**Report：NLP\_HW2 成功大學\_N96134645\_許宸華**

**Running environment：Colab Python version：Colab**

在回答報告的問題前，我想先大概說明我更動了模型的那些地方。

首先是結果呈現一張含有 文字, 螢幕擷取畫面, 字型, 行 的圖片

自動產生的描述

可以看到模型表現非常差，準確率是0，測試的結果也一蹋糊塗。

一張含有 文字, 字型, 行, 數字 的圖片

自動產生的描述

我是使用助教給的模版進行coding，在觀察訓練過程的時候，有時候loss值的下降會不如預期，因此有調整學習率去改善，但會發現即使loss下降到0結果仍是很差，因此我有想說會不會是在設置id的時候有問題，因為我的padding跟eos放在id的最後面，但調整後是沒有差別的。一張含有 文字, 螢幕擷取畫面, 字型, 行 的圖片

自動產生的描述

因此想要去增強模型，在layer層開啟了bidirectional，讓模型可以雙向的去學習，也增加dropout層防止overfittng(但我的訓練可以說是毫無fitting QAQ)

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

最後覺得是dataset應該有寫錯，讓資料變得奇怪。

**一、What impact does using different learning rates have on model training?**

一張含有 文字, 字型, 螢幕擷取畫面, 數字 的圖片

自動產生的描述在這次實驗中，我總共使用了4種不同的learning rate ，分別是0.001、0.005、0.0001、0.01，但直到做報告的現在，我的模型表現依然很差，準確率甚至是0%，唯一在更改學習率有發現的只有loss值的收斂速度不一樣。  
理論上來說過大的學習率會導致結果發散，loss值也會飄忽不定，然而太小的學習率也會導致模型更新過慢或者loss值無法降到最低，所以據我觀察0.05的學習率是最好的，雖然說我的準確率仍是0%，但我分析不是學習率的問題。

**二、If you use RNN or GRU instead of LSTM, what will happen to the quality  
of your answer generation? Why?**

**三、If we construct an evaluation set using three-digit numbers while the  
training set is constructed from two-digit numbers, what will happen to  
the quality of your answer generation?**

**四、If some numbers never appear in your training data, what will happen to  
your answer generation?**

**五、Why do we need gradient clipping during training?**