NLP HW2

NCKU F64091130 楊采語

|  |  |  |
| --- | --- | --- |
| 訓練模型 | Running environment | Python version |
| System: Ubuntu 22.04  CPU: intel i5 13500  GPU: Nvidia RTX 4090  CUDA: 12.4 | Python 3.8.5 |
| 其他 | Colab | Colab |

大部分code皆有與ChatGPT、Claude討論過再修改

1. What impact does using different learning rates have on model training?

一般來說，較小的learning rates的訓練過程比較穩定(loss不會忽大忽小)，但訓練時間會比較久，而較大的learning rates初期的訓練速度很快，但比較不穩定(loss會忽大忽小)，甚至可能遇到無法收斂的情況。

一張含有 文字, 螢幕擷取畫面, 字型, 黑色 的圖片

自動產生的描述

而上圖為我這次用50%的資料集測試的結果，結果發現learning rates為0.001的是最為適合的，不僅loss最小，訓練速度也稍快一些，但在穩定性卻意外發現learning rates為0.01時是最小的，推測是因為他剛好卡在local min的地方，所以loss很穩定，但卻沒找到最佳解。

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

如下圖測試出的比較表格可以看到，RNN比起GRU跟LSTM loss明顯大很多，而且增加訓練次數也沒有顯著的準確率提升。而GRU跟LSTM會較適合本次的任務，兩者的準確率都蠻高的，而且有穩定在提升，不過LSTM的成效還是最好的。我想主要還是因為本次任務需要模型去捕捉運算符號的加減關係，這需要模型要較長的記憶能力才能達成，所以RNN的成效相對不佳，但在GRU可能就能大概負荷本次任務的訓練長度，可能在算式較長的某幾個數據表現比較不好而已。

|  |  |  |
| --- | --- | --- |
| RNN | GRU | LSTM |
| 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 | 一張含有 文字, 字型, 印刷術, 螢幕擷取畫面 的圖片  自動產生的描述 | 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 |
| 一張含有 文字, 字型, 螢幕擷取畫面, 印刷術 的圖片  自動產生的描述 | 一張含有 文字, 字型, 螢幕擷取畫面, 印刷術 的圖片  自動產生的描述 | 一張含有 文字, 字型, 螢幕擷取畫面, 印刷術 的圖片  自動產生的描述 |
| 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 | 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 | 一張含有 文字, 字型, 螢幕擷取畫面, 印刷術 的圖片  自動產生的描述 |
| 一張含有 文字, 字型, 螢幕擷取畫面, 圖形 的圖片  自動產生的描述 | 一張含有 文字, 字型, 螢幕擷取畫面, 印刷術 的圖片  自動產生的描述 |  |
| 一張含有 文字, 字型, 螢幕擷取畫面, 圖形 的圖片  自動產生的描述 | 一張含有 文字, 字型, 螢幕擷取畫面, 印刷術 的圖片  自動產生的描述 |  |

1. If 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 we generation?

在我的模型中，不太能成功解三位數的數學，模型在原本驗證集的準確度高達88.40%，但將計算式換成都是三位數的輸入時，準確度下降來到28.60%了。

|  |  |
| --- | --- |
| 原始 | 計算3位數 |
|  |  |

1. If some numbers never appear in your training data, what will happen to your answer generation?

在本次的實驗結果中，我將訓練集中有7的model都拿掉再進行訓練，可以發現有7的答案都預測不出來了，而在其他資料也可能因為都少了7的計算導致準確度下降許多。

|  |  |
| --- | --- |
| 原始 | 拿掉7 |
|  | 一張含有 文字, 螢幕擷取畫面, 字型, 設計 的圖片  自動產生的描述 |

1. Why do we need gradient clipping during training?

因為RNN/LSTM中的梯度會隨時間步累積，可能會出現梯度爆炸問題，而且較長的序列可能導致梯度指數增長，gradient clipping可以限制梯度範圍，防止參數更新過大，讓模型訓練更穩定。另外因為這個算術運算任務的序列長度不一、且計算符有依賴關係，所以需要用到長期記憶更容易出現梯度指數增長的問題，故使用gradient clipping來限制梯度計算。

下圖是我拿不同gradient clipping的值進行測試的圖片，可以發現不同的gradient clipping參數在本次的任務中影響不太大，可能本次任務的序列不算到太長，較不會出現梯度爆炸的情況。

