April 9, 2020

Text Generation with RNNs

Homework Assignment by prof.Huaque

Sangjee Dondrub
sangjeedondrub@live.com
Qinghai Normal University

Contents

1 Github

https://github.com/sangjeedondrub/text-gen-homework

2 Environment

```
pip install --upgrade tensorflow
```

```
import tensorflow as tf

print(tf.__version__)
```

```
2.1.0
```

3 Data

```
head -n 10 lu-drub-gong-gyan.txt
```

དབུ་མའི་ཟབ་གནད་སྟིང་ཡོར་དྲིལ་བའི་ལེགས་བཤད་ཀུ་སུམ་དཊོར་སྐྱབ་དགོངས་རྒྱན་ཞེས་བྱ་བ་བཞུགས་སོ།།
དགེ་འདུན་ཆོས་འཕེལ།
བདུད་ཀྱི་མཚོན་ཆ་རྩྭབ་མོ་དག་ལ་འཛམ་མཉེན་མེ་ཏོག་ལན་དུ་སྤྱིན།།
ལྷས་ཅིན་ཁྲོས་པའི་སྐྱེགས་རྟོ་འཕངས་ཆེ་མི་སྣའི་བརྡུལ་ཞུགས་དང་དུ་བླངས།།
རང་གི་དབུ་ལ་སྤང་མིག་ཙམ་ཡང་བལྟ་བར་མི་ནུས་ཤཱཀྱའི་སྲས།།
འཇིགས་རུང་འཕོར་བའི་དགྲ་ཆེན་སྐྱོབས་པའི་ཐོགས་སུ་ཤེས་ལྡན་སུ་ཡིས་བཀུར།།
དགེ་ལེགས་ཞི་བསིལ་བདུད་རྩིའི་འབྱུང་གནས་ཐར་པའི་ལམ།།
འདུན་པས་ཐུ་ཞེབ་མ་ནོར་མཆོག་པ་སྲིད་པའི་མིག།
ཆོས་རྗེ་སྣ་བའི་སེང་གེ་མཆོག་ནི་འདིའོ་ཞེས།།
འཕེལ་འགྲིབ་མེད་པར་རིགས་སྣ་ཚོགས་ཀྱིས་རྟག་ཏུ་གུས།།
གང་གི་མཁྱེན་པའི་ཉི་མའི་དཀྱིལ་འཁོར་ཀུན་ཏུ་བཟང་པོའི་མཁའ་དབྱིངས་ནས།།

Total number of lines

```
wc -l lu-drub-gong-gyan.txt
```

```
918 lu-drub-gong-gyan.txt
```

Table 1. As Character

Total number of characters	81674
Total vocab	65
Total Patterns	81664

Table 2. As Syllable

Total number of syllables	21755
Total vocab	1189
Totoal pattern	21752

4 Models

Character model

```
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
embedding_1 (Embedding)      (64, None, 256)           16640
_________________________________________________________________
gru_1 (GRU)                  (64, None, 1024)          3938304
_________________________________________________________________
dense_1 (Dense)              (64, None, 65)            66625
=================================================================
Total params: 4,021,569
Trainable params: 4,021,569
Non-trainable params: 0
```

Syllable model

```
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
embedding_4 (Embedding)      (1, None, 256)            304384
_________________________________________________________________
gru_4 (GRU)                  (1, None, 1024)           3938304
_________________________________________________________________
dense_4 (Dense)              (1, None, 1189)           1218725
=================================================================
Total params: 5,461,413
Trainable params: 5,461,413
Non-trainable params: 0
_________________________________________________________________
```

5 Train and evaluate

```
git clone https://github.com/sangjeedondrub/text-gen-homework --depth=1
cd text-gen-homework
```

Train character-level model

```
python text-gen.py
```

Train and evaluate syllable-level model

```
python text-gen.py --use_syllable
```

The evaluation results will be saved to `sample.syllable.txt` and `sample.char.txt` files