**四、研究方法及實驗步驟**

* **研究方法**
* **以BERT對原始文本進行預訓練針對特定下游任務進行微調**

BERT 之運作過程主要基於未標註或只有少量標註之文本數據進行微調以適用於特定下游任務。  
運作過程包括以下三個主要步驟:

1. 準備原始文本數據:

文本數據包含未標註或少量標註之文本，透過數據清理將文本中空白標題的範例與 XML/html 之無效標籤去除，同時將超出BERT模型中預設序列長度的範本去除，並以將小於序列長度之向量補0，以符合預訓練文本固定序列長度讀入。

1. 將原始文本轉換成BERT相容之輸入格式:

文本進行預處理過程中對句子開頭向量位置加入分類符 [CLS]，並以[SEP]區分前一句與下一句。再以中文 BERT 對文本進行斷詞，如圖5。



圖 5,BERT 成對句子編碼示意圖, 來源: BERT 論文

1. 於BERT頂層加入新的Layer進行微調使其適用於特定下游任務:

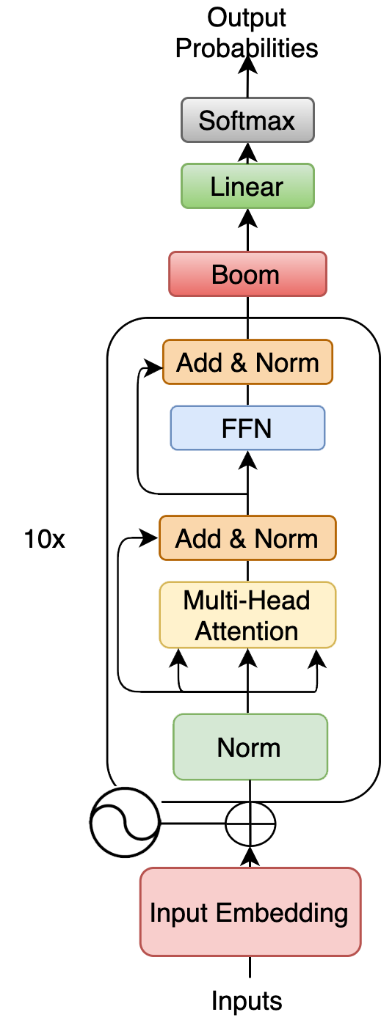
 對BERT模型進行微調包含利用下游任務的目標函式從頭訓練分類器並微調BERT參數，以訓練完的BERT加上線性分類器最大化當前下游任務的目標，如圖6。

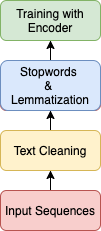
圖 6,BERT分類層示意圖, 來源: Faiza Khattak

透過遷移學習，新增的分類器大多參數均來自已經預訓練的BERT模型，因此實際上需要重新訓練之參數量較少。微調的過程則依不同下游任務需求加入不同線性分類器。

* + - **實驗步驟  
      在基於Transformer之模型中挑選對情感分類資料集表現最佳的模型做為主要分類器與進行後續修改**

實驗模型主要以基於Transformer的四種模型進行比較，此使用 BERT, RoBERTa, XLNet, DistilBERT 對情感分類數據集做分類。

**** 挑選效能最佳之 RoBERTa 做為主要模型，並以Boom Layer改造前饋層做高維度轉換，提取更多的文本特徵與降低整個矩陣運算量。

****

**對文本數據進行預處理 模型架構圖: EDM-RoBERTa**

****  
  
**<<以EDM-RoBERTa與其他實驗樣本對IMDb Dataset做二分類情感分類****<<以EDM-RoBERTa與其他實驗樣本對SST-5: Rotten Tomatoes Dataset做多維度情感分類**  
**<< EDM-RoBERTa之訓練細節與效能比較**

|  |  |
| --- | --- |
| Google Colaboratory Pro | GPU: NVIDIA Tesla V100-SXM2-16GB |
| **CPU:** Intel Xeon(R) @2.00GHz  **OS:** Ubuntu 18.04.5 LTS  **RAM:** 32GB |
| MacBook Pro (16-inch Late 2019) | **GPU:** AMD Radeon Pro 5600M-8GB-HBM2 |
| **CPU:** Intel Core i9-9980HK @2.4GHz **Dual Boot OS:** Ubuntu 18.04.5 LTS with macOS Catalina 10.15.7 (19H2)  **RAM:** 64GB |

實驗過程以 SST-2 與 SST-5 資料集驗證模型效能，實驗環境如下方表格，訓練期間以 Wiki-Text103 與 CC-NEWs對 EDM-RoBERTa 進行預訓練，後續以模型對 SST-3 及 SST-5 進行監督式微調完成模型訓練。