

MLDS HW4 Report

Members

D04944007 高瑞宗

R05922139 林子芃

R04921106 陳彥谷

B02902105 廖瑋中

Environment

我們用自己實驗室的Azure雲端計算資源：

Name	Cores	Memory	OS	GPU
msslbgpu1	6	56G	Ubuntu 16.04	Tesla K80, 12G Ram

Library We Used

無。

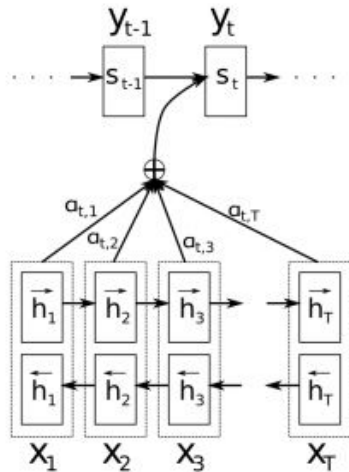
Datasets

在 slides 中，助教提供了三種 dataset 給我們使用。但是，經過使用後發現，twitter 的 dataset 太過口語與複雜，造成 model 難以學習。而 Cornell 的 dataset 雖然品質不錯，但是資料量太小。因此，最後我們主要是使用 Open Subtitles 這個 dataset 來訓練我們的 model。

Model Description

seq2seq + attention + anti-LM + RL

seq2seq + attention



我們使用了和作業二同樣的 seq2seq + attention model，並且使用 tensorflow 內建的 seq2seq 和 attention 來實作，就可以做到基本的對話。可是這樣的話會有幾個問題，一是train出來機率最高的回覆都會是幾個generic的response，二是回答有可能都很簡短無趣。

anti-LM (Minimum Mutual Information)

令 $S = \{s_1, s_2, s_3, \dots\}$ 為 input sequence, $T = \{t_1, t_2, t_3, \dots\}$ 為 response sequence。
正常 seq2seq 的 objective function 會是：

$$\hat{T} = \arg \max_T \{ \log p(T|S) \}$$

這樣做的話很容易造成train出來的model會產生很統一性的回覆，像是「I'm sorry.」或「I don't know.」。所以 anti-LM 的作法是把 objective function 換成：

$$\hat{T} = \arg \max_T \{ \log p(T|S) - \lambda \log p(T) \}$$

後面加上 $-\lambda \log p(T)$ 為產生統一性回覆的 penalty，而 λ 是控制這個 penalty 的大小的變數。因為太常出現的 response 會因為這個作法受到 penalty，所以基本上就可以解決 generic response 的問題。

Reinforced Learning

這裡 RL 做了兩件事：

- **Sentence Length Reward** : 簡單來說就是在上述的 objective function 中加入句子長度的reward, 越長的句子reward越高。

`all_prob = all_prob_ts - args.antiilm * all_prob_t + args.n_bonus * dptr`

- **Information Flow** : 如果意思太相近的句子一直連續出現, 就會受到penalize。這樣是為了讓聊天的人不會一直連續得到類似的回覆。

$$r_2 = -\log \cos(h_{p_i}, h_{p_{i+1}}) = -\log \cos \frac{h_{p_i} \cdot h_{p_{i+1}}}{\|h_{p_i}\| \|h_{p_{i+1}}\|}$$

h_{p_i} 和 $h_{p_{i+1}}$ 為在 p_i 和 p_{i+1} 時間得到的兩個句子的 vector representation, 而 reward 為這兩個 vector 的 cosine similarity 的 negative log。

Final Objective Function

$$T = \operatorname{argmax} \{ \log p(T|S) - \log p(T) * \lambda + \text{sentence_length} * \text{length_reward} - \log \cos(h_{p_i}, h_{p_{i+1}}) \}$$

How to improve performance

1. 我們試著前處理一下資料, 使得原始文本資料中一些縮寫或是口語用法統一成正式用法。並且, 我們過濾掉一些無謂的符號, 讓整個文本更加乾淨一點。如此一來, 在建立字典的時候, 就可以少掉許多奇怪的符號與縮寫, 讓字典的品質更好。
2. 一開始我們使用助教提供給我們的三種 dataset。但是經過使用後發現, twitter 的 dataset 太過口語與複雜, 造成 model 難以學習。而 Cornell 的 dataset 雖然品質不錯, 但是資料量太小。因此, 最後我們決定使用 Open Subtitles 這個 dataset 來做訓練。因為挑選過了 dataset, 也讓我們 model 學習的更好了。
3. 在讀入 source 和 target 的 dialog 時有加入 bi-directional model, 所以會有顛倒的 Q&A pair。可以用 question 來 train answer 也可以用 answer 來 train question。

Experiment Settings and Observations

Model	seq2seq	seq2seq + anti-lm + RL
steps	405500	~400000 (seq2seq + anti-lm) ~200000 (seq2seq + anti-lm + RL)
batch size	64	64 (seq2seq + anti-lm) 1 (seq2seq + anti-lm + RL)
training time	~5days	~5days

learning rate	0.05	0.05
vocabulary size	80000	80000
number of layers	4	4
features in each layer	128	128

以下為單純 seq2seq 和 seq2seq+anti-lm+RL 的比較：

```

Hello. : (2017-06-15 17:31:34.365904)
(1) -> hello ?
How are you? : (2017-06-15 17:31:34.736139)
(1) -> i ' m sorry .
What is your name? : (2017-06-15 17:31:36.729045)
(1) -> i ' m sorry .
What are you talking about? : (2017-06-15 17:31:37.267526)
(1) -> i ' m sorry .
Do you like to play basketball? : (2017-06-15 17:31:37.878812)
(1) -> no .
Tell me about yourself. : (2017-06-15 17:31:38.545308)
(1) -> i ' m sorry .
What is your favorite sport? : (2017-06-15 17:31:39.146809)
(1) -> i don ' t know .
Do you love me? : (2017-06-15 17:31:39.809946)
(1) -> i ' m sorry .
Are you hungry? : (2017-06-15 17:31:40.191229)
(1) -> no .
What is your favorite thing? : (2017-06-15 17:31:40.809335)
(1) -> i ' m sorry .
Give me some food please. : (2017-06-15 17:31:41.429953)
(1) -> i ' m sorry .

```

seq2seq

```

Hello. : (2017-06-15 17:56:09.230226)
(1) -> hello .
How are you? : (2017-06-15 17:56:09.614437)
(1) -> yeah .
What is your name? : (2017-06-15 17:56:11.751194)
(1) -> i ' m sorry .
What are you talking about? : (2017-06-15 17:56:12.473273)
(1) -> no .
Do you like to play basketball? : (2017-06-15 17:56:13.091184)
(1) -> yeah .
Tell me about yourself. : (2017-06-15 17:56:13.779300)
(1) -> i ' m not gonna be .
What is your favorite sport? : (2017-06-15 17:56:14.486895)
(1) -> yeah .
Do you love me? : (2017-06-15 17:56:15.217868)
(1) -> no .
Are you hungry? : (2017-06-15 17:56:15.582427)
(1) -> no .
What is your favorite thing? : (2017-06-15 17:56:16.267783)
(1) -> yeah .
Give me some food please. : (2017-06-15 17:56:16.983056)
(1) -> i ' m not gonna be .

```

seq2seq + anti-lm + RL

可以看出有 anti-lm 和 RL 的 model 中，「I'm sorry」出現的次數明顯減少。雖然有句子長度的 reward function，但是在這幾個對話中還是看不出明顯的差別。

Team Division

Member	Work
D04944007 高瑞宗	Basic seq2seq
R05922139 林子芃	Data preprocess + seq2seq + RL + report
R04921106 陳彥谷	-
B02902105 廖瑋中	Data preprocess + seq2seq + RL + report