MLDS 2017 Spring HW4 - Seq2Seq + Reinforcement Learning

B03901056 孫凡耕 B03901070 羅啟心
B03901032 郭子生 B03901003 許晉嘉

1 Environment

OS	CPU	CPU Memory	GPU	GPU Memory
`Arch linux 4.10`	`i7 3.6 GHz`	`32 GB`	`GTX 1080 Ti`	`11 GB`

2 Data Sets

- Open subtitles(`http://opus.lingfil.uu.se/download.php?f=OpenSubtitles/en.tar.gz`)
- Movie subtitles(`http://www.mpi-sws.org/~cristian/Cornell_Movie-Dialogs_Corpus.html`)

3 Model description

1. Seq2Seq 模型：
   典型的 encoder-decoder 的模型，也就是利用 LSTM 把不同長度的輸入，轉化為固定大小的 state，然後將此 state 傳遞給另個 LSTM，並依序輸出當前最佳解直到遇到 `<eos>` 爲止。
   模型所使用的參數如下：
   - learning rate = 0.5
   - learning rate decay factor = 0.99
   - max gradient norm = 5.0
   - batch size = 64
   - vocab size = 100000
   - size of each model layer = 256, 512, 1024
   - number of layers = 4
2. Reinforment Learning 模型:
   我們的模型參考自 Adversarial Learning for Neural Dialogue Generation [*] 這篇 paper，並改寫自該篇 paper 所提供的程式 [†]，模型所使用的參數如下：
   (a) Generator 與 Discriminator update 的比例爲 Generator 一次及 Discriminator 四次。
   (b) Generator: 大部分參與與上方 Seq2Seq 相同。
   - dropout = 0.5

   (c) Disciminator:
   - Hierarchical encoder[‡]
   - size per layer = 512
   - number of layers = 4
   - learning rate = 0.2
   - dropout = 0.5
   - max gradient norm = 5

[*] Jiwei Li, 2017, Adversarial Learning for Neural Dialogue Generation
[†] https://github.com/liuyuemaicha/Adversarial-Learning-for-Neural-Dialogue-Generation-in-Tensorflow
[‡] Jiwei Li, 2015, A Hierarchical Neural Autoencoder for Paragraphs and Documents

(d) Reward function: 跟 GAN 的概念一樣，Generator 讀進來一個句子之後，會產生相對應的回答，Discriminator 再根據回答計算有多少機率是人或機器所產生的句子，若某個回答越接近人所產生的句子，則 reward 越接近 1，反之，則越接近 0。

以 x 表示前兩句話，y 表示 Generator 所產生的句子，以 $Q_+(\{x, y\})$ 表示爲人所產生的機率。則 reward function 爲 $J(\theta) = \mathbb{E}_{y\sim p(y|x)}(Q_+(\{x, y\})|\theta)$。則以 likelihood ratio trick 近似後可得 $\Delta J(\theta) \approx [Q_+(\{x, y\}) - b(\{x, y\})]\Delta \sum_t \log p(y_t|x, y_{1:t-1})$。

此外，爲了增加 reward 對於 Generator 的影響，對於 Generator 所產生的每個子句，都會以蒙地卡羅搜尋五次，以五次的平均作爲這個子句的 reward。最後，爲了讓 Generator 能持續產生好的句子，而不是突然找不到好的方向，每次 update 完 Discriminator 及 Generator 後，還會在 true data 上對 Generator 進行 update (Teacher forcing)。

4 Improvement

1. 在 Seq2Seq 模型中，我們嘗試在相同的 data set 之下，去調整模型的大小，也就是每層 layer 中所含的 cell 數量，對於不同 cell 數量，perplexity 下降的速度可以從 Figure 1 以及 Figure 2 中看出。對於相同的 data set，如果一層的 cell 數量越多，所能蘊含的資訊量便越多，因此 cell 數量越多的情況下，perplexity 下降的速度便會越快。也就是說，模型中的大小越大，訓練的速度也會較爲快速。

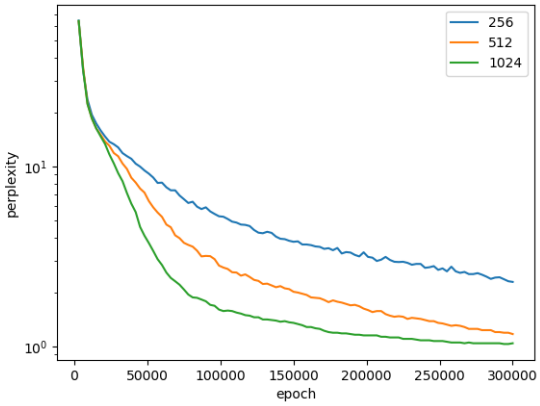

Figure 1: 以 Movie subtitles 作爲 data set

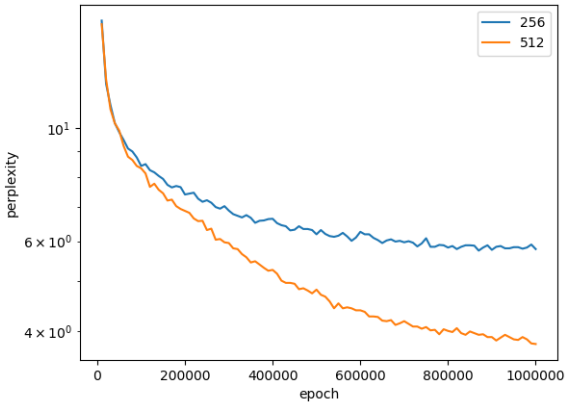

Figure 2: 以 Open subtitles 作爲 data set

2. 我們以 movie subtitles 作爲 data set 時，將太短（少於三個字）以及太長（長於四十個字）的句子先刪除，以及出現特殊符號（非字母數字）的句子也剔除，之後便剩下約莫 12 萬句對，大約剩下原先的一半。同時將單字量調整至三萬及五萬，所訓練出來的結果如下：

單字量	30000	50000
how are you?	hello , secure .	fine . domestic .
how old are you?	I 'm a little nervous .	gibarian . buckle wallace .
Where are you from?	my room 's here .	she 's a lawyer owen owen
What's your name?	you know my name .	gibarian fidget .
Sounds great!	what 's he got ?	muskets . buckle jeff .
Looks funny!	I 've got a trunk	i dined to see.
What's that?	half-red , half-black –	denning 's log .

從上表可以看出，將短的句子去掉，可以使輸出更爲有趣，將長的句子去掉，可以加速訓練的過程，但由於句子的減少，單字量也必須減少，否則，會使模型無法訓練起來。

3. 在 Reinforcement Learning 的模型中，雖然我們所參考的 paper [§] 中表示，Discriminator 與 Generator 更新的比例爲 5:1，但是我們發現 Generator 更新次數較多有利於讓 Generator 回答的結果較爲符合。這是由於每次更新 Generator 都會做 Teacher Forcing，因此能使 Generator 產生更爲

[§]Jiwei Li, 2017, Adversarial Learning for Neural Dialogue Generation

合理的句子。

輸入句子	輸出回應
how are you?	fine , macaulay .
how old are you?	i 'm thirty-seven .
Where are you from?	my russian _UNK .
What's your name?	someone else . called me .
Sounds great!	he 's eating the realm !
Looks funny!	you once you 're onto .
What's that?	half-red , half-black –

5 Experiment

1. 在 Seq2Seq 模型中，我們實驗將相同 data set 相同模型下，不同 perplexity 時，模型所作出的回應如下：

   (a) data set 爲 movie subtitles 、模型大小爲 4×512：

perp	輸入句子	輸出回應	perp	輸入句子	輸出回應
30	how are you?	no .	10	how are you?	fine , fine .
	how old are you?	no .		how old are you?	older .
	Where are you from?	no .		Where are you from?	west city .
1	how are you?	fine , fine .			
	how old are you?	twenty-eight .			
	Where are you from?	california .			

   (b) data set 爲 movie subtitles 、模型大小爲 4×256：

perp	輸入句子	輸出回應	perp	輸入句子	輸出回應
30	how are you?	no .	10	how are you?	i ' m fine .
	how old are you?	i have him .		how old are you?	o .
	Where are you from?	i have him .		Where are you from?	california .
1	how are you?	fine .			
	how old are you?	thirty-five .			
	Where are you from?	meet me .			

   從不同的 perplexity 之間可以看出，明顯地，perplexity 越低，所回答出的句子越佳，其中模型大小爲 4×512 的部分在 perplexity 爲 1 時，所回答出的句子幾乎完全可以視爲人話。但模型大小爲 4×256 的部分在 perplexity 爲 1 時，仍然有若干的句子回答不佳。我們認爲這是由於模型大小較小，因此所能儲存的資訊量較小的緣故所致。

2. 在 Seq2Seq 模型中，我們嘗試以相同的模型，以不同的 data set 作爲訓練資料，來比較以不同 data set 訓練後的模型，對於同樣的問句會有如何的回答。(以下結果皆爲 perplexity≤ 3 的情況)

   (a) 以下爲模型大小爲 4×256 的結果：

輸入句子	open subtitles	movie subtitles
how are you?	i ' m fine .	fine .
how old are you?	00 .	thirty-five .
Where are you from?	i ' m from the new york .	meet me .
What's your name?	i ' m your name .	star .
Sounds great!	what ?	yeah , i got it !
Looks funny!	you ' re a good man .	is this your shovel and your husband ?
What's that?	what ?	what ?

   (b) 以下爲模型大小爲 4×512 的結果：

輸入句子	open subtitles	movie subtitles
how are you?	good .	fine , fine .
how old are you?	00 .	twenty-eight .
Where are you from?	you ' re from texas .	southern california .
What's your name?	you know what ?	lisette .
Sounds great!	no .	let ' s get out of here .
Looks funny!	you know what ?	what ?
What's that?	what ?	what do you mean ?

從以上的結果可以看出，在模型大小不夠大的時候，以 open subtitles 及 movie subtitles 爲 data set 的結果相去不遠，又以 open subtitles 的部分較爲像人所作出的回應。但將模型大小擴大之後，可以看出 open subtitles 的部分並沒有顯著的進步，因此可以推斷出 open subtitles 的資料量也許較爲簡單，而將模型擴大之後，movie subtitles 的結果便有十分顯著的進步，幾乎所有的回應都有相當程度的貼近人話。可以推斷，movie subtitles 的資訊量較爲完整，但也需要較大的模型。

3. 在 Reinforcement Learning 的部分，由於我們 data 的量並不夠多，加上 model pretrained 的部分也不夠多，以及我們的 RL model 是基於前兩句來推測下一句，但實際上我們的資料是一句對一句，因此，模型實際上並不太正確。而且當 Discriminator 更新次數比較少的時候，便會出現答非所問的現象，這個現象在訓練越久便會越明顯！下表爲 Reinforcement Learning Generate 更新次數與 Discriminator 更新次數爲 1:1 時的結果。

輸入句子	輸出回應
how are you?	head for bridge .
how old are you?	i didn't see her .
Where are you from?	my russian boys .
What's your name?	william simpson . robert rath .
Sounds great!	sally , kitchen .
Looks funny!	i 'm sorry ... lee lother 's a very ...
What's that?	i 'm getting .

4. 因爲 Generator 與 Discriminator 需要保持平衡的狀態，我們所訓練出來的模型，所觀察到的 Reward per sentence 大致上維持在 0.4 左右。然而，如果去掉 Teacher Focing，則 Generator 即使在有 pretrain 的情況下，仍然很難獲得 Reward。如 Figure 3 所示，Reward 基本都維持一致。但不確定原因爲何如 Figure 4 所示 Teacher Loss 會上升。

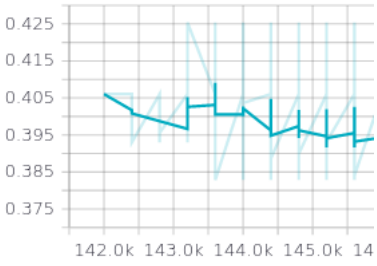

Figure 3: Reward

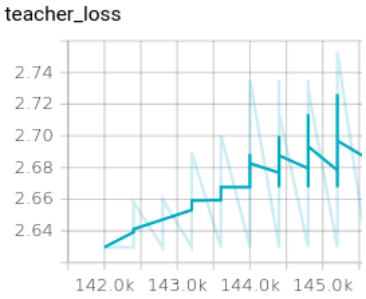

Figure 4: Teacher Loss

6 Team division

孫凡耕	RL、分配組內工作、教導組員
羅啟心	協助餘項事務
郭子生	協助餘項事務
許晉嘉	Seq2Seq、統整撰寫報告、跑實驗