DD2424 - Assignment 4 - Bonus

Thi Thuy Nga Nguyen

1 Read in data

Data was read from JSON files of Donald Trump's tweets which contains 36307 tweets. Each tweet contains informations such as source, id, text, time, ... but I only need *text* for this assignment. *Text* includes ASCII characters, html links, emoticons, Chinese, Japanese words, ... For the training purpose, I converted every possible characters to ASCII and removed the other non-ASCII ones. Sometimes there is url link in a tweet which I am not interested in to train my model. I thus removed them. The number of characters (vocab_size) is 94 at the end.

Since to be able to train tweet by tweet I added "@\n" (\n is new line) at the end of each tweet.

2 Train RNN

For Trump's tweets, I set a hard limit of 140 characters when synthesizing (sampling) text and the hprev was reinitialized to 0 at the end of each training tweet. Since training on tweet, I reduced the sequence length to 16 in order to keep more training data. There are 1116 tweets which length is shorter than 25 while there are 255 ones which length is shorter than 16.

I trained my RNN as above with 100 hidden states, the sequence length is 16 and using AdaGrad optimizer with learning rate $\eta = 0.1$, $\epsilon = 1e - 8$. After 5000000 updates I got the smooth loss after each 20000 updates in figure 1.

2.1 Evolution of the text synthesized

With the same setting as above, RNN gave 140 synthesized characters as follows:

- Before the first update step:

```
1   A#/XP(rY|LEo!5 Xi?)#_]h−PK5{gl:∗aZ3(Tw@J}x:Bao}K="0'"/8s[QQ=x]B∗0LF
    }{[B,qui38|zFn&O_Wj/1,Ua{
2   8NE[åĂŹM?RkK'oJ;LvmBaJ:CKA.3YjX&j3D w+0HWL+{ZRe14D
3
```
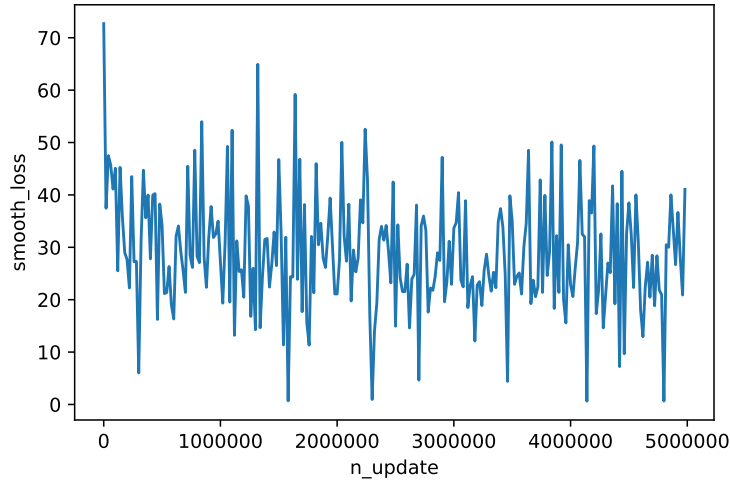
- Before 100000th update step:

Figure 1: The smooth loss after every 20000 updates after training RNN for 5000000 updates with $m = 100, seq\_length = 16$, and AdaGrad optimizer with $\eta = 0.1, \epsilon = 1e - 8$.

```
1    @Misihete and modan now divionbfreman shoustens adering of foren
     wour faathonar! T's 15n @realDonaldTrump Merosnors hay berrom
     toon Nov dal
2
```

- Before 200000th update step:

```
1     ing never, shiads an bully stolly, way you in Dall: To Stada,
     preside workAa cheme in the USA mave the 8/+. Pook TI9%. bat
     makivis hean wher
2
```

- Before 300000th update step:

```
1    alDonaldTrump witer do to deadal a med sade pennouls, bulks
     ollaka! #Malysnims all on the havined in to our toughe on Sture."
     @realDonaldTrum
2
```

- Before 400000th update step:

```
1     ein amnew a Sails and hossed persbark!"
2   Stigh say Trump Got Affeesoinetiendevesting fohshardanding bay the
     @fexterThous. #frid what shouts
3
```

- Before 500000th update step:

```
1    ashment Libwe and see for the need Go. Trump "Trump hat of nned.
      Must cenores @
2   Laid stardan 10...  − bele my the ObamaCare arus trump need
3
```

- Before 600000th update step:

```
1    @Tusteral Asgrin angoust we loven biggets byJonks capstondefcers
     you inthat: @MEshinghc52: @realDonaldTrump he skengle raile
     presillagrespet
2
```

- Before 700000th update step:

```
1    redJTrertbattent ocher my Oy wially the will times &amp; fight!
     Migsengaldyfe to we the plectiefficaâĂŹs − suce peaddiodistions
     stazz. and wher
2
```

- Before 800000th update step:

```
1    t in at Ass more w/ Obama slaising so lavis She, of their thit
     how bad, lacconately, @TNNCLAgraziki ont).− Trump, we when
     eceismative love e
2
```

- Before 900000th update step:

```
1    e no a burtery of or our the od President. I Past at 2ode &gatce.
     T@ark one becting and IS: out have in a reavie littary aboosesed
     great in
2
```

- Before 1000000th update step:

```
1    edits" Trump has eratial frumply most appront ever soo Robverdey.
     In besping Americare." have with (formly to Iramme. Send and
     morn agial wr
2
```

2.2   Text synthesized from the best model

After training 5 million updates, the lowest loss after each epoch was ∼26 and the 4 ×
140 characters synthesized from the best model (corresponding to the lowest loss) is:

```
1   ior our frart every monsiores &amp; vote tedeneded fringle to to and
     sponitecterthtasters, have of thought with Obala would of Crolping has
     P
2
3   alither comarns0 the great iccuntersiem an Mid everyone puile President
     everens new is  Mumpting State our eight and tarver I. President suc
```

```
4
5   @Jamece__Obama Es @BarackObama. Thanks Fake comes elersed, phays
      sttimes of hears"  Obama changy a is to helpe. RookObance dget, Obama'
      s ewi
6
7   llishicoun China"  $301:    @Seedmanes: @scamer, Obama I. Will as the
      newday badly to our inclo starticlable take so have to Get Flart." Why
```