# Information Retrieval and Applications- HW6

# Transformer model

M10915045 施信宏

這次的作業是實作當今耳熟能詳的NLP界的巨人，BERT模型。在還未公布作業6的時候，便對這個網路有很大興趣，也是第一次使用pytorch實作一個大型的網路，在這之前都只是跑跑小模型及簡單分類。而這次的code全都是自己手刻出來的，過程遇到的問題相當多，大部分都在於切資料以及pytorch讀取資料的方式，如dataset，dataloader的使用，經過這作業後也對這兩個class有更深的認識，最讓人煩心的便是，跑一次training或是跑一次testset的預測都十分的久，都要將近一小時左右的時間，但似乎助教後面公布baseline的寫法執行的速度更快，不需一小時即可跑完，且切割資料的方式頗有水準，不像我是將CSV檔全部切割成自己想要的格式後，再由dataloader去讀取，助教是直接使用給予的dataset直接用pandas做修改，不愧是kaggle 專家。最後心累的地方便是找不到alpha，測試集與驗證集的alpha不一樣，在第一版baseline為0.45時，怎麼測就是差baseline0.05分，重跑一次結果或是切割資料都要很久，幸好後面因為公布baseline code降低baseline後，才暫時結束了BERT惡夢。之後期末結束會來看一下助教的思考方式與寫法，想必對我也有很大的幫助，也謝謝這次作業讓我熟悉了pytorch及了解了深度網路在實作上究竟是如何，比起以前只讀理論時有大幅度的進步。