則是我更改了無數次超參數後，最後發現還是只調整 learning rate有相對較好的結果。

|  |  |  |  |
| --- | --- | --- | --- |
| **TrainingArguments** | **SST2** | **CoLA** | **MRPC** |
| **num\_train\_epochs** | 5 | | |
| **learning\_rate** | 5e-4 | 1e-3 | 5e-4 |
| **per\_device\_train\_batch\_size** | 16 | | |
| **per\_device\_eval\_batch\_size** | 16 | | |
| **gradient\_accumulation\_steps** | 1 | | |
| **warmup\_steps** | 500 | | |
| **weight\_decay** | 0.01 | | |
| **evaluation\_strategy** | Epoch | | |
| **save\_strategy** | Epoch | | |
| **save\_total\_limit** | 10 | | |
| **seed** | 42 | | |

1. 而相對於full-finetunig來說，bitfit需要用相對較大的learnig rate來做訓練，因為bitfit模型相對來說較小，需要較大的learning rate協助。

|  |  |  |  |
| --- | --- | --- | --- |
| **TrainingArguments**  **(full-finetuning)** | **SST2** | **CoLA** | **MRPC** |
| **learning\_rate** | 1e-5 | 1e-5 | 1e-5 |

下圖結果為三個情況下的比較，藍色為使用上方表格，也就是在正常情況下的bitfit，其實可以發現僅相對full-finetune的情況分數來的相對較低一些，但是當在bitfit使用與full-finetuning一樣相對較低的learning rate時(下圖紅色)，訓練結果就十分的糟糕。

一張含有 文字, 螢幕擷取畫面, 圖表, 行 的圖片

自動產生的描述

1. **PEFT Comparison**
   1. 下圖為bitfit、lora與full-finetunig三者的評測結果比較圖以及使用之超參數。

|  |  |  |  |
| --- | --- | --- | --- |
| **TrainingArguments** | **SST2** | **CoLA** | **MRPC** |
| **learning\_rate**  **(full-finetuning)** | 1e-5 | 1e-5 | 1e-5 |
| **learning\_rate (bitfit & lora)** | 5e-4 | 1e-3 | 5e-4 |
| **num\_train\_epochs** | 5 | | |
| **per\_device\_train\_batch\_size** | 16 | | |
| **per\_device\_eval\_batch\_size** | 16 | | |
| **gradient\_accumulation\_steps** | 1 | | |
| **warmup\_steps** | 500 | | |
| **weight\_decay** | 0.01 | | |
| **evaluation\_strategy** | Epoch | | |
| **save\_strategy** | Epoch | | |
| **save\_total\_limit** | 10 | | |
| **seed** | 42 | | |

|  |  |  |  |
| --- | --- | --- | --- |
| **LoraConfig** | **SST2** | **CoLA** | **MRPC** |
| **r** | 8 | | |
| **lora\_alpha** | 16 | | |
| **lora\_dropout** | 0.01 | | |
| **bias** | none | | |
| **task\_type** | SEQ\_CLS | | |

一張含有 文字, 圖表, 螢幕擷取畫面, 行 的圖片

自動產生的描述

由上柱狀圖可以發現，相對而言，在使用lora模型會有相對較好的表現，且通常會相對full-finetunig的訓練節果來說還更好一些。而bitfit的訓練結果則是略遜於兩者。

* 1. 根據lora的架構設定以及實作後的結果可以發現r的不同會影響到能訓練的parameter的數量，但是對於訓練的結果而言，影響並沒有那麼的大，而且r的大小也並非是越大越好，或許對於模型以及欲訓練的資料集而言，都有最適合他們的rank(r)。

下表為在不同r的情況下parameter數量的改變。

|  |  |
| --- | --- |
| R=8 | |
| trainable params | 296,450 |
| all params | 109,780,228 |
| trainable% | 0.2700395193203643 |
| R=16 | |
| trainable params | 591,362 |
| all params | 110,075,140 |
| trainable% | 0.537234837947969 |
| R=32 | |
| trainable params | 1,181,186 |
| all params | 110,664,964 |
| trainable% | 1.0673531687951392 |

下圖為在不同r下，不同訓練集的評分結果。

一張含有 文字, 圖表, 螢幕擷取畫面, 行 的圖片

自動產生的描述

下表為使用之超參數

|  |  |  |  |
| --- | --- | --- | --- |
| **TrainingArguments** | **SST2** | **CoLA** | **MRPC** |
| **learning\_rate** | 5e-4 | 1e-3 | 5e-4 |
| **num\_train\_epochs** | 5 | | |
| **per\_device\_train\_batch\_size** | 16 | | |
| **per\_device\_eval\_batch\_size** | 16 | | |
| **gradient\_accumulation\_steps** | 1 | | |
| **warmup\_steps** | 500 | | |
| **weight\_decay** | 0.01 | | |
| **evaluation\_strategy** | Epoch | | |
| **save\_strategy** | Epoch | | |
| **save\_total\_limit** | 10 | | |
| **seed** | 42 | | |

|  |  |  |  |
| --- | --- | --- | --- |
| **LoraConfig** | **SST2** | **CoLA** | **MRPC** |
| **lora\_alpha** | 16 | | |
| **lora\_dropout** | 0.01 | | |
| **bias** | none | | |
| **task\_type** | SEQ\_CLS | | |